International Journal for Research in Education

Volume 44
Issue 3 *Vol.44, Issue 3 - October 2020*

Article 5

2020

# A Feasibility Study for Developing a Computerized Adaptive Form of Raven's Colored Progressive Matrices Test for Omani Children Based on the Item Response Theory

Ehab Mohammed Omara
*Sultan Qaboos University, Oman and Ain Shams University, Egypt*, ehab@squ.edu.om

علي مهدي كاظم
amkazem@squ.edu.om

Follow this and additional works at: https://scholarworks.uaeu.ac.ae/ijre

Part of the Educational Assessment, Evaluation, and Research Commons, and the Educational Psychology Commons

## Recommended Citation

Omara, Ehab Mohammed and 2020) علي مهدي ,كاظم) "A Feasibility Study for Developing a Computerized Adaptive Form of Raven's Colored Progressive Matrices Test for Omani Children Based on the Item Response Theory," *International Journal for Research in Education*: Vol. 44 : Iss. 3 , Article 5.
Available at: https://scholarworks.uaeu.ac.ae/ijre/vol44/iss3/5

This Article is brought to you for free and open access by Scholarworks@UAEU. It has been accepted for inclusion in International Journal for Research in Education by an authorized editor of Scholarworks@UAEU. For more information, please contact j.education@uaeu.ac.ae.

UAEU كلية التربية College of Education
جامعة الإمارات العربية المتحدة United Arab Emirates University
IJRE

المجلة الدولية للأبحاث التربوية
International Journal for Research in Education

Vol. (44), issue (3) October 2020 المجلد (44) العدد (3) أكتوبر 2020

# A Feasibility Study for Developing a Computerized Adaptive Form of Raven's Colored Progressive Matrices Test for Omani Children Based on the Item Response Theory

# دراسة جدوى تطوير صورة تكيفية محوسبة من اختبار المصفوفات المتتابعة الملونة لرافن للأطفال العمانيين باستخدام نظرية الاستجابة للمفردة

| Received تاريخ الاستلام | 06/08/2019 | Accepted تاريخ القبول | 16/01/2020 | Published تاريخ النشر | 1/10/2020 |
|---|---|---|---|---|---|

DOI : http://doi.org/10.36771/ijre.44.3.20-pp142-181

**Dr. Ehab Mohammed Naguib Omara**
College of Education, Sultan Qaboos University, Oman
College of Education, Ain Shams University, Egypt
ehab@squ.edu.om

**د. إيهاب محمد نجيب عمارة**
كلية التربية – جامعة السلطان قابوس – سلطنة عمان
كلية التربية – جامعة عين شمس – جمهورية مصر العربية

**Prof. Ali Mahdi Kadhim**
College of Education, Sultan Qaboos University, Oman

**أ.د. علي مهدي كاظم**
كلية التربية – جامعة السلطان قابوس – سلطنة عمان

حقوق النشر محفوظة للمجلة الدولية للأبحاث التربوية
ISSN 2519-6146 (Print) ISSN 2519-6154 (Online)

# A Feasibility Study for Developing a Computerized Adaptive Form of Raven's Colored Progressive Matrices Test for Omani Children Based on the Item Response Theory

## Abstract

The current study investigated the feasibility of developing a computerized adaptive form of Raven's Colored Progressive Matrices test, one of the most important culture-free intelligent tests, using the Item Response Theory. The test consists of 36 items divided into three groups. The data used in the current study were adapted from the study by Kadhim et al. (2008), which included 1042 subjects, aged 5 to 10 from both genders and distributed into 11 Omani governorates. Item Response Theory assumptions were met and then the mirtCAT package was used to evaluate a computerized adaptive form of the test. Raven's test items were compatible with the three-parameter model which was used to scale the test items. The Maximum Fisher Information method was used to select items in the adaptive form. The full and the adaptive forms were compared to each other across the different simulated conditions in the current study. The results indicated that using 17 items of the adaptive form could accurately estimate the subjects' abilities without a substantial loss of information. The previous finding is a preliminary indication of the possibility of developing an adaptive form of the colored progressive matrices test that can be used in various assessments that practitioners may need to assess, classify or diagnose children.

*Keywords*: Item Response Theory, computerized adaptive test, colored progressive matrices test

المجلة الدولية للابحاث التربوية- جامعة الإمارات العربية المتحدة
المجلد ( 44 ) العدد ( 3 ) أكتوبر 2020

# دراسة جدوى تطوير صورة تكيفية محوسبة من اختبار المصفوفات المتتابعة الملونة للأطفال العمانيين باستخدام نظرية الاستجابة للمفردة


## مستخلص البحث

هدفت الدراسة الحالية إلى دراسة جدوى تطوير صورة تكيفية محوسبة من اختبار رافن للمصفوفات المتتابعة الملونة باستخدام نظرية الاستجابة للمفردة، لكونه واحداً من أهم اختبارات الذكاء المتحررة ثقافياً، ويتكون الاختبار من ست وثلاثين مفردة مقسمة إلى ثلاث مجموعات. وقد تم الاعتماد على عينة تقنين الاختبار في دراسة كاظم وآخرون (2008) على البيئة العمانية، والتي بلغت 1042 مفحوصاً من الجنسين موزعين على 11 محافظة وعلى الفئات العمرية (5 – 11). ولتحقيق الهدف من الدراسة تم التحقق من افتراضات نظرية الاستجابة للمفردة، ومن ثم استخدمت البرمجية mirtCAT في تقييم الصورة التكيفية باستخدام منهجية المحاكاة. وقد أظهرت مفردات اختبار رافن مستوى جيد من الملاءمة للنموذج الثلاثي البارامتر الذي استخدم في تدريج مفردات الاختبار. وباستخدام طريقة انتقاء المفردات وفقاً للمعلومات القصوى مع أطوال مختلفة من المفردات لإنهاء الاختبار، تم تقييم إمكانية تكوين صورة تكيفية من الاختبار بالمقارنة بين قدرات المفحوصين المقدرة وفقًا للاختبار ككل والصور التكيفية، ولقد أظهرت النتائج درجة جيدة من التقارب بين مستويات القدرة باستخدام الصورة التكيفية بعدد مفردات يصل إلى 17 مفردة دون فروق جوهرية في المعلومات ودقة تقدير القدرة. وتُعد النتيجة السابقة مؤشراً أولياً على إمكانية تطوير صورة تكيفية من اختبار رافن للمصفوفات المتتابعة الملونة يمكن استخدامها في التقييمات المختلفة التي قد يحتاجها المسؤولون في تقييم أو تصنيف أو تشخيص الأطفال.


*الكلمات المفتاحية:* نظرية الاستجابة للمفردة، الاختبار التكيفي المحوسب، اختبار المصفوفات المتتابعة الملونة

## المقدمة والخلفية النظرية

يهتم الباحثون والمتخصصون في مجال القياس النفسي بقياس سمات الإنسان وقدراته، ويأتي في مقدمة ذلك الذكاء، الذي يشغل المهتمين بالكشف عن الفروق الفردية بين الطلبة. ويعتبر الذكاء واحدًا من العوامل المهمة المحددة للسلوك الإنساني والمسؤولة عن الفروق الفردية بين الأفراد، وهو ما جعله محط أنظار المهتمين من الباحثين بدراسة السلوك الانساني وأسبابه، وكذلك لكونه أحد أهم المؤشرات المستخدمة في تحديد الموهوبين أو فرز الأطفال ذوي صعوبات التعلم. وتحتل اختبارات الذكاء حيزًا كبيرًا من الاهتمام للاعتماد عليها بشكل أساسي في قياس الذكاء وتكوين تصور أولي وسريع عما يتميز به الأفراد من قدرات. ويعتبر قياس الذكاء واحداً من من التحديات التي تواجه الباحثين في مجال القياس النفسي بشكل عام، وخاصة مع تعدد النظريات التي تفسره وتطورها عبر الزمن، إلا أنه قد ظهرت على الساحة العديد من اختبارات الذكاء والتي تنوعت من حيث التوجه النظري الذي تنطلق منه، وكذلك من حيث طريقة تطبيقها (فردية، وجماعية) وكذلك وفقاً لمحتواها (لفظية، وغير لفظية).

والمتتبع لحركة التطور التي مرت بها أدوات قياس الذكاء يجد العديد من الاختبارات التي يشيع استخدامها والتي لا تزال تستخدم إلى وقتنا الحالي، فمنها على سبيل المثال لا الحصر، اختبار ستانفورد بينية Stanford-Binet ومقياس وكسلر Wechsler ، وهي من الاختبارات الفردية التي لا يزال يعتمد عليها حتى الآن في مجال تشخيص ضعاف القدرة وذوي صعوبات التعلم وكذلك فئات المتفوقين والوهوبين (متولي و القحطاني، 2016).

ومن بين الاختبارات الجمعية الأكثر شهرة أيضاً اختبار كاليفورينا للنضج العقلي California Test of Mental Maturity (CTMM) واختبار القدرات العقلية الأولية Primary Mental Abilities (PMA) واختبار أوتيس-لينون Otis-Lenon Mental Ability Test (OLMAT) واختبار المصفوفات المتتابعة لرافن Raven’s Progressive Matrices Test (RPMT). ويتضح من الأدبيات المتخصصة أن اختبارات الذكاء بنوعيها الفردية والجمعية يمكن الاعتماد عليها في التشخيص والكشف عن الفروق الفردية في القدرات، وفي الوقت نفسه يشير (Pierson, Kilmer, Rothlisberg, & Mclntosh, 2012) إلى إمكانية استخدام صورًا مختصرة من هذه الاختبارات لقياس الذكاء.

ويعتبر اختبار المصفوفات المتتابعة من اختبارات الذكاء الجمعية غير اللفظية، والذي أعده عالم النفس الإنجليزي رافن Raven في ثلاثينيات القرن الماضي، وقد لقي هذا الاختبار اهتماماً غير مسبوق بين جميع أدوات الفرز المبدئي والذي أظهرت نتائجه ارتباطًا مرتفعًا مع نتائج اختبارات الذكاء الفردية (البرصان و بخيت، 2015)، ولهذا يستخدم كمحك فعال في الدراسات التي تهتم بفرز

المجلة الدولية للابحاث التربوية- جامعة الإمارات العربية المتحدة
المجلد ( 44 ) العدد ( 3 ) أكتوبر 2020

الطلبة الموهوبين كما في(Carman, 2011; Wellisch & Brown, 2012) .وقد طور رافن ثلاثة اختبارات وهي اختبار المصفوفات المتتابعة المعياري Standard Progressive Matrices، اختبار المصفوفات المتتابعة المتقدمة Advanced Progressive Matrices، اختبار المصفوفات المتتابعة الملون Colored Progressive Matrices.

وقد انطلق رافن عند بنائه لاختبارات المصفوفات المتتابعة الثلاثة من نظرية العاملين لسبيرمان والتي تعتقد أن الذكاء يمثل القدرة العقلية العامة للفرد وفي أنه يتكون من عامل عام واحد وعامل أو أكثر خاص، ويرى سبيرمان أن الذكاء ينعكس في القدرة على استنباط العلاقات والمتعلقات والتي تظهر في القدرة على اكتشاف الروابط الحقيقية والخفية بين الأشياء (Bass, 2000). ويشير (علام، 2000) أن اختبارات الذكاء التي تبنى وفق التوجه العاملي الذي يفترضه سبيرمان تنتهي بتحديد درجة واحدة كلية للذكاء تعبر عما أطلق عليه سبيرمان العامل العام الذي يؤثر في جميع العمليات العقلية بنسب متفاوتة، مع عدم التركيز على العوامل النوعية الخاصة التي تشترك مع هذا العامل العام وتختلف بطبيعة الحال باختلاف العملية العقلية المستخدمة في الاختبار .

واعتمد رافن في مفردات اختبارات المصفوفات المتتابعة بشكل عام على قياس القدرة على إدراك العلاقات بين الوحدات المجردة (وهي العلاقات بين الأشكال الهندسية التي استخدمها في صورة مصفوفات)، وتتدرج العلاقات التي يعتمد عليها الاختبار إذ تبدأ بعلاقات التشابه والاختلاف بين الوحدات والتي تعتمد على دقة التمييز والملاحظة حتى تصل إلى العمليات العقلية العليا متمثلة في إدراك الروابط المنطقية التي تحتاج للقدرة على التفكير الاستدلالي. وهو من الاختبارات المتحررة من أثر الثقافة (Kail, 2004). ويؤكد رافن (2004) على أن اختبارات المصفوفات المتتابعة تعد الأكثر شيوعاً وتفضيلاً في قياس القدرة العقلية المجردة أو غير اللفظية، حيث تقيس الذكاء السائل غير المرتبط بأي ثقافة (في الحسن ورحمة، 2017).

ويشهد ميدان التقويم التربوي تقدماً ملموساً وذلك مع التطور التكنولوجي الذي يمثل السمة الرئيسية للعصر الحالي. فمنذ منتصف القرن الماضي، ظهرت بدايات نظرية الاستجابة للمفردة Item Response Theory لتواجه العديد من الصعوبات والتحديات التي لم تتمكن نظرية الاختبارات الكلاسيكية Classical Test Theory من علاجها. ومن أهم التطبيقات التي ارتبطت بهذا التقدم بناء بنوك الأسئلة والاختبارات التكيفية المحوسبة Computerized Adaptive Tests (CAT)والتي تعتمد بشكل أساسي على استخدام الحواسيب ليس فقط كوحدات تخزينية ولكن لتطبيق وإدارة عملية التقييم.

وتعتبر نظرية الاستجابة للمفردة من النظريات المهمة في بناء وتطوير المقاييس والاختبارات والمواكبة للتطور التكنولوجي المستمر والمتسارع من منتصف القرن الماضي، وتنطلق

هذه النظرية من فكرة أن أداء المفحوص على مفردات اختبار يقيس سمة معينة يتحدد بما يمتلكه هذا المفحوص من هذه السمة، ولذا فمن ضمن مسمياتها نظرية السمات الكامنة. وتتضمن هذه النظرية عدداً من النماذج النظرية التي تحاول أن تصف العلاقة بين مستويات السمة واحتمال الاستجابة على المفردات الاختبارية بأنواعها المختلفة ثنائية التدريج Dichotomous Items أو متعددة التدريج Polytomous Items، وقد بين هامبلتون وسوامينثان ( Hambleton & Swaminathan, 1985) أن من أهم ما تتميز به نظرية الاستجابة للمفردة عن النظرية الكلاسيكية للاختبارات هو إمكانية تقدير قدرة المفحوصين بشكل مستقل عن عينة المفردات التي تطبق عليهم، وكذلك إمكانية تقدير خصائص المفردات الاختبارية بشكل مستقل عن عينة الأفراد الذين يطبق عليهم الاختبار وهو ما أكدت عليه كاظم (1988) بأن هذه النظرية قد جاءت لتحل مشكلة الموضوعية في القياس، والتي فسرتها بعدم تأثر قياس الظاهرة موضع الاهتمام باختلاف الأداة المستخدمة في القياس.

ويوضح هامبلتون وسوامينثان (Hambleton & Swaminathan, 1985) أكثر نماذج نظرية الاستجابة للمفردة شيوعاً مع المفردات ثنائية التدريج للاختبارات أحادية البعد، وهي النماذج الأحادية والثنائية والثلاثية البارامترات، حيث تختلف هذه النماذج في عدد البارامترات التي تفترضها لتصف العلاقة بين مستوى السمة أو القدرة التي يمتلكها الفرد وخصائص المفردات. ففي النموذج أحادي البارامتر يوضح المنحنى المميز للمفردة العلاقة بين مستوى السمة أو القدرة واحتمال الاستجابة عليها بافتراض اختلاف بارامتر الصعوبة Item Difficulty (b) بين المفردات في الاختبار والذي يحدد موضع المفردة على متصل القدرة، ويضيف النموذج ثنائي البارامتر معلماً آخراً للمنحنى المميز للمفردة وهو معامل التمييز Item Discrimination (a) والذي يمثل ميل المنحنى عند النقطة المقابلة لموضع المفردة على متصل القدرة، ويعكس هذا المعلم مستوى تمييز المفردة بين المرتفعين والمنخفضين في مستوى القدرة. ويضيف النموذج الثلاثي البارامتر معلم التخمين Item Guessing (c) والذي يمثل احتمال الإجابة الصحيحة على السؤال اعتماداً على التخمين أي عند أدنى مستويات القدرة (Baker, 2001).

وتختلف الاختبارات التكيفية المحوسبة بشكل واضح عن الاختبارات الخطية Linear Testsوهي الاختبارات التقليدية والتي تعتبر الأكثر استخداماً حتى الآن، ويطلق عليها أحياناً اختبارات الورقة والقلمPaper-and-pencil Tests ، وهي تلك الاختبارات التي يجيب فيها جميع المفحوصين على جميع أسئلة الاختبار بغض النظر عن صعوبة تلك المفردات وكذلك بغض النظر عن مستوى المفحوص. كما وتختلف أيضاً الاختبارات التكيفية المحوسبة عن الاختبارات المحوسبة Computer-based Tests بشكل عام، حيث يمكن أن تستخدم الحواسيب في تطبيق الاختبارات ولكن بصورة خطية أيضاً كما هو الحال في اختبارات الورقة والقلم ( Magis, Yan, & von Davier, 2017).

المجلة الدولية للأبحاث التربوية- جامعة الإمارات العربية المتحدة
المجلد ( 44 ) العدد ( 3 ) أكتوبر 2020

ولا تعتبر فكرة الاختبارات التكيفية حديثة العهد بالاختبارات والمقاييس، إذ ترجع إلى اختبار بينيه للذكاء في بدايات القرن الماضي، وكان للتقدم التكنولوجي الفضل في ظهور مفهوم الاختبارات التكيفية المحوسبة والتي طورت الفكرة الكلاسيكية لتحقق مستوى أعلى من الدقة والموضوعية في عملية التقييم، كما أن شيوع هذا النوع من الاختبارات ارتبط كثيراً بظهور نظرية الاستجابة للمفردة (Lord, 1980)، حيث تنطلق الفكرة من تدريج المفردات المكونة للاختبارات باستخدام هذه النظرية للتعرف على مواصفاتها وخصائصها الإحصائية، والتي يعتمد عليها في بنك الأسئلة الذي يمثل قاعدة البيانات في الاختبار التكيفي المحوسب. ويقوم الحاسوب باختيار مفردة وفق قاعدة معينة وغالباً ما تكون متوسطة الصعوبة لتقدم للمفحوص في البداية ومن ثم استخدامها في تقدير مستوى القدرة الأولي للفرد، وبناء على المستوى الأولي يتم اختيار المفردة التالية التي تقدم للمفحوص والتي سيتم بعدها إعادة تقدير مستوى القدرة للفرد مرة أخرى، وتستمر هذه الخطوات حتى يتحقق محك يتم اختياره لإنهاء تطبيق الاختبار وتحديد القدرة النهائية للفرد ومستوى دقة تقديرها (Wang & Lu, 2018). لهذا فإن هذه الاختبارات تقدم للمفحوص فقط تلك المفردات التي تناسب مستواه وتستبعد بقية المفردات البعيدة عنه. ويرى كل من (Hambleton & Swaminathan, 1985) أن اعتماد فكرة عمل هذه الاختبارات على نظرية الاستجابة للمفردة يعود لسببين رئيسيين وهما تحرر تقديرات القدرة للأفراد عن خصائص عينة المفردات المستخدمة في تقدير القدرة (Item free) وكذلك تحرر الخصائص الإحصائية للمفردات عن خصائص عينة الأفراد المستخدمة في تقدير خصائص المفردات (Person free).

وبشكل عام، يتضمن تطوير الاختبارات التكيفية المحوسبة مجموعة من العناصر وهي بنك الأسئلة الذي يعتبر بمثابة مخزن أو مستودع الأسئلة الكلي والذي يتم انتقاء مفردات الاختبار التكيفي منه، ومستوى بداية الاختبار، وأساليب اختيار المفردات في الصورة التكيفية Item Selection Procedures، وأساليب تقدير القدرة Ability Estimation Methods، وكذلك القاعدة التي ينتهي بها الاختبار أو محك إيقاف الاختبار (Weiss & Kingsbury, 1984).

ويشير بويد (Boyd, 2003) إلى أن بنك الأسئلة لا يتضمن فقط مجموعة من الأسئلة تعمل كمخزن لانتقاء الأسئلة في الاختبار التكيفي، ولكن يجب أن يتم تدريجها وفق أحد نماذج نظرية الاستجابة للمفردة التي تلائم مفردات الاختبار، وفي الوقت نفسه يجب أن تتوزع مفردات البنك وفق مستوى صعوبتها على متصل القدرة، حيث تعتمد دقة عملية التقدير لقدرة المفحوص على دقة توزيع المفردات على هذا المتصل. كما يوضح المرجع نفسه وجود عدد من الطرق المختلفة التي يمكن استخدامها في انتقاء المفردات من بنك الأسئلة وهي طريقة الاختيار العشوائي Random Selection Method، وطريقة المعلومات القصوى Maximum Information Selection Method، والطريقة البيزية في انتقاء المفردات Bayesian Item Selection Method، وتعد

الطريقتين الأخيرتين الأكثر استخداماً في انتقاء المفردات للاختبارات التكيفية المحوسبة.

كما ويوضح (Stafford, Runyon, Casabianca, & Dodd, 2018) عدداً من المحكات التي يمكن استخدامها لإيقاف تطبيق الاختبار التكيفي، وهي محك الطول الثابت للاختبار Fixed Lengthحيث يقدم لكل مفحوص نفس العدد من الأسئلة الاختبارية التي يحددها الفاحص، وهي تفتقد للاتساق في دقة تقديرات القدرة للمفحوصين، حيث يتوقع اختلاف دقة تقدير المفحوص بحسب موقعة على متصل القدرة، وتتطلب محاولات رفع الدقة باستخدام هذه الطريقة زيادة عدد الأسئلة وهو ما يعني عدم جدوى تكييف الاختبار وكذلك زيادة نسبة التعرض لكل سؤال في بنك الأسئلة وهو أيضاً الأمر الذي ينبغي الانتباه له عند تكييف الاختبارات ( Leroux & Dodd, 2014)، ومحك الطول المتغير للاختبار Variable Length والذي يعتمد على الوصول لمستوى معين من دقة التقدير لقدرة المفحوص وفقاً للخطأ المعياري للتقدير، ووفقاً لهذا المحك لا يتعرض جميع المفحوصين لنفس العدد من المفردات، ولكن بالقدر الذي يضمن تحقيق مستوى الدقة المحدد. ويشير (Babcock & Weiss, 2012) إلى دقة طريقة الإيقاف وفقاً للقيمة المطلقة للتغير في القدرة- Absolute Change in Theta والتي تعتبر من طرق الإيقاف وفقاً لمحك الطول المتغير للاختبار- حيث يتوقف الاختبار التكيفي عندما يكون الفرق المطلق للتغير في قيمة القدرة المقدرة أقل من قيمة محددة.

وللاختبارات التكيفية عدة مميزات مقارنة بالاختبارات الخطية أو التقليدية (اختبارات الورقة والقلم) ترصدها الأدبيات، منها سهولة إدارة المفردات الاختبارية وإتاحة درجة الفرد والتغذية المرتدة عن الأداء بشكل فوري، وكذلك الحفاظ على مستوى دافعية المفحوص، حيث لا يتعرض لعدد كبير من المفردات، وكذلك خفض معدلات القلق لديه، بالإضافة إلى النقطة الأهم وهي دقة تقدير مستوى القدرة للمفحوص أو مستوى السمة لديه ( Linacre, 2000; Rezaie & Golshan, 2015) ولم تعد السياقات التربوية فقط هي المجال الوحيد لتطبيقات نظرية الاستجابة للمفردة والاختبارات التكيفية المحوسبة، حيث ظهرت دراسات متنوعة لتطوير هذا النوع من الاختبارات في المجالات الطبية والتي تسعى للوصول إلى أكبر قدر من المعلومات في وقت قليل وعدد أقل من المفردات كما هو الحال مثلاً في دراستي ( Liegl et al., 2019; McCabe, Gross, & Bulut, 2018).

ويمكن تطبيق فكرة الاختبارات التكيفية المحوسبة على الأنواع المختلفة من المفردات الاختبارية سواء كانت ثنائية التدريج Dichotomous Items أو متعددة التدريج Polytomous Itemsوفقاً للنماذج المناسبة لكل منهما بين نماذج نظرية الاستجابة للمفردة، كما ويمكن تطبيقها أيضاً على الاختبارات التي تقيس سمة واحدة وهي الاختبارية أحادية البعد وفقاً لنماذج الاستجابة

المجلة الدولية للابحاث التربوية- جامعة الإمارات العربية المتحدة
المجلد ( 44 ) العدد ( 3 ) أكتوبر 2020

للمفردة الخاصة بها Unidimensional IRT models أو مع الاختبارات متعددة الأبعاد بنماذجها الخاصة بها والتي تعتبر تطويراً لنماذج الاستجابة للمفردة في الاختبارات أحادية البعد Multidimensional IRT models (Magis, Yan, & von Davier, 2017).

وتهتم الدراسة الحالية بمدى إمكانية تطوير صورة تكيفية محوسبة من اختبار رافن للمصفوفات المتتابعة الملونة وذلك باستخدام نظرية الاستجابة للمفردة، ويعتبر اختبار رافن للمصفوفات المتتابعة الملونة واحداً من أهم اختبارات الذكاء التي شاع استخدامه بسبب تصميمه المبتكر والذي يلغي أهم العقبات الثقافية والتعليمية في إعداد الاختبارات بشكل عام، وهي اللغة ومحتوى الاختبار والقراءة والسرعة (Bass, 2000). ويؤكد (Lynn & Vanhanen, 2002) على أن اختبار رافن من أكثر الاختبارات التي شاع استخدامها في الدراسات عبر الثقافات المختلفة في قارات العالم الخمس. ولقد اهتمت العديد من الدراسات بتقنين اختبارات رافن في البيئة العربية ومنها دراسة ابن زرقين (2016) على البيئة الجزائرية، وكاظم وآخرون (2008) والتي اهتمت بتقنين الاختبار في البيئة العمانية ودراسة آل ثاني (2002) التي قننت الاختبار لطلاب وطالبات المرحلة الابتدائية في قطر، ودراسة عيد (1999) في الإمارات العربية المتحدة، وفي السودان كما في دراسة الخطيب ومصطفى وحسين (2006) وفي العراق كما في العاني (1989) وفي الكويت كما في القرشي (1987)، وفي اليمن عند الهيتي والغانم والزبيدي والنود (1995).

**مشكلة الدراسة**

تعكس الأدبيات محاولات لتطبيق فكر الاختبارات التكيفية المحوسبة في البيئة العربية، وتعتبر تلك المحاولات قليلة إذا ما قورنت بنظيرتها في البيئة الأجنبية، ويوضح ثومبسون و ويس (Thompson & Weiss, 2011) أن الاهتمام الأكبر لمعظم الدراسات التي تناولت الاختبارات التكيفية المحوسبة كان لعقد مقارنات سواء بينها وبين الاختبارت الخطية أو بتطبيقها على اختبارات بعينها. كما أنه وفقًا لما يشير إليه ثومبسون وويس (Thompson & Weiss, 2011) لم تقدم نتائج وتوصيات البحوث السابقة نموذج عام يمكن الاسترشاد به في بناء تلك الاختبارات، وهو ما حاول أن يقدمه بوضعه لإطار عام يفيد في تطوير أي أنظمة تقييم تعتمد على الاختبارات التكيفية المحوسبة .

وفي العالم العربي محاولات متنوعة لتطبيق الاختبارات التكيفية المحوسبة فقد هدفت دراسة عبيدات (2008) إلى فحص فاعلية الاختبار التكيفي المحوسب في تقدير القدرة العقلية باستخدام مصفوفات رافن بصورها الثلاثة (الملونة، المعيارية، المتقدمة)، وقد أشارت النتائج إلى أن قاعدة إنهاء الاختبار بعدد محدد من المفردات توفر تقديرات للقدرة أدق ودالة معلومات أعلى من قاعدة أدنى خطأ معياري، وأن كليهما يوفر تقديرات أدق من الاختبارات الخطية، بالإضافة إلى

ذلك كانت 70% من مفردات الاختبار الخطي كافية لانهاء الاختبار عند أقل مستوى للخطأ المعياري.

كما قام كل من عودة وعبيدات (2013) بدراسة فاعلية الاختبار التكيفي المحوسب باستخدام مصفوفات رافن الملونة والمعيارية والمتقدمة والتي استخدمت لتشكيل بنك للأسئلة لفحص فاعلية الاختبار التكيفي المحوسب في دقة تقدير القدرة العقلية باستخدام مصفوفات رافن باختلاف مجموعة من طرق التقدير وقواعد إنهاء الاختبار.

وهدفت دراسة البرصان وبخيت (2015) إلى الكشف عن فاعلية القياس التكيفي في الكشف عن الموهوبين باستخدام اختبار رافن للمصوفات المتتابعة المعياري، وقد أشارت نتائج الدراسة إلى فاعلية الكشف عن الموهوبين باستخدام الصورة التكيفية وإلى ارتباط مرتفع بين القدرة المقدرة من الاختبار التكيفي وتلك التي تم تقديرها باستخدام الاختبارات التقليدية الخطية.

كما هدفت دراسة البرصان (2012) إلى الكشف عن فاعلية القياس التكيفي المبني من مفردات ذات إجابة منتقاة ثنائية التدريج ومفردات ذات إجابة منشأة متعددة التدريج بالاعتماد على بيانات مولدة باستخدام برمجية متخصصة، وقد أظهرت النتائج ارتباط مرتفع بين القدرة المقدرة بالاختبار التكيفي وتلك المقدرة من الاختبار الخطي، بالإضافة إلى فاعلية القياس التكيفي المبني باستخدم نوعي المفردات ثنائية ومتعددة التدريج.

وتهتم الدراسة الحالية بمدى إمكانية تطوير صورة تكيفية محوسبة من اختبار رافن للمصفوفات المتتابعة الملونة في البيئة العمانية باستخدام نظرية الاستجابة للمفردة ووفقاً للإطار الذي اقترحة ثومبسون وويس (Thompson & Weiss, 2011)، وتستند في ذلك على مجموعة من المبررات هي:

1- لا توجد دراسات -في حدود علم الباحثين- اهتمت بتطوير اختبارت تكيفية محوسبة في البيئة العمانية بشكل عام يمكن استخدامها لقياس القدرة العقلية العامة، في الوقت الذي تزيد فيه الحاجة إلى مثل هذه الاختبارت وخاصة لأغراض التشخيص والتقييم والتي يؤخذ بناء عليها قرارات قد تكون في غاية الأهمية بالنسبة للأفراد.
2- لم تهتم الدراسات العربية -في حدود علم الباحث- بدراسة جدوى تكييف اختبار المصفوفات المتتابعة الملونة، وإن استخدم بشكل جزئي في بنك الأسئلة الذي شكله كل من عودة وعبيدات (2013) من ثلاثة اختبارات لرافن وهي المصفوفات المتتابعة الملونة والمعيارية والمتقدمة، في حين اهتمت دراسة يوليانتو (Yulianto, 2007) بالمصفوفات المتتابعة المتقدمة، واهتمت دراسة البرصان والبخيت (2015) باختبار المصفوفات المتتابعة المعياري.

المجلة الدولية للأبحاث التربوية- جامعة الإمارات العربية المتحدة
المجلد ( 44 ) العدد ( 3 ) أكتوبر 2020

3- احتياج المتخصصين في وزارة التنمية الاجتماعية من المهتمين بتشخيص ذوي الاحتياجات الخاصة إلى اختبارات ذكاء موثوقة تمكنهم من القيام بأغراض التصنيف بشكل دقيق ومن ثم تقديم الدعم المناسب لهم ولأسرهم، بالإضافة إلى الحاجة لتقليل عدد المفردات التي تستخدم لتقييم هؤلاء الطلبة، حيث يمكن أن تتكرر مرات التقييم والمتابعة لهم من وقت لآخر لمعرفة مدى التقدم والتحسن الذي يحدث لهم.

4- حاجة العاملين في وزارة التربية والتعليم لاختبارات ذكاء مقننة وموضوعية يمكن الوثوق فيها عند تصنيف فئات الطلبة وخاصة في المرحلة الابتدائية للتمييز بين الفئات المختلفة (العاديين، ذوي الصعوبات التعلم، والموهوبين).

5- كون اختبار رافن للمصفوفات المتتابعة الملونة من أهم اختبارت الذكاء على مستوى العالم والتي تنطلق من أساس نظري واضح وهو نظرية العاملين لسبيرمان، والذي يلقى اهتماماً واسعاً في الدراسات عبر الثقافات المختلفة، ما دفع العديد من الباحثين في البيئتين الأجنبية والعربية إلى تقنينه.

6- تأتي الدراسة الحالية والتي تهتم بإمكانية إعداد صورة تكيفية من اختبار رافن للمصفوفات المتتابعة الملونة مكملة للمجهود الذي قام به الفريق البحثي كاظم وآخرون (2008) والذي اهتم بتقنين الاختبار في صورته العادية (الورقة والقلم) على البيئة العمانية وكان ذلك ضمن مشروع بحثي ممول داخلياً من جامعة السلطان قابوس.

ويمكن حصر مشكلة الدراسة الحالية في الإجابة عن السؤال الرئيس التالي:

ما مدى إمكانية تطوير صورة تكيفية من اختبار رافن للمصفوفات المتتابعة الملونة على البيئة العمانية؟

ويندرج من هذا السؤال ثلاثة أسئلة فرعية وهي:

1- ما مدى تحقق الافتراضات الرئيسية لنظرية الاستجابة للمفردة في مفردات الاختبار؟
2- هل يمثل بنك أسئلة اختبار رافن للمصفوفات المتتابعة الملونة مستودعاً كافياً لاستخدامه لأغراض القياس التكيفي؟
3- ما مستوى فاعلية الصورة التكيفية المحوسبة من اختبار رافن للمصفوفات المتتابعة الملونة؟

وسوف يعتمد الباحثان في العمل الحالي على منهجية محاكاة البيانات الواقعية التي استخدمت في العديد من الدراسات والتي وضحها (Thompson & Weiss, 2011) عندما قدم إطاراً مقترحاً لبناء الاختبارات التكيفية المحوسبة، وأيضاً كما في دراسة (McCabe et al., 2018) والتي وضحت طريقة التعرف على مدى إمكانية تطوير اختبار تكيفي محوسب لتقييم الأداء البدني للمرضى وفقاً لاستجاباتهم على بعض المفردات. وتوضح البحوث السابقة أن دراسات المحاكاة

بطريقة "مونت كارلو" تمكن الباحثين والمهتمين بتطوير الاختبارات التكيفية المحوسبة من تقدير طول الاختبار ودقة تقديرات القدرة التي يتم التوصل إليها بالاختبار في صورته التكيفية، كما تساعد أيضاً في تحديد أنسب الطرق للتحكم في اختيار المفردات الاختبارية ومحكات التوقف في نهاية الاختبار، بالإضافة إلى معرفة الحجم المطلوب لبنك الأسئلة ومعدل التعرض للمفردات Item exposure الأمر الذي قد يكون بالغ الأهمية عندما تكون سرية المحتوى واحداً من الأمور المهمة كما هو الحال في اختبارات القبول وغيرها.

وتمر مراحل التعرف على مدى ملاءمة استخدام الاختبار التكيفي المحوسب بمجموعة من المراحل يعتبر أهمها ما تم توضيحه بالتفصيل في الإطار المقترح الذي قدمه ( Thompson & Weiss, 2011) وهي:

1- دراسة الجدوى والتخطيط وقابلية التطبيق.
2- تطوير محتوى بنك الأسئلة.
3- تدريج أو معايرة مفردات بنك الأسئلة قبلياً.
4- تحديد مواصفات الاختبار التكيفي المحوسب النهائي.
5- النشر الفعلي للاختبارات التكيفية المحوسبة .

ونظراً لأن بنك الأسئلة الذي تعتمد عليه الدراسة الحالية هو اختبار رافن للمصفوفات المتتابعة الملونة وهو ما يتم تناوله في المرحلة الثانية في البرامج التي ستبدأ في إعداد وبناء أو تطوير محتوى بنك الأسئلة، وأيضاً نظراً للحاجة الماسة لدراسة جدوى مثل هذه الاختبارات كما هو موضح في مبررات الدراسة الحالية والتي تم عرضها سلفاً وهو الأمر الذي يتم تدارسه في المرحلة الأولى، فإن الدراسة سوف تغطي المرحلتين الثالثة والرابعة من بين المراحل الخمسة السابقة، والتي يمكن اعتبارها الأساسية للحكم على إمكانية إعداد اختبارات تكيفية محوسبة من اختبار رافن للمصفوفات المتتابعة الملونة والتي يمكن من خلالها الحكم على صلاحية نشر هذا الاختبارات (المرحلة الخامسة) ضمن منصات الاختبارات التكيفية سواء التي تعمل على شبكة الانترنت أو تلك التي يتم برمجتها وفقاً لاحتياجات المؤسسات التي تستخدمها، ولعل منصة FastTest واحدة من أشهرها والتي تنتجها مؤسسة Assessment ويتابعها ويطورها عدد كبير من الخبراء والمتخصصين في مجال القياس والتقويم والتحليلات المتقدمة ذات الصلة بنظرية الاستجابة للمفردة.

**أهداف الدراسة**

تهدف الدراسة الحالية بشكل أساسي إلى دراسة إمكانية إعداد صورة تكيفية محوسبة من اختبار رافن للمصفوفات المتتابعة الملونة، وفقاً لإطار العمل الذي تناولته دراسة ( Thompson

المجلة الدولية للأبحاث التربوية- جامعة الإمارات العربية المتحدة
المجلد ( 44 ) العدد ( 3 ) أكتوبر 2020

(& Weiss, 2011) ودراسة (McCabe et al., 2018)، وفي ضوء هذا الهدف العام تسعى الدراسة إلى مايلي:

1- التعرف على إمكانية تدريج مفردات اختبار المصفوفات المتتابعة الملونة باستخدام نظرية الاستجابة للمفردة.
2- التحقق من ملاءمة ومناسبة بنك أسئلة الاختبار لإجراءات ومتطلبات الاختبارات التكيفية المحوسبة.
3- تقييم أداء الاختبار التكيفي المحوسب في ضوء عملية المحاكاة التي تعتمد على البيانات الواقعية الحقيقية.

**أهمية الدراسة**

تبرز أهمية الدراسة الحالية كونها الدراسة الأولى –في حدود علم الباحثين- والتي تهتم بدراسة جدوى تطوير اختبار تكيفي محوسب في سلطنة عمان بشكل عام، وبالتحديد لقياس القدرة العقلية العامة، وهو ما يتماشى مع احتياجات وزارتي التربية والتعليم والتنمية الاجتماعية والتي تعتمد على الاختبارات بشكل واضح في العديد من الإجراءات المهمة والضرورية سواء في الجوانب التشخيصية والتصنيفية أو في التعرف المبكر على القدرات والمواهب أو الصعوبات والتي يمكن من خلالها تطوير ممارسات تتناسب واحتياجات الأفراد.

كما ويعتبر البحث الحالي ضرورة في ضوء التطور المستمر الذي تشهده أنظمة التعليم في سلطنة عمان وبخاصة نظام التعليم الأساسي، وكذلك أنظمة التقييم، والذي يتوقع أن يتربط به العديد من التحديات التي تسعى سلطنة عمان بشكل مستمر أن تواجهها بالاعتماد على البحث العلمي والذي يشجعه السلطان قابوس بن سعيد راعي النهضة الحديثة في سلطنة عمان.

كما تبرز أهمية الدراسة أيضاً في كونها تمثل تطبيقاً عملياً لدراسة جدوى استخدام الاختبارت التكيفية المحوسبة مع اختبار مثل اختبار رافن، وهو ما قد يفتح الباب أمام دراسات مماثلة على اختبارات أو مقاييس أخرى، أو حتى لتنتقل الفكرة إلى أنظمة التقييم الأكثر اتساعاً في السلطنة كما هو الحال على سبيل المثل في اختبارات الثانوية العامة.

**مصطلحات الدراسة**

**الذكاء.** للذكاء تعريفات متعددة، وسوف يتم عرض التعريف الذي يتناسب مع التوجه النظري الذي يعتمد عليه اختبار رافن للمصفوفات المتتابعة وهو الذكاء وفق نظرية العاملين لسبيرمان، حيث يعرف بالقدرة على استنباط العلاقات والمتعلقات والتي تظهر في اكتشاف الروابط الحقيقية والخفية للأشياء، سواء أكانت مدركة أو كانت مجردة (مخائيل، 2005).

**ويعرف إجرائياً** بالدرجة التي يحصل عليها المفحوص على اختبار رافن للمصفوفات المتتابعة الملونة CPM والتي تعرف بمجموع درجات إجابات المفحوص الصحيحة على أسئلة الاختبار والتي يبلغ عددها 36 سؤال، حيث تدل الدرجة المرتفعة على الذكاء المرتفع في حين تشير الدرجة المنخفضة إلى الذكاء المنخفض.

**اختبار المصفوفات المتتابعة الملونة لرافن.** وهو اختبار صمم لقياس الذكاء بطريقة غير لفظية، وهو من الاختبارات المتحررة من تأثير الثقافة، ويستخدم مع الأطفال في الفئة العمرية (5-11)، ويتكون من عدد 36 مفردة (كاظم وآخرون، 2008).

**الاختبار التكيفي المحوسب.** ويسمى الاختبار المفصل tailored حيث يعتمد على تقديم مفردة اختبارية للمفحوص تستخدم في تقدير مستواه بشكل أولى ثم استخدام هذا التقدير في اختيار المفردة التالية التي تتناسب ومستواه، ويستمر ذلك حتى يتوقف الاختبار بتحديد تقدير لقدرة أو مستوى سمة المفحوص (Mizumoto, Sasao, & Webb, 2017).

## حدود الدراسة

تقتصر الدراسة الحالية على دراسة إمكانية تكييف اختبار رافن للمصفوفات المتتابعة الملونة في ضوء بيانات عينة أرشيفية استخدمت لتقنين الاختبار عام 2008، وهي عينة ممثلة للأطفال العمانيين من الجنسين في الفئة العمرية 5-11 سنة، وهي بيانات مشروع بحثي ممول ضمن المنح الداخلية لجامعة السلطان قابوس (IG/EDU/PSYC/05/01). ووفقاً لذلك فإن صلاحية الدراسة الحالية وإمكانية تعميم نتائجها ترتبط بالحدود السابقة.

## الدراسات السابقة ذات الصلة

يظهر في الأدبيات عدداً من الدراسات التي تناولت اختبار المصفوفات المتتابعة الملونة، وتقنينها في العديد من الدول العربية والأجنبية، وفي هذه الجزء سيتم عرض أهم الدراسات التي تناولته في البيئتين ثم ينتهي العرض بتلك الدراسات التي اهتمت بالاختبارات التكيفية بشكل عام وبتلك التي اهتمت بتكييف اختبارات رافن على وجه الخصوص.

فقد قام كوستينبادر و نجاري (Costenbader & Ngari, 2001) بتقنين اختبار المصفوفات المتتابعة الملونة على عينة من كينيا بلغ عددها 1370 من الأطفال في المرحلة العمرية من 6-10 سنوات ومن الجنسين، وهي عينة ممثلة للمجتمع الكيني، تم التحقق من الثبات بطريقتين وهما إعادة الاختبار على عينة بلغ عددها 50 طالباً وطالبة بفاصل زمني أسبوعين بين التطبيقين،

المجلة الدولية للابحاث التربوية- جامعة الإمارات العربية المتحدة
المجلد ( 44 ) العدد ( 3 ) أكتوبر 2020

ومعامل ثبات ألفا، وكان معامل ثبات القسم "أ" بطريقة إعادة الاختبار 0.55، والقسم "أ ب" بلغ ثباته 0.74، والقسم الثالث "ب" كان ثباته 0.65، كما كان ثبات الاختبار ككل مرتفعاً وقيمته 0.84، في حين بلغت قيمة معامل ثبات ألفا لاختبار المصفوفات المتتابعة الملونة 0.87، وهي بشكل عام معاملات ثبات مرتفعة.

ودراسة لينسترووم (Linstrom, 2002) التي قننت اختبار المصفوفات المتتابعة الملونة على عينة من جنوب إفريقيا حجمها 2400 من الأطفال في المرحلة العمرية 6-12 سنة. وتم استخراج المعايير لكل فئة من الفئات العمرية المستهدفة، والتي أكدت على أن معايير الاختبار للأطفال في البيئات الحضرية لا تختلف عن نظيرتها في العينة الريفية. كما كانت المعايير المستخرجة من الأطفال ممن يعمل أولياء أمورهم في أعمال مهنية أو إدراية أو فنية مرتفعة عند مقارنتها بالمعايير المستخرجة من الأطفال لأولياء أمور بدون عمل أو من العمال العاديين.

وفي دراسة (Bandeira, Alves, Giacomel & Lorenzatto, 2004) والتي استهدفت استخراج المعايير المئينية لعينة برازيلية من الجنسين حجمها 779 (386 من الذكور، و393 من الإناث) والمقارنة بين مدارس مدينتي بورتو اليجري وساوباولو، ودراسة العلاقة بين الدرجات على اختبار المصفوفات المتتابعة الملونة لرافن والعمر الزمني للمفحوصين. وأوضحت نتائج هذه الدراسة وجود فروق بين المتوسطات في المدينتين، كما وجدت زيادة مطردة في متوسطات الدرجات مع الزيادة في مستوى العمر الزمني للمفحوصين .

ودراسة (Uno, Shinya, Haruhara, & Kaneko, 2005) والتي درست اختبار المصفوفات المتتابعة الملونة لرافن على عينة من الأطفال اليابانيين حجمها 644، من الصف الثاني إلى السادس الابتدائي، ممن لديهم إعاقات سمعية من مستويات متفاوتة، وأظهرت النتائج مستويات ثبات جيدة للاختبار ومعاملات ارتباط مرتفعة مع اختبار ويكسلر للأطفال والذي استخدم كمحك للتحقق من صدق درجات اختبار رافن، وقد أكدت نتائج الدراسة على صلاحية استخدام اختبار المصفوفات المتتابعة الملونة لرافن مع عينة الأطفال من ذوي الإعاقات السمعية.

كما قام كوتون وآخرون (Cotton, et al., 2005) بدراسة لتقنين اختبار المصفوفات المتتابعة الملونة لرافن على عينة من الأطفال الاستراليين بلغ حجمها 618 من الجنسين يتراوح أعمارهم بين 6-11 عاماً. كما استهدفت المقارنة بين معايير الاختبار المشتقة من هذه العينة وعينة أخرى أرشيفية من عام 1975. وقد تراوحت معاملات ثبات ألفا كرونباخ بين 0.76 و 0.88 ولطريقة التجزئة النصفية بين 0.81 و 0.90 وهي معاملات ثبات مرتفعة. كما وأظهرت النتائج فروقاً ذات دلالة إحصائية بين الأطفال من الجنسين وفق فئات العمر المختلفة ماعدا في فئة 6 سنوات، وكانت الفروق لصالح مجموعة الذكور والذين كان لهم المتوسط الأعلى. ومن النتائج المهمة لهذه الدراسة

استقرار معايير الاختبار المستخرجة من عينة الدراسة عند مقارنتها بالمعايير المشتقة من العينة الأرشيفية والتي سبقتها بحاولي 30 عاماً.

وفي عام 2006، قام الخطيب وآخرون بتقنين اختبار المصفوفات المتتابعة الملونة في السودان على عينة بلغ عدد أفرادها 1683 من طلبة الحلقة الأولى بمرحلة التعليم الأساسي تراوحت أعمارهم من 6-9 سنوات. وقد أظهرت نتائج معاملات الثبات درجة جيدة من ثبات الاختبار على العينة السودانية تراوحت بين 0.72 و 0.91 باستخدام معامل ثبات ألفا، وبين 0.63 و 0.83 باستخدام معادلة سبيرمان براون، وبين 0.63 و 0.81 باستخدام معادلة جتمان. وقد ظهرت معاملات الارتباط بين درجات عينة الدراسة في أقسام الاختبار الثلاثة دالة عند مستوى 0.01، كما وجدت علاقة ارتباطية دالة إحصائياً بين الدرجات على الاختبار والعمر الزمني للمفحوصين في عينة التقنين، وانتهت الدراسة بوضع معايير لتفسير الدرجات في فئات العمر المستهدفة.

كما قام حماد (2012) بدراسة استهدفت تقنين اختبار المصفوفات المتتابعة الملونة على عينة من البيئة الفلسطينية وذلك لاشتقاق معايير الاختبار للأطفال في فلسطين، مع مقارنة المعايير مع نظيرتها في بعض الدول العربية والأجنبية. وقد اختيرت لذلك عينة عشوائية طبقية حجمها 1258 من الأطفال في الفئة العمرية (5.6-11.6) عاماً، موزعين على المحافظات التعليمية لقطاع غزة، رفح، خانيونس، الوسطى، غزة، شمال غزة بواقع (654 من الذكور، و604 من الإناث)، وقد أظهرت النتائج أن معاملات الثبات على مستوى الدرجة الكلية لأقسام الاختبار تتراوح بين 0.59 و 0.82 وعلى مستوى الدرجة الكلية للاختبار كانت تتراوح بين 0.79 و 0.90 وهي معاملات ثبات جيدة للاختبار بشكل عام. كما أظهرت النتائج دلالات على صدق التكوين الفرضي للاختبار باستخدام محكات مثل اختبار رسم الرجل لجودانف وكذلك مستوى التحصيل لعينة الدراسة. كما تشابهت المعايير المشتقة من عينة الدراسة مع معايير دول أخرى عربية مثل السعودية والكويت وأجنبية مثل البرازيل وبريطانيا، مع وجود بعض الاختلافات في عدد من الفئات العمرية عبر الدول.

وتؤكد الدراسات السابق ذكرها على انتشار وشيوع اختبار المصفوفات المتتابعة الملونة لرافن في كل من البيئتين العربية والأجنبية، والذي ظهرت أدلة متسقة على مستويات ثبات مرتفعة له في كل البيئات التي تناولتها تلك الدراسات، وعلى فئات العمر المختلفة، كما وأشارت بعض الدراسات التي قارنت بين المعايير المشتقة من بلدان مختلفة مثل دراسة حماد 2012 إلى وجود درجة عالية من التشابه بين المعايير للاختبار في تلك الدول مع وجود بعض الاختلافات في بعض الفئات العمرية، بالإضافة إلى استخدام الاختبار مع عينات من الأسوياء وغيرهم من ذوي الإعاقات، وخاصة الإعاقات السمعية.

ومن الدراسات العربية التي تناولت فكرة الاختبارات التكيفية المحوسبة بشكل عام دراسة

المجلة الدولية للابحاث التربوية- جامعة الإمارات العربية المتحدة
المجلد ( 44 ) العدد ( 3 ) أكتوبر 2020

محمد وكاظم (2015) والتي اهتمت بالمقارنة بين تقديرات قدرة الأفراد عند تطبيق مقياس الاستدلال اللفظي في صورتيه التكيفية المحوسبة والخطية على عينة من طلبة المدارس المصرية، وقد أظهرت نتائج هذه الدراسة تعادلاً بين تقديرات القدرة للمحفوصين باستخدام نوعي التطبيق (الخطي/ التكيفي المحوسب).

ودراسة ربابعة (2012) التي اهتمت باستقصاء أثر عدد المفردات المرشحة للتقديم من بنك الأسئلة في فاعلية أسلوب كنغزباري-زارا لضبط تقديم المفردات متعددة الخطوات في الاختبارات التكيفية المحوسبة، واستخدم في الدراسة بيانات طلبة في اختبارين للتفكير اللفظي واسعي النطاق، وقد تم إنتاج مجموعتين من بيانات المحكاة في ضوء قيم البارامترات الحقيقية، وقد أشارت النتائج إلى تأثير طول الاختبار الثابت على فاعلية أسلوب كنغزباري-زارا، حيث تبين أن معدل ضبط تقديم المفردة كان لصالح 20 سؤال، وفي حالة نقصان عدد الأسئلة من 20 إلى 16 تزيد قيمة الانحراف المعياري، وخلصت الدراسة إلى أنه كلما زاد عدد أسئلة الاختبار الثابت زاد ضبط تعرض المفردة وزادت كذلك دقة القياس.

وكذلك دراسة مسلم (2011) والتي اقترحت أسلوباً لاختيار المفردات وتقدير القدرة في الاختبارات التكيفية المحوسبة. واعتمدت الدراسة على توليد البيانات وفق شروط محددة لاختبار فاعلية الأسلوب المقترح، وقد أظهرت النتائج ارتباطات مرتفعة ودالة إحصائياً بين قيم تقديرات القدرة بالطريقة المقترحة والطرق الأخرى وهي طريقة الأرجحية العظمى والطرق البيزية. وفيما يخص مؤشرات دقة القياس فقد جاءت متقاربة للطريقة المقترحة مع بقية الطرق عند مستويات القدرة المختلفة. ووفقاً للكفاءة، أظهرت النتائج تقارب بين متوسط عدد الأسئلة اللازمة للوصول لخطأ معياري (0.25) في تقدير القدرة، وذلك وفقاً لطرق التقدير المختلفة وبالتحديد عند أطراف متصل القدرة .

وفي تركيا تناولت دراسة أوزورت و أوزورت (Özyurt & Özyurt, 2015) تطوير نظام تكيفي محوسب لاستخدامة في نظام التعليم الثانوي التركي واختبار ثبات هذا النظام، ويغطي هذا النظام خمسة موضوعات في الرياضيات للصف الحادي عشر. وقد تم تطبيق النظام على عينة صغير بلغ حجمها 48 من طلبة الصف الحادي عشر في ترابزون بتركيا. وأظهرت نتائج التطبيق مستويات مرتفعة من الثبات تتراوح بين (0.88-0.93) في الموضوعات الخمسة التي يتضمنها النظام. وقد أكدت النتائج على إمكانية وفعالية استخدام نظم الاختبارات التكيفية في المناهج التركية.

ويلاحظ من الدراسات السابقة تركيزها على فاعلية الاختبارات التكيفية المحوسبة مقارنة بالخطية كما في محمد وكاظم (2015) أو تقديم مقترحات لطرق اختيار للمفردات و تقدير القدرة

أكثر من تركيزها على تطوير اختبارات تكيفية محددة مثل دراسة مسلم (2011)، أو تطوير أنظمة تقييم مدرسي تكيفية الأمر الذي حاولت دراسة أوزورت و أوزورت (Özyurt & Özyurt, 2015) التركيز عليه، وتباينت تلك الدراسات بين الاعتماد على بيانات حقيقية فقط كما في دراسة محمد وكاظم (2015)، أو بيانات مولدة حاسوبياً في ضوء بعض الشروط مثل دراسة مسلم (2011) وربابعة (2012).

ومن الدراسات التي سعت إلى تكييف اختبار المصفوفات المتتابعة لرافن دراسة البرصان وبخيت (2015) والتي هدفت إلى الكشف عن فاعلية القياس التكيفي في الكشف عن الموهوبين باستخدام اختبار رافن للمصفوفات المتتابعة المعياري، وذلك على عينة حجمها 404 من أربعة مدارس بشمال الرياض في المملكة العربية السعودية تتضمن طلبة من العاديين والموهوبين (المصنفين وفق كشوف الموهوبين المعتمدة في وزارة التربية والتعليم بالسعودية) بواقع 304 طالب من العاديين و 100 من الطلبة الموهوبين، في الصفوف من الرابع إلى السادس ومستويات عمر زمني من 9 إلى 11 عاماً. وقد أظهرت النتائج معاملات ارتباط مرتفعة بين تقديرات القدرة وفقاً للاختبار الخطي والاختبار التكيفي بلغت 0.91 وبمتوسط فرق في القدرة مقداره 0.55 بوحدة اللوجيت، كما أظهرت النتائج أيضاً تقارب نسبة الطلبة المصنفين كموهوبين باستخدام الاختبار الخطي ونظيره التكيفي، وهو الأمر الذي يدعم فاعلية تكييف اختبار رافن المعياري في فرز أو تصنيف الطلبة الموهوبين وذلك عند مقارنته بالاختبار الخطي.

كما وأجرى عودة وعبيدات (2013) دراسة عن فاعلية الاختبار التكيفي المحوسب في تقدير القدرة العقلية باختلاف طريقة تقدير القدرة (طريقة الترجيح الأقصى، وطريقة التقدير البعدي الأقصى) وقواعد إنهاء الاختبار(عدد الأسئلة المحدد، والخطأ المعياري الأدنى) باستخدام مصفوفات رافن (الملونة والمعيارية والمتقدمة لتشكل جميعها بنك الأسئلة للاختبار التكيفي) على عينتين بلغ حجم الأولى 2695 من طلبة كلية التربية جامعة اليرموك، وذلك لأغراض تدريج بنك الأسئلة المستخدم، كما بلغ حجم العينة الثانية 638 من طلبة الكلية لأغراض تطبيق الاختبارات التكيفية المحوسبة. وبشكل عام أظهرت النتائج زيادة دقة تقديرات القدرة من خلال الاختبارات التكيفية مقارنة بنظيرتها الخطية مع اختلاف طريقتي تقدير القدرة المستخدمة، وبزيادة في مقدار المعلومات لطريقة الترجيح الأقصى مقارنة بطريقة التقدير البعدي الأقصى، كما كانت نتائج التقدير وفقاً لمحك انهاء الاختبار بعدد محدد من الأسئلة أعلى دقة مقارنة بمحك الخطأ المعياي الأدنى، وقدمت وأوصى الباحثون بضرورة تفعيل واستخدام الاختبارات التكيفية المحوسبة في تطبيقات مختلفة ومتنوعة لما أظهرته من فاعلية.

كما أجرى يوليانتو (Yulianto, 2007) دراسة على عينة حجمها 298 من طلبة الجامعة

في كلية علم النفس بالجامعة الاندونيسية، بهدف دراسة فاعلية تكييف اختبار المصفوفات المتتابعة المتقدمة لرافن ومقارنتها بالاختبار الخطي وصورة محوسبة عادية من نفس الاختبار. وفي الصورة التكيفية من الاختبار تم الاعتماد على محك بداية الاختبار كما وضحه كل من امبريتسون وريس (Embretson & Reise, 2000) بالاختيار لسؤال عشوائي ذو مستوى صعوبة في الفترة (-0.5 ، +0.5)، ومحك انتقاء مفردات يعتمد على المعلومات القصوى ومحك لإيقاف الاختبار وفقاً للخطأ المعياري أقل أو يساوي (0.4). أظهرت النتائج عدم وجود فروق ذات دلالة إحصائية بين الدرجات على التطبيقين (بالورقة والقلم) والتطبيق التكيفي للاختبار، بينما وجدت فروق في الأداء بين التطبيقين المحوسب العادي والورقة والقلم لصالح الأول، وأكدت نتائج هذه الدراسة على فاعلية استخدام الصورة التكيفية من اختبار المصفوفات المتتابعة المتقدم لرافن لاختصاره في وقت التطبيق وتحقيقه درجة جيدة من الدقة بتطبيق متوسط 12 سؤال من إجمالي عدد الأسئلة الكلية والبالغ عدده 36 سؤال، وذلك عند مقارنته بكل من الصورتين المحوسبة العادية ونتائج اختبار الورقة والقلم.

ومن العرض السابق يوجد عدد قليل من الدراسات في البيئة العربية –في حدود علم الباحثين- يهتم بتكييف اختبار رافن بصوره الثلاث، ففي دراسة العودة وعبيدات (2013) كون بنكاً للأسئلة من الاختبارات الثلاثة للمصفوفات المتتابعة، في حين هدفت دراسة البرصان وبخيت (2015) إلى تكييف اختبار رافن للمصفوفات المتتابعة المعياري، وبشكل عام فهي محاولات محدودة، ويلاحظ من العرض السابق للأدبيات ما يلي:

1- لم يستخدم اختبار المصفوفات المتتابعة الملونة في تلك الدراسات إلا كجزء من بنك الأسئلة الذي شكله عودة وعبيدات (2013) من ثلاث اختبارات وهي المصفوفات المتتابعة الملونة والمعيارية والمتقدمة، وهو مالا يعتبر دليلاً كافياً على إمكانية تكييف النسخة الملونة من المصفوفات المتتابعة لرافن وحدها.
2- تعتمد الدراسات السابقة على بيانات حقيقية بتطبيق الاختبارات ودراسة فعالية النسخ التكيفية في ضوء تلك البيانات، وهو ما يحول دون دراسة العديد من الظروف التطبيقية للاختبارات التكيفية مثل طول الاختبار وطرق إيقاف التطبيق وغيرها في نفس الوقت .
3- لا توجد دراسة عربية –في حدود علم الباحثين- تعتمد على منهجية المحاكاة في دراسة إمكانية تكييف الاختبارات وفقاً للمنهجية التي ذكرها كل من ثومبسون و ويس ( Thompson & Weiss, 2011)، والتي تتيح دراسة ظروف تطبيقية متنوعة للاخبتارات التكيفية بخلاف الدراسات الامبريقية.
4- لا توجد -في حدود علم الباحثين- محاولات لتكييف اختبارات أو مقاييس نفسية أو عقلية في البيئة العمانية.

وتختلف الدراسة الحالية في أنها تستهتدف إمكانية تكييف اختبار المصفوفات المتتابعة الملونة على وجه الخصوص، بالاعتماد على منهجية المحاكاة للبيانات في ضوء تقديرات حقيقية لبارامترات المفردات، ووفقاً للطريقة التي يصفها ثومبسون و ويس ( Thompson & Weiss, 2011)، حيث يمكن في ضوء ذلك دراسة عدداً من الظروف التطبيقية للاختبارات التكيفية وتأثيرها على فعالية ودقة نتائج الصورة التكيفية من الاختبار.

## الطريقة والإجراءات

### منهجية الدراسة

اتبعت الدراسة الحالية المنهج الوصفي، إذ تسعى لوصف الخصائص الإحصائية لاختبار رافن للمصفوفات المتتابعة الملونة باستخدام نظرية الاستجابة للمفردة، وذلك لاستخدام خصائص هذه المفردات والأفراد الواقعيين في محاكاة البيانات الحقيقية للاختبار وتوليد بيانات لها نفس الخصائص للتعرف على مدى إمكانية اعداد اختبار تكيفي محوسب من اختبار رافن للمصفوفات المتتابعة الملونة وفق مجموعة من البرمجيات المخصصة لذلك، وهي تعتبر دراسة تحليل ثانوي للبيانات المستخدمة في عينة تقنين اختبار المصفوفات المتتابعة الملونة في دراسة كاظم وآخرون (2008).

### عينة الدراسة

تعتمد الدراسة الحالية على عينة تقنين اختبار رافن للمصفوفات المتتابعة الملونة في البيئة العمانية في دراسة كاظم وآخرون (2008) الممولة ضمن المنح الداخلية لجامعة السلطان قابوس، والتي بلغت 1042 طفلاً وطفلة من 70 مدرسة في جميع مناطق السلطنة، وموزعين على 11 محافظة وعلى الفئات العمرية (5 – 11) بواقع (136، 126، 137، 142، 154، 170، 177) على الترتيب، والتي تم اختيارها عشوائياً لتمثل الأطفال العمانيين من الجنسين بواقع (534 من الذكور، 508 من الإناث) في الفئة العمرية المستهدفة.

### أداة الدراسة

لتحقيق الهدف العام من البحث الحالي تم استخدام اختبار رافن للمصفوفات المتتابعة الملونةRaven’s Colored Progressive Matrices ، وهو اختبار يتكون من ست وثلاثين مفردة مقسمة إلى ثلاث مجموعات مرتبة بحيث تتشابه المجموعتان الأولى والثالثة مع نظيراتها في اختبار رافن العادي Standard Progressive Matrices بينما تعتبر المجموعة الوسطى ذات مستوى متوسط من الصعوبة بين الأولى والثالثة. ويقيس الاختبار تفصيلياً العمليات العقلية للأطفال بين عمر 5 – 11 سنة ومع كبار السن، ويتضمن الاختبار مصفوفات ملونة يمكن استخدامها مع الأطفال العاديين والمتخلفين عقلياً. وتتكون كل مصفوفة من شكل أو نمط ناقص

المجلة الدولية للابحاث التربوية- جامعة الإمارات العربية المتحدة
المجلد ( 44 ) العدد ( 3 ) أكتوبر 2020

يقوم المفحوص باستكماله بالاختيار من بين ستة من البدائل توجد أسفل كل مصفوفة بحيث يكمل البديل الصحيح النمط بشكل سليم. وتستخدم هذه المصفوفات كاختبار قوة وليس سرعة (بمعنى أن المفحوص لن يقف عقبة بينه وبين الاستجابة على المفردة سوى قدرته على التفكير دون التأثر بعوامل السرعة أو التقيد بزمن للاجابة) ولذلك فان هذا الاختبار يقيس القدرة على التفكير المنظم للفرد.

**ثبات الاختبار.**

تحقق كاظم وآخرون (2008) من ثبات درجات الاختبار على البيئة العمانية بعدة طرق، وهي ثبات إعادة تطبيق الاختبار على عينة (52) طفلاً وطفلة تتراوح أعمارهم بين 5-11 عاماً، وبفاصل زمني من 14 إلى 20 يوماً، وكانت قيمة معامل الارتباط بين الدرجات في التطبيقين مساوية لـ 0.56، وباستخدام معامل ثبات ألفا كرونباخ تراوحت قيم معاملات الثبات للدرجات في فئات العمر المختلفة بين 0.81 و 0.90 وهي تشير إلى قيم ثبات مرتفعة للاختبار في جميع الفئات العمرية. كما استخدم معامل ثبات التجزئة النصفية، وقد تراوحت معاملات سبيرمان-براون بين 0.71 و 0.86 وهي كذلك تؤكد على مستوى ثبات جيد لاختبار رافن للمصغوفات المتتابعة الملونة في البيئة العمانية .

**صدق الاختبار.**

تحقق كاظم وآخرون (2008) من صدق الاختبار وفقاً للمحك (تلازمياً) حيث تم استخدام اختبار أوتيس-لينون على عينة من (55) طفلاً وطفلة كمحك للحكم على صدق الاختبار، كما تم جمع بيانات عن مستوى تحصيل الأطفال في مقررات الرياضيات واللغة العربية واللغة الانجليزية والعلوم، مع تطبيق استمارة تقدير المعلمات اللاتي يقمن بتدريس هؤلاء الأطفال لمدة لا تقل عن ستة أشهر، وقد بلغت قيمة معامل الارتباط بين درجات الاختبار والتحصيل في المقررات الأربعة بشكل عام 0.41، وهي قيمة دالة إحصائياً عند مستوى أقل من 0.01. وكانت قيمة معامل الارتباط بين درجات اختبار رافن والدرجات في اختبار أوتيس –لينون 0.64 وهي قيمة دالة إحصائياً عند مستوى أقل من 0.01، وكان معامل الارتباط مع تقديرات المعلمات مساوياً لـ 0.41، وهي قيمة دالة إحصائياً عند مستوى أقل من 0.01. وتعتبر نتائج هذه المؤشرات مجتمعة دليلاً على صدق درجات اختبار المصفوفات المتتابعة الملونة على العينة العمانية. كما قدمت دراسة كاظم وآخرون أيضاً أدلة على صدق بناء الاختبار، كتمايز مستوى أداء الطلبة في اختبار رافن باختلاف مستويات تحصيلهم، وكذلك باختلاف فئات العمر.

**المعالجة الإحصائية**

تمر الدراسة الحالية بمجموعة مختلفة من التحليلات الإحصائية من أجل الإجابة على الأسئلة البحثية وهي:

1- اختبار أحادية البعد باستخدام برنامج Winsteps والذي يوفر إمكانية الكشف عن نسبة

التباين الكلي المفسر بالمفردات، وحجم التباين غير المفسر بالعامل الأول، بالإضافة الى التحليل العاملي التوكيدي وذلك للتعرف على مدى تحقق أحادية البعد كونها أحد الافتراضات الأساسية التي تقوم عليها نظرية الاستجابة للمفردة.

2- التحقق من مدى ملاءمة الأفراد والمفردات لأحد نماذج الاستجابة للمفردة باستخدام مؤشر مربع كاي.
3- تدريج مفردات بنك الأسئلة باستخدام نموذج الاستجابة للمفردة.
4- دراسة خصائص الاختبار التكيفي باستخدام المحاكاة المعتمدة على بيانات واقعية.

وسوف يتم الاعتماد على مجموعة من البرمجيات لعمل المعالجات الإحصائية المطلوبة وهي برنامج SPSS-22 وحزمة برمجيات نظرية الاستجابة للمفردة في منصة R وهي (mirt) وكذلك حزمة البرمجيات الرئيسية لاختبار خصائص الاختبارات التكيفية في منصة R وهي (catR).

## نتائج الدراسة ومناقشتها

فيما يلي يتم عرض نتائج أسئلة الدراسة مع تقديم التفسيرات الخاصة بها ومناقشتها، وذلك على النحو التالي:

### الإجابة على السؤال الأول

ما مدى تحقق الافتراضات الرئيسية لنظرية الاستجابة للمفردة في مفردات الاختبار؟

وللإجابة على هذا السؤال تم التحقق من الافتراضات الرئيسية لنظرية الاستجابة للمفردة وهي: أحادية البعد، والاستقلال الموضعي، والتحرر من السرعة، وملاءمة الأفراد والمفردات لنماذج الاستجابة للمفردة، وفيما يلي بيان تفصيلي لكل محور

### أ- أحادية البعد.

تم التحقق من هذا الافتراض كأحد الافتراضات المهمة لنظرية الاستجابة للمفردة باستخدام برنامج Winsteps والذي يوفر إمكانية التحقق من أحادية البعد في مفردات الاختبار، ولقد أشارت النتائج إلى ما يلي:

جدول 1

*نتائج الكشف عن أحادية البعد لاختبار المصفوفات المتتابعة الملونة*

| نسبة التباين الكلي المفسر من البيانات | حجم التباين المفسر بالعامل الأول المستخلص من مصفوفة البواقي | نسبة التباين المفسر بالعامل الأول المستخلص من مصفوفة البواقي |
|---|---|---|
| %91 | 2,5 | %0.6 |

المجلة الدولية للأبحاث التربوية- جامعة الإمارات العربية المتحدة
المجلد ( 44 ) العدد ( 3 ) أكتوبر 2020

ويتضح من النتائج الموضحة بجدول 1 تحقق فرضية أحادية البعد في اختبار رافن للمصفوفات المتتابعة الملونة بقدر يسمح باستخدام نماذج نظرية الاستجابة للمفردة وذلك وفقاً لما يراه(Linacre, 2006, p. 272) ، حيث تزيد نسبة التباين الكلي المفسر من البيانات عن (20%) وينخفض كل من حجم التباين المفسر بالعامل الأول المستخلص من مصفوفة البواقي عن (3) ونسبة هذا التباين عن (5%) وهي المحكات المحددة لهذا الاختبار. مما لا يترك مجالًا لظهور أبعاد أخرى مؤثرة بقوة على استجابة الأفراد لمفردات الاختبار، وبالتالي تحقق فرضية أحادية البعد بدرجة تكفي لإتمام باقي الإجراءات.

كما استخدم التحليل العاملي التوكيدي Confirmatory Factor Analysis (CFA) من خلال البرمجية Amos-22 والمعدة لتطبيقات نمذجة المعادلة البنائية Structural Equation Modeling (SEM) لاختبار البنية الأحادية البعد لاختبار رافن للمصفوفات المتتابعة الملونة. كما استخدمت مجموعة من مؤشرات حسن المطابقة للحكم على مدى ملاءمة النموذج النظري للبيانات والتي تم تلخيصها في جدول 2 والذي يوضح قيم هذه المؤشرات للنموذج الأحادي البعد.

وقد أسفرت نتائج التحليل العاملي التوكيدي عن درجة جيدة من المطابقة بين النموذج والبيانات وذلك وفقاً لمؤشرات حسن المطابقة المتعارف عليها والتي يوضحها جدول 2، حيث جاءت مؤشرات حسن المطابقة في المدى المثالي المقبول لها ( Byrne, 2006; Schweizer, 2010). ويوضح جدول 3 التشبعات المعيارية لمفردات الاختبار على العامل العام الذي يمثل الدرجة الكلية للذكاء وقد تراوحت قيم التشبعات بين 0.22 و0.78، وتزيد قيمة التشبعات المعيارية لنسبة 80% من المفردات عن 0.40 الأمر الذي يشير إلى درجة جيدة من التشبع على العامل العام، وهو ما يدعم تحقق أحادية البعد بقدر كاف في اختبار المصفوفات المتتابعة الملونة لرافن.

جدول 2

*مؤشرات حسن المطابقة للنموذج أحادي البعد لاختبار المصفوفات المتتابعة الملونة لرافن*

| مؤشرات المطابقة | القيمة | شروط قبول النموذج |
|---|---|---|
| مربع كاي χ2 | 1848.99 | عدم الدلالة الإحصائية لقيمة مربع كاي |
| مربع كاي المعياري ($\chi^2$/df) | 3.20 | أن تكون النسبة محصورة بين (1-5) |
| مؤشر جذر متوسط مربع الخطأ التقريبي (RMSEA) | 0.046 | أقل من 0,05 (تكون المطابقة جيدة) |
| مؤشر المطابقة المقارن (CFI) | 0.915 | CFI ≥ 0,90 (تكون المطابقة جيدة) |
| مؤشر توكر لويس (TLI) | 0.904 | TLI ≥ 0,90 (تكون المطابقة جيدة) |
| مؤشر جذر متوسط مربعات البواقي (SRMR) | 0.057 | أقل من 0,08 (تكون المطابقة جيدة) |

جدول 3

*قيم التشبعات المعيارية لمفردات اختبار المصفوفات المتتابعة الملونة*

| المفردة | التشبع المعياري | المفردة | التشبع المعياري | المفردة | التشبع المعياري |
|---|---|---|---|---|---|
| 1 | 0.22 | 13 | 0.32 | 25 | 0.33 |
| 2 | 0.28 | 14 | 0.44 | 26 | 0.56 |
| 3 | 0.37 | 15 | 0.40 | 27 | 0.66 |
| 4 | 0.39 | 16 | 0.67 | 28 | 0.68 |
| 5 | 0.35 | 17 | 0.70 | 29 | 0.74 |
| 6 | 0.44 | 18 | 0.78 | 30 | 0.57 |
| 7 | 0.60 | 19 | 0.73 | 31 | 0.54 |
| 8 | 0.36 | 20 | 0.78 | 32 | 0.69 |
| 9 | 0.59 | 21 | 0.69 | 33 | 0.65 |
| 10 | 0.70 | 22 | 0.59 | 34 | 0.63 |
| 11 | 0.50 | 23 | 0.55 | 35 | 0.63 |
| 12 | 0.46 | 24 | 0.47 | 36 | 0.50 |

**ب- الاستقلال الموضعي.**

ويتم الكشف عن هذا الافتراض من افتراضات نظرية الاستجابة للمفردة من خلال حساب معاملات الارتباط الجزئية Partial Correlation بين كل زوج من أزواج المفردات عند مستويات الدرجة المختلفة، بحيث إذا جاءت قيم هذه المعاملات قريبة بشكل كبير من الصفر فان ذلك يعتبر مؤشرًا على تحقق افتراض الاستقلال الموضعي. ولقد استخدم برنامج SPSS لحساب معاملات الارتباط الجزئية بين المفردات مع ضبط الدرجة الكلية على الاختبار.

وجاءت النتائج لتشير إلى أن معظم قيم معاملات الارتباط كانت أقل من (0.15) في حين جاءت بعض معاملات الارتباط أكثر من (0.2) والتي لوحظت بين المفردات شديدة السهولة كما هو الحال في المفردات (1، 2، 3، 4، 5)، ولذلك قرر الباحثان الإبقاء عليها حفاظاً على عدد المفردات في بنك الأسئلة، وكذلك لأن وجود العلاقة بينها كان بسبب زيادة نسبة الاجابة الصحيحة عليها عن 97% وليس بسبب ارتباط في المحتوى، وتوصي الدراسة الحالية بإعادة التحقق من ذلك على عينات تسمح بتباين الاستجابات على تلك المفردات.

**ج- التحرر من السرعة.**

لا يعتبر اختبار رافن للمصفوفات المتتابعة الملونة من اختبارات السرعة، ولكنه وكما تم التوضيح سابقاً من اختبارات القوة التي تركز على قياس القدرة على التفكير المنظم، وهو ما يؤكد على أن استجابات الطلبة في عينة الدراسة لم تتأثر بعامل الوقت.

المجلة الدولية للأبحاث التربوية- جامعة الإمارات العربية المتحدة
المجلد ( 44 ) العدد ( 3 ) أكتوبر 2020

**د- التأكد من ملاءمة الأفراد والمفردات لنماذج الاستجابة للمفردة.**

قبل البدء في هذا الإجراء تم التأكد من عدم وجود البيانات التامة في مفردات الاختبار، حيث لا يوجد مفردات أجاب عليها جميع أفراد العينة بشكل خاطئ أو بشكل صحيح، وفي الوقت نفسه تم حذف عدد 17 حالة من بين أفراد العينة أجابوا على جميع المفردات بشكل صحيح. وللتحقق من ملاءمة الأفراد والمفردات لنماذج الاستجابة للمفردة تم الاعتماد على محك المقارنة بين مستويات الملاءمة العامة للنماذج الثلاثة الأحادي والثنائي والثلاثي البارامتر وذلك بالاعتماد على مؤشر(AIC) Akaike Information Criterion الذي تشير القيمة الأقل له إلى النموذج الأكثر ملاءمة للمفردات والأفراد.

جدول 4

*المقارنة بين ملاءمة البيانات لنماذج الاستجابة للمفردة*

| النموذج | AIC |
|---|---|
| الأحادي | 1879.72 |
| الثنائي | 1561.08 |
| الثلاثي | 1339.09 |

وتشير النتائج في الجدول السابق إلى أن النموذج الثلاثي هو الأكثر ملاءمة ومناسبة للبيانات وهو ما أكدته قيم مربع كاي لملاءمة المفردات للنماذج الثلاثة حيث كانت الأقل بالنسبة للنموذج الثلاثي بمتوسط (13.72) يليه، النموذج الثنائي بمتوسط (15.11) ثم الأحادي البارامتر بمتوسط (26.49).

ولذلك يخلص الباحثان من النتائج السابقة إلى تحقق الافتراضات الأساسية لنماذج الاستجابة للمفردة، بالإضافة إلى وجود مستوى مقبول من الملامة مع النموذج ثلاثي البارامتر، ولذلك سيتم استخدامه في عملية معايرة وتدريج مفردات اختبار رافن للمصفوفات المتتابعة الملونة.

**الإجابة على السؤال الثاني**

هل يمثل بنك أسئلة اختبار رافن للمصفوفات المتتابعة الملونة مستودعاً كافياً وملائماً لاستخدامه لأغراض القياس التكيفي؟

يحتاج الاختبار التكيفي إلى بنوك أسئلة كبيرة حتى تسمح بإجراءات تطبيق الاختبارات التكيفية والتي تعتمد على اختيار المفردات المناسبة والملائمة لمستوى المفحوص. ويرى الباحثان أن مثل هذا الشرط سيحصر استخدام فكر الاختبارات التكيفية المحوسبة فقط مع أنظمة التقييم التي تعتمد على بنوك أسئلة ضخمة، وسيمنع الاستفادة منها في قياس القدرات العقلية أو السمات

النفسية والتي يصعب إعداد مثل هذه البنوك الضخمة فيها، وهو ما أكده ( Magis, Yan, & von Davier, 2017) بأنه في الواقع يصعب تحقيق الظروف التي يتواجد فيها هذا الشرط في بنوك الأسئلة، ولذلك يمكن القول بأن بنك الأسئلة يكون كافياً للاستخدام في الاختبارات التكيفية إذا كان يتضمن مفردات تتباين في مستويات صعوبتها ولا تتجمع عند مستويات محددة فقط من القدرة، ولهذا يسعى الباحثان إلى معرفة تحقق هذا الشرط في بنك المفردات في اختبار رافن للمصفوفات المتتابعة الملونة، وذلك بمعايرة مفرداته وفقاً للنموذج الثلاثي البارامتر ثم التعرف على توزيع المفردات من حيث مستويات صعوبتها على متصل القدرة.

ولقد تم الاعتماد على الحزمة (ltm) التي يتم استخدامها من خلال منصة R في معايرة مفردات بنك الأسئلة باستخدام النموذج الثلاثي البارامتر، ويوضح الجدول (5) قيم معاملات التمييز والصعوبة والتخمين لمفردات بنك الأسئلة.

جدول 5

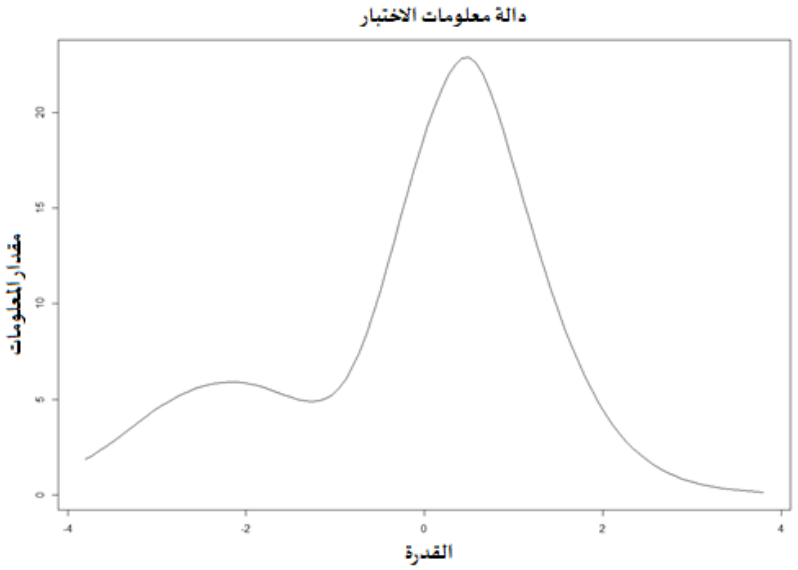
*قيم بارامترات الصعوبة والتمييز والتخمين لمفردات بنك الأسئلة*

| المفردة | الصعوبة | التمييز | التخمين | المفردة | الصعوبة | التمييز | التخمين |
|---|---|---|---|---|---|---|---|
| 1 | -3.38 | 1.41 | 0.01 | 19 | -0.18 | 2.67 | 0.17 |
| 2 | -3.01 | 2.28 | 0.01 | 20 | 0.15 | 2.75 | 0.06 |
| 3 | -2.57 | 2.25 | 0.00 | 21 | 0.41 | 2.49 | 0.11 |
| 4 | -2.36 | 1.77 | 0.00 | 22 | 0.59 | 2.19 | 0.18 |
| 5 | -2.06 | 2.19 | 0.00 | 23 | 0.49 | 1.73 | 0.19 |
| 6 | -1.75 | 1.98 | 0.00 | 24 | 1.21 | 3.05 | 0.12 |
| 7 | -0.39 | 1.57 | 0.08 | 25 | -1.72 | 1.81 | 0.59 |
| 8 | -0.32 | 0.80 | 0.20 | 26 | -0.70 | 2.33 | 0.36 |
| 9 | -0.28 | 1.85 | 0.21 | 27 | -0.51 | 3.41 | 0.30 |
| 10 | 0.12 | 2.54 | 0.15 | 28 | -0.07 | 3.07 | 0.30 |
| 11 | 1.26 | 2.42 | 0.09 | 29 | 0.09 | 2.62 | 0.12 |
| 12 | 1.53 | 1.86 | 0.05 | 30 | 0.53 | 2.02 | 0.21 |
| 13 | -1.92 | 2.51 | 0.43 | 31 | 0.87 | 2.96 | 0.16 |
| 14 | -1.25 | 2.33 | 0.44 | 32 | 0.54 | 3.05 | 0.06 |
| 15 | -2.17 | 1.42 | 0.00 | 33 | 0.55 | 3.04 | 0.11 |
| 16 | -0.02 | 2.89 | 0.27 | 34 | 0.48 | 4.02 | 0.22 |
| 17 | -0.04 | 2.58 | 0.20 | 35 | 0.72 | 3.05 | 0.08 |
| 18 | 0.01 | 2.87 | 0.09 | 36 | 1.63 | 2.18 | 0.07 |

المجلة الدولية للأبحاث التربوية- جامعة الإمارات العربية المتحدة
المجلد ( 44 ) العدد ( 3 ) أكتوبر 2020

ويتضح من قيم معاملات الصعوبة في جدول 5 أنها تتوزع عبر متصل القدرة، حيث تتراوح قيمها بين (-3.38 – 1.63) وهو ما يمكن اعتباره مؤشراً جيداً على توافر تغطية مناسبة لمتصل القدرة يمكن الاعتماد عليها في التحرك قدماً وتكملة إجراءات التحقق من خصائص الاختبار التكيفي، ويوضح الشكل (1) دالة المعلومات لبنك الأسئلة لاختبار رافن للمصفوفات المتتابعة الملونة، وهي توضح تغطية الاختبار فعلياً لمستويات من القدرة منخفضة ولكن بمقدار معلومات أقل مقارنة بمقدار المعلومات في الفترة من (-1 ألى 2).

ويبرر انخفاض قيمة المعلومات خارج الفترة (-1، 2) بانخفاض القدرة التمييزية للمفردات التي تقع فيها عبر متصل القدرة، حيث يلاحظ وصول معامل التمييز لأعلى مستوياته للمفردات في تلك الفترة مقارنة بخارجها، الأمر نفسه يمكن أنه يفسره عدد المفردات، حيث يلاحظ وجود النسبة الأكبر من المفردات داخل هذه الفترة (26 مفردة)، في حين يوجد عدد قليل من المفردات خارجها (10 مفردات)، ويلاحظ كذلك ارتفاع نسبة التخمين لتصل (0.44، 0.43، 0.59) عند مستوى قدرة (-1.25، -1.92، -1.72) على الترتيب وهو ما قد ينعكس على دقة تقدير القدرة خارج تلك الفترة، وقد أكدت دراسة (Paek, 2015) على تأثر ثبات التقديرات بزيادة مستوى تخمين المفردات، حيث أشارت إلى انخفاض معاملات الثبات مع زيادة قيمة التخمين. وفي الوقت نفسه يرتبط زيادة قيمة المعلومات للاختبار بشكل عام بمناطق تجمع المفردات على متصل القدرة، وكذلك قمية تمييز تلك المفردات، حيث يعتمد مقدار المعلومات الذي تقدمه المفردة عند موضعها على متصل القدرة على مستوى تمييزها وهو الأمر الذي أكد عليه بيكر (Baker, 2001).




*شكل 1*. دالة معلومات اختبار رافن للمصفوفات المتتابعة الملونة

**الإجابة على السؤال الثالث.**

ما مستوى فاعلية الصورة التكيفية المحوسبة من اختبار رافن للمصفوفات المتتابعة الملونة؟

للإجابة على هذا السؤال استخدم الباحثان الحزمة (catR) والتي تستخدم من خلال منصة التحليل R، وهي تمكن من عمل محاكاة للبيانات الواقعية التي استخدمت في معايرة مفردات البنك بالإضافة إلى خصائص المفردات والأفراد في العينة، كما تمكن هذه الحزمة من اختبار عدد من الظروف التي يمكن من خلالها وصف مدى فاعلية الاختبار التكيفي المحوسب مثل محكات إنهاء الاختبار التكيفي (طول الاختبار، أو دقة تقدير القدرة) أو محكات اختيار المفردات. وسوف تقوم الدراسة الحالية بالتحقق من فعالية الاختبار التكيفي المحوسب تحت شروط متنوعة لإيقاف الاختبار (طول الاختبار: 4، 8، 12، 16، 18، والخطأ المعياري لتقدير القدرة: 0.25، 0.30، 0.35) مع استخدام طريقة المعلومات القصوى Maximum Fisher Information لانتقاء المفردات.

جدول 6

*يوضح أداء اختبار رافن للمصفوفات المتتابعة في صورته التكيفية تحت شروط الدراسة المختلفة*

| التصميم | طول | الخطأ المعياري | متوسط الخطأ | الدقة | Bias | عدد | متوسط عدد |
|---|---|---|---|---|---|---|---|
| 1 | 4 | | 0.53 | 0.81 | 0.09- | - | - |
| 2 | 8 | | 0.38 | 0.85 | 0.09- | - | - |
| 3 | 12 | | 0.33 | 0.87 | 0.08- | - | - |
| 4 | 16 | | 0.30 | 0.90 | 0.05- | - | - |
| 5 | 18 | | 0.31 | 0.91 | 0.05- | - | - |
| 6 | | 0.25 | 0.28 | 0.93 | 0.05- | 34-13 | 22 |
| 7 | | 0.30 | 0.31 | 0.92 | 0.04- | 33-9 | 17 |
| 8 | | 0.35 | 0.35 | 0.90 | 0.03- | 27-6 | 12 |
| 9 | 4 | 0.25 | 0.53 | 0.83 | 0.06- | 4-4 | 4 |
| 10 | 12 | 0.25 | 0.33 | 0.86 | 0.10- | 12-11 | 11.6 |
| 11 | 18 | 0.25 | 0.30 | 0.90 | 0.07- | 18-15 | 15.5 |
| 12 | 4 | 0.30 | 0.55 | 0.85 | 0.10- | 4-4 | 4 |
| 13 | 12 | 0.30 | 0.34 | 0.87 | 0.08- | 12-8 | 10 |
| 14 | 18 | 0.30 | 0.32 | 0.90 | 0.09- | 17-9 | 12.5 |
| 15 | 4 | 0.35 | 0.52 | 0.76 | 0.10- | 4-4 | 4 |
| 16 | 12 | 0.35 | 0.37 | 0.89 | 0.07- | 12-6 | 8.4 |
| 17 | 18 | 0.35 | 0.36 | 0.89 | 0.04- | 17-6 | 10.2 |

يوضح جدول 6 معدلات دقة الصور التكيفية للاختبار في الحالات المختلفة، والتي تمثل قيم معامل الارتباط بين تقديرات القدرة باستخدام الصورة التكيفية وتلك المقدرة من الاختبار

المجلة الدولية للابحاث التربوية- جامعة الإمارات العربية المتحدة
المجلد ( 44 ) العدد ( 3 ) أكتوبر 2020

ككل، والتي توضح أن دقة الدرجة المقدرة من الصورة التكيفية تزيد من 0.81 إلى 0.91 عند زيادة عدد مفردات الاختبار من 4 إلى 18 مفردة في الصورة التكيفية، وفي الوقت نفسه يلاحظ انخفاض القيمة المطلقة للتحيز من 0.09 إلى 0.05، وهو ما يشير إلى أن استخدام 50% من مفردات اختبار رافن للمصفوفات المتتابعة عند تطبيقه في الصورة التكيفية كافي لتحقيق درجة عالية من الدقة في درجات المفحوصين تصل إلى 0.91. وهو ما يدعم جدوى الصورة التكيفية من اختبار رافن للمصفوفات المتتابعة الملونة.

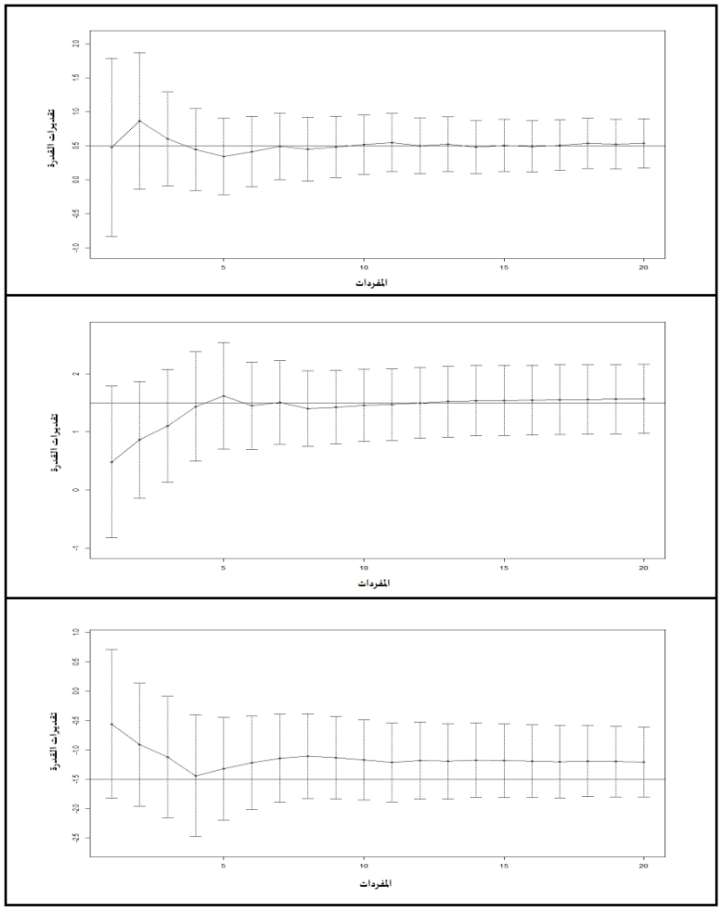
وبالنظر إلى النتائج في ضوء محك إنهاء الاختبار وفق مستوى الخطأ المعياري لتقدير القدرة، يمكن ملاحظة أن استخدام محك الخطأ المعياري 0.35 يحقق مستوى دقة في تقدير قدرة المفحوصين بالصورة التكيفية وبمتوسط عدد مفردات 12 مفردة يصل إلى 0.90 وهو يشابه مستوى الدقة التي تم التوصل إليه باستخدام محك الإنهاء (18 مفردة)، كما يعكس هذا الجزء من الجدول أيضاً أن زيادة مستوى الدقة إلى 0.93 يتطلب زيادة ملموسة في عدد المفردات تصل إلى (22 مفردة).

وفي تصميمات الدراسة من 9 إلى 17 لم يلاحظ الباحثان تأثير واضح أو ملموس للدمج بين محكي إنهاء الاختبار، حيث تقترب كثيراً درجات الدقة مع نظيرتها من حيث طول الاختبار في التصميمات من 1 إلى 5، بل ويلاحظ انخفاض في معدلات ثبات الاختبار أو دقة تقديرات القدرة عند المقارنة مع نظيراتها في التصميمات من 6 إلى 8.

وإجمالاً يرى الباحثان أن إمكانية الوصول إلى مستوى ثبات يصل إلى 0.9 كحد أدنى عند استخدام محك انهاء بطول اختبار من 16 إلى 18 مفردة (بتخفيض في عدد المفردات يصل إلى أكثر من 50%) أو بتحديد الخطأ المعياري لتقدير القدرة بين 0.30 و0.35 ، وتقترب قيمة الثبات التي تمكن الوصول إليها بالصورة التكيفية في دراسة المحاكاة الحالية مع نظيرتها في دراسة البرصان وبخيت (2015) مع الانتباه إلى اختلاف الاختبارين المستخدمين، حيث استخدم اختبار المصفوفات المتتابعة المعياري في تلك الدراسة. كما ويقترب طول الاختبار التكيفي المقترح في الدراسة الحالية لتحقيق الدرجة المشار إليها من الدقة مع نتائج دراسة يوليانتو (Yulianto, 2007) والذي حددها بـ 12 مفردة من أصل 36 في اختبار المصفوفات المتتابعة المتقدم لرافن (وهو يشابه اختبار المصفوفات المتتابعة الملونة في عدد المفردات الكلية)، ويرى الباحثان أن هذه النتيجة يمكن أن تكون مؤشراً قوياً على جدوى استخدام الاختبار التكيفي للمصفوفات المتتابعة الملونة. كما وتعتبر أيضاً دافعاً لمحاولة نشر اختبار رافن للمصفوفات المتتابعة الملونة على منصات الاختبارات التكيفية.

ويرى الباحثان في الدراسة الحالية أن اعتماد محك إيقاف الاختبار بتحديد مستوى للخطأ

المعياري لتقدير القدرة أكثر دقة إذا ما استخدم بشكل عام، حيث يحافظ بمستوياته الثلاثة على درجة دقة تزيد عن أو تساوي 0.9، وهذا بشكل عام يتماشى مع ما ذكره (Stafford, et al., 2018) حول سلبيات تحديد عدد ثابت للمفردات الأمر الذي يزيد من نسبة ظهور السؤال في الصورة التكيفية إذا ما قورن باختيار طريقة تحديد الخطأ المعياري لتقدير القدرة والذي لا يحدد عدداً ثابتاً من الأسئلة في الصورة التكيفية، وهذا وفقاً لما أشار إليه (Leroux & Dodd, 2014).




*شكل 2*. مسار تقديرات القدرة لثلاثة أفراد افتراضيين من مستويات قدرة مختلفة وهي (0.5، 1.5، -1.5) على الترتيب، باستخدام طول الاختبار "20" كقاعدة لإنهاء الاختبار التكيفي

وتوضح الرسوم الثلاثة في الشكل 2 إمكانية الوصول إلى تقديرات مستقرة للقدرة بدرجة كبيرة بعد تطبيق 15 مفردة من الاختبار التكيفي على المفحوصين، حيث تتناول الرسوم أمثلة لأداء ثلاثة أفراد افتراضيين من مستويات متباينة في القدرة (0.5، 1.5، -1.5)، ويوضح الشكل ارتفاع

المجلة الدولية للأبحاث التربوية- جامعة الإمارات العربية المتحدة
المجلد ( 44 ) العدد ( 3 ) أكتوبر 2020

خطأ التقدير لدرجة المفحوص الثالث، وهو ما يمكن توقعه فعلاً بالنظر إلى دالة معلومات بنك أسئلة اختبار رافن للمصفوفات المتتابعة الملونة، والذي يظهر إنخفاض ملحوظ في مقدار المعلومات عند هذا المستوى من القدرة، ولذلك تجدر الإشارة إلى أن فعالية الاختبار التكيفي المحوسب لا يجب النظر إليها بعيداً عن دالة معلومات بنك الأسئلة الذي يعتمد عليه الاختبار التكيفي.

وبشكل عام، فإنه للتحقق من مدى إمكانية تطوير صورة تكيفية محوسبة من الاختبار كان من الضروري التحقق أولا من إمكانية تدريج الاختبار باستخدام نظرية الاستجابة للمفردة، ثم الانتقال إلى تقييم آداء الصورة التكيفية من خلال محاكاة التطبيق بالاعتماد على نتائج التدريج في الخطوة السابقة باستخدام البرمجية المطورة بلغة البرمجة R وهي الحزمة .mirtCAT ولقد تم التحقق من شرط أحادية البعد باستخدام التحليل العاملي وفقاً لبرنامج winsteps والذي أكد على وجود عامل عام وراء استجابات المفحوصين على الاختبار أكدته نتائج التحليل العاملي التوكيدي أيضاً وهو يتفق مع ما أكد عليه عودة وعبيدات (2013) وكذلك (Cikrikci-Demirtasli, 2002) حول أحادية البعد والاستقلال الموضعي للمفردات ما يعني صلاحية اختبار رافن لتطبيقات نماذج الاستجابة للمفردة.

وقد أظهرت مفردات اختبار رافن مستوى جيد من الملاءمة للنموذج الثلاثي البارامتر إحدى نماذج نظرية الاستجابة للمفردة ثانية التدريج dichotomous والذي اعتمدته الدراسة الحالية في تدريج مفردات الاختبار. وباستخدام طريقة انتقاء المفردة وفقاً لمحك المعلومات القصوى Maximum Fisher Information مع أطوال مختلفة من المفردات لانهاء الاختبار، تم تقييم إمكانية تكوين صورة تكيفية من عدد أقل من المفردات وذلك بالمقارنة بين قدرات المفحوصين المقدرة وفق العدد الكلي من المفردات والصورة التكيفية، ولقد أظهرت نتائج التقييم درجة جيدة من التقارب بين مستويات القدرة المقدرة باستخدام العدد الكلي للمفردات وتلك المقدرة بالصورة التكيفية بعدد مفردات يصل إلى 17 مفردة دون فقدان كمية جوهرية من المعلومات والدقة في تقدير القدرة. كما كان الاعتماد على محك إيقاف الاختبار بتحديد مستوى للخطأ المعياري بين 0.25 و 0.35 يصل إلى درجة أفضل من الثبات والدقة في تقدير القدرة، وكذلك متوسط عدد المفردات، والذي يصل إلى ما بين 12 و مفردة17 إذا كانت مستويات الخطأ المعياري لتقدير القدرة في الفترة بين 0.30 و 0.35.

وتعتبر النتيجة السابقة مؤشراً أولياً على إمكانية تطوير صورة تكيفية من اختبار رافن للمصفوفات المتتابعة الملونة، واستخدامها في التقييمات المختلفة التي قد يحتاجها المسؤلون في تقييم أو تصنيف أو تشخيص الأطفال، وخاصة عند التعامل مع الأطفال المتأخرين عقلياً. ويرى الباحثان أن تطوير مثل هذه الصور من الاختبارات يحدث نقلة كبيرة في ميدان الاختبارات

والمقاييس وخاصة تلك التي تتكون من عدد كبير من المفردات، حيث يمكن من خلالها الوصول إلى تقييمات دقيقة متوائمة مع الفرد باستخدام أقل عدد ممكن من المفردات الاختبارية، وهو ما يعالج أحد القضايا التي تؤثر في صدق استجابات المفحوصين وهي طول الاختبار .

وتجدر الإشارة إلى أن المعايرة الحالية لمفردات اختبار رافن في ضوء العينة المشار إليها لا تعتبر كافية لتقديم خصائص دقيقة للمفردات، حيث توجد بعض المفردات التي لم تتباين استجابات المفحوصين عليها بشكل كاف للوصول لتقديرات دقيقه لبارامتراتها، وهو ما قد يغير من دالة معلومات الاختبار وكذلك إمكانية تقديم تقديرات أكثر دقة للمفحوصين في مدى أكثر اتساعاً من الممثل في الشكل 1. لذلك يوصي الباحثان بضرورة توسيع نطاق العينة المستخدمة في معايرة مفردات الاختبار باستخدام نظرية الاستجابة للمفردة حتى تسمح بمستويات أعلى من التباين في الأداء على جميع مفردات الاختبار. كما يوصي الباحثان أيضاً بضرورة الانتقال إلى مرحلة استخدام الصورة التكيفية من اختبار المصفوفات المتتابعة بوضعه على المنصات المعدة لذلك الغرض، وذلك لما أظهرته من إمكانية الوصول لنتائج دقيقة بعدد أقل من المفردات الاختبارية الملائمة للمفحوصين، بالإضافة إلى إمكانية الاعتماد على إحدى محكين لإنهاء الاختبار وهما بتحديد طول اختبار من 16 إلى 18 مفردة (بتخفيض في عدد المفردات يصل إلى أكثر من 50%)، أو بتحديد الخطأ المعياري لتقدير القدرة بين 0.30 و0.35 وهو ما يحقق مستوى من الدقة يصل إلى 90% على الأقل مع متوسط عدد مفردات أقل، الأمر الذي يشير إلى مدى جيد من الدقة في اتخاذ القرار ونسبة أقل للتعرض لمفردات الاختبار.

كما ويوصي الباحثان بضرورة استخدام عينة حديثة لتقنين الاختبار مرة أخرى ومقارنة معاييره الحديثة بالمعايير المشتقة من عينة 2008، وكذلك للتحقق من صدق نتائج الصورة التكيفية التي يوصى بإعدادها في ضوء نتائج الدراسة الحالية بمقارنتها مع النتائج للصورة الكاملة من الاختبار في ضوء بيانات واقعية.

## التمويل

لم يحصل هذا البحث على أي تمويل من الجهات الممولة للبحوث العلمية، لكنه اعتمد على بيانات ارشيفية من بيانات مشروع بحثي ممول ضمن المنح الداخلية لجامعة السلطان قابوس (IG/EDU/PSYC/05/01).

## تضارب المصالح

أفاد الباحثون بعدم وجود تضـارب في المصـالح فيما يتعلق بالبحث، والملكية الفكرية، ونشر هذا البحث.

المجلة الدولية للأبحاث التربوية - جامعة الإمارات العربية المتحدة
المجلد ( 44 ) العدد ( 3 ) أكتوبر 2020

## المراجع

ابن زرقين، محمد (2016). *تقنين اختبار المصفوفات المتتابعة المتقدم لجون رافن على الطلبة الجامعيين: دراسة ميدانية على عينة من طلبة جامعة قاصدي مرباح ورقلة* (رسالة ماجستير غير منشورة). كلية العلوم الإنسانية والاجتماعية بجامعة محمد بوضياف، الجزائر.

آل ثاني، العنود مبارك بن أحمد (2002). تقنين اختبار المصفوفات المتتابعة العادي لرافن على طلاب وطالبات المرحلة الابتدائية بمدينة الدوحة بدولة قطر. *مجلة مركز البحوث التربوية*، 21، 234-292.

البرصان، إسماعيل سلامة (2012). فاعلية القياس التكيفي باستخدام فقرات ذات إجابة منتقاة وفقرات ذات إجابة منشأة. *مجلة جامعة الملك سعود العلوم التربوية والدراسات الإسلامية، 24*(4)، 1487 – 1518.

البرصان، إسماعيل سلامة وبخيت، صلاح الدين فرح (2015). فاعلية القياس التكيفي في الكشف عن الموهوبين باستخدام اختبار رافن للمصفوفات المتتابعة المعياري. *رسالة التربية وعلم النفس*، (50)، 75-97.

الحسن، أسماء عدنان؛ ورحمة، عزيزة (2017). الخصائص السيكومترية لاختبار المصفوفات المتتابعة المعيارية (SPM) لرافن على تلاميذ الفئة العمرية (9-12) سنة ودراسة الفروق بينهم في ضوء عدة متغيرات. *مجلة جامعة البعث، 39*(72)، 11- 49.

حماد، إبراهيم مصطفى علي (2012). *تقنين اختبار المصفوفات المتتابعة الملون في البيئة الفلسطينية (رسالة ماجستير غير منشورة)*. كلية التربية، الجامعة الإسلامية بغزة، فلسطين.

الخطيب، محمد الأمين مصطفة، ومصطفى، مهيد محمد المتوكل، وحسين، أسماء ميرغني (2006). *تقنين اختبار المصفوفات المتتابعة الملون لتلاميذ الحلقة الأولى بمرحلة الأساس بولاية الخرطوم*. الخرطوم: شركة مطابع السودان المحدودة.

ربابعة، حسين أحمد عايد (2012). *أثر عدد الفقرات المرشحة للتقديم من بنك أسئلة في فاعلية أسلوب كنغرباري-زارا لضبط تقديم الفقرات متعددة الخطوات في الاختبارات التكيفية المحوسبة (رسالة دكتوراة غير منشورة)*. كلية التربية، جامعة اليرموك، الأردن.

العاني، علاء الدين جميل (1989). *بناء معايير عراقية لاختبار المصفوفات المتتابعة الملون* (رسالة ماجستير غير منشورة). كلية التربية، جامعة بغداد، العراق.

علام، صلاح الدين محمود (2000). *القياس والتقويم التربوي والنفسي: أساسياته وتطبيقاته وتوجهاته المعاصرة*. القاهرة: دار الفكر العربي.

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

عيد، أحمد (1999). *اختبار المصفوفات المتتابعة الملونة (دليل الاستخدام)*. وزارة التربية والتعليم: دولة الإمارات العربية المتحدة.

القرشي، عبد الفتاح (1987). *اختبار المصفوفات المتتابعة الملونة (دليل الاستخدام)*. الكويت: دار القلم.

كاظم، أمينة (1988). *استخدام نموذج راش في بناء اختبار تحصيلي في علم النفس وتحقيق التفسير الموضوعي للنتائج*. مطبوعات جامعة الكويت.

كاظم، علي مهدي والزبيدي، عبد القوي سالم والصارمي، عبد الله ويوسف، حسن يوسف والجمالية، فوزية عبد الباقي والمشهداني، سكرين ابراهيم والبلشوية، سناء سبيل والخروصي، حسين على والبوسعيدي، أميمة بدر والبحرانية، وداد عبد الله، والفورية، سعاد مبروك (2008). تقنين اختبار رافن للمصفوفات المتتابعة الملونة على الأطفال العمانيين في المرحلة العمرية بين 5 – 11 سنة. *دراسات نفسية، 18*(3)، 391-429.

الهيتي، خلف نصار، والغانم، عزة محمد، والزبيدي، عبدالقوي سالم، والنود، يحي على (1995). *تقنين اختبار المصفوفات المتتابعة الملون لرافن على أطفال مرحلة التعليم الأساسي في الجمهورية اليمنية للعمر 6-11 سنة*. صنعاء: منظمة الأمم المتحدة للطفولة (اليونسيف).

متولي، فكري لطيف؛ والقحطاني، شتوي مبارك (2016) *صعوبات التعلم للمبتكرين والموهوبين*. القاهرة: مكتبة الأنجلو المصرية.

محمد، مصطفى أحمد قاسم؛ كاظم، أمينة محمد (2015) المقارنة بين الاختبار الورقي والاختبار المحوسب في تقدير القدرة على الاستدلال اللفظي. *مجلة البحث العلمي في التربية، 1*(16)، 443-464.

مخائيل، امطانيوس (2005). *القياس النفسي*. دمشق: منشورات جامعة دمشق.

مسلم، فؤاد عبد الرازق محمود (2011). *أسلوب مقترح لاختيار الفقرات وتقدير القدرة في الاختبارات التكيفية المحوسبة (رسالة دكتوراة غير منشورة)*. كلية التربية، جامعة اليرموك، الأردن.

Al-Ani, A. J. (1989). *Building Iraqi standards for Raven's Colored Progressive Matrices Test* (Unpublished master thesis) (In Arabic). College of Education, University of Baghdad, Iraq.

Al-Bursan, I. S. (2012). The efficacy of adaptive measurement by using selected-answer paragraphs and constructed-answer paragraphs (In Arabic). *Journal of Educational Sciences and Islamic Studies- King Saud University*, *24* (4), 1487-1518.

Al-Bursan, I. S., & Bakhit, S. F. (2015). The effectiveness of adaptive

measurement in detecting gifted persons using Raven's Standard Progressive Matrices Test (In Arabic). *Education and Psychology Mission*, *50*, 75-97.

Al-Hassan, A. A., & Rahma, A. (2017). Psychometric properties of Raven's Standard Progressive Matrices Test on students of the age group (9-12) years, and studying the differences between them in light of several variables (In Arabic). *Al-Baath University Journal*, *39* (72), 11-49.

Al-Hiti, K. N., Al-Ghanem, A. M., Al-Zubaidi, A. S., & Al-Nud, Y. A. (1995). *Standardization of the Raven's Colored Progressive Matrices Test for children of basic education in the Republic of Yemen, ages 6-11 years old* (In Arabic). Sana'a: The United Nations Children's Fund (UNICEF).

Al-Khatib, M. M., Mustafa, M. M., & Husayn, A. M. (2006). *Standardization of the Raven's Progressive Matrices Test for students in the first stage of the basic stage in Khartoum State* (In Arabic). Khartoum: Sudan Printing Press Company Ltd.

Allam, S. M. (2000). *Educational and psychological measurement and evaluation: its basics, applications and contemporary trends* (In Arabic). Cairo: Dar Al-Fikr Al-Arabi.

Al-Qurashi, A. (1987). *Raven's Colored Progressive Matrices Test* (User Manual) (In Arabic). Kuwait: Dar Al-Qalam.

Al-Thani, A. M. (2002). Standardization of Raven's Progressive matrices test for elementary school students in Doha, State of Qatar (In Arabic). *Journal of the Center for Educational Research*, (21), 234-292.

Babcock, B., & Weiss, D. J. (2012). Termination criteria in computerized adaptive tests: Do variable-length CATs provide efficient and effective measurement. *Journal of Computerized Adaptive Testing*, 1, 1–18.

Baker, F. B. (2001). *The basics of item response theory*. ERIC Clearinghouse on Assessment and Evaluation, College Park, MD.

Bandeira, D. R., Alves, I. C. B., Giacomel, A. E., & Lorenzatto, L. (2004). Matrizes progressivas coloridas de Raven - escala especial: Normas para Porto Alegre, RS. *Psicologia em Estudo*, *9*(3), 479-486. https: //dx.doi.org/10.1590/S1413-73722004000300016

Bass, N. (2000). *The Raven's Colored Progressive Matrices Test: A pilot study for the establishment of normative data for Xhosa-speaking Primary School pupils in the Grahamstown region* (Unpublished master thesis). Rhodes University, Grahamstown.

Boyd, A. M. (2003). *Strategies for controlling testlet exposure rates in computerized adaptive testing systems*. (Unpublished PhD Thesis), The University of Texas at Austin.

Byrne, B. M. (2006). *Structural equation modeling with EQS and EQS/Windows: Basic concepts, applications, and programming*. New York and London: Psychology Press.

Carman, C. A. (2011). Adding personality to gifted identification: Relationships among traditional and personality-based constructs. *Journal of Advanced Academics*, *22*, 412–446.

Cikrikci-Demirtasli, N. (2002). A study of Raven Standard Progressive Matrices Test's item measures under classic and item response models: An empirical comparison. Ankara University, *Journal of Faculty of Educational Science, 35*(1-2), 71-79.

Costenbader, V., & Ngari, S. M. (2001). A Kenya Standardization of the Ravens Coloured Progressive Matrices. *School Psychology International, 22*(3), 258–268. doi: 10.1177/0143034301223003

Cotton, S. M., Kiely, P. M., Crewther, D. P., Thomson, B., Laycock, R., & Crewther, S. G. (2005). A normative and reliability study for the Raven's Coloured Progressive Matrices for primary school aged children from Victoria, Australia. *Personality and Individual Differences, 39*(3), 647–659. doi: 10.1016/j.paid.2005.02.015

Eid, A. (1999). *Raven's Colored Progressive Matrices Test* (User Manual) (In Arabic). Ministry of Education: The United Arab Emirates.

Hambleton, R., & Swaminathan, H. (1985). *Item response theory. Principles and Application*. Boston: Kluwer-Nighoff Publishing.

Hammad, I. M. (2012). *Standardization of the Raven's Colored Progressive Matrices Test in the Palestinian environment* (unpublished master thesis) (In Arabic). College of Education, Islamic University of Gaza, Palestine.

Ibn-Zarqin, M. (2016). *Standardization of John Raven's Advanced Progressive Matrices Examination for Undergraduate Students: A study on a sample of Qasdi Marbah and Ouargla University students* (unpublished master thesis) (In Arabic). Faculty of Humanities and Social Sciences at Mohamed Boudiaf University, Algeria.

Kadhim, A. M., Zubaidi, A. S., Al-Sarmi, A., Yusuf, H. Y., Al-Jamalia, F. A., Al-Mashhadani, S. I., Al-Balshwiyya, S. S., Al-Kharousi, H. A., Al-Busaidi, O. B., Al-Bahraniya, W. A., & Al-Furiya, S. M. (2008). Standardization of the Raven's Colored Progressive Matrices Test for Omani children aged 5-11 years (In Arabic). *Psychological Studies*, *18* (3), 391-429.

Kail, R. V. (2004). Children and their Development, third edition. Library of Congress person Education, INC.

Kazem, A. (1988). *Using Rasch model in constructing an achievement test in psychology and achieving an objective interpretation of the results* (In Arabic). Kuwait University Publications.

Leroux, A. J., & Dodd, B. G. (2014). A comparison of stopping rules for computerized adaptive screening measures using the rating scale model. *Journal of Applied Measurement*, 15, 213–226.

Liegl, G., Petersen, M., Groenvold, M., Aaronson, N., Costantini, A., Fayers, P., & Nolte, S. (2019). Establishing the European norm for the health-related quality of life domains of the computer-adaptive test EORTC CAT Core. *European Journal of Cancer, 107*, 133-141. doi:10.1016/j.ejca.2018.11.023

Linacre, J. M. (2006). Data variance explained by measures. *Rasch Measurement Transactions, 20*, 1045–1047.

Linacre, J. M. (2000). Computer-adaptive testing: A methodology whose time has come. *Rasch Measurement Transactions, 69*, http://www.rasch.org/memo69.pdf

Linstrom, A. (2002, July). Variance in colored progressive matrices (CPM) scores with ethnicity, economics circumstances and living conditions in South Africa. *Paper presented at the 25th International Congress of Applied Psychology in Singapore*.

Lord, F. M. (1980). *Applications of item response theory to practical test* problems. Hillsdale, NJ: Erlbaum.

Lynn, R. and Vanhanen, T. (2002). *IQ and the wealth of nations*. Westport, Connecticut: Praeger.

Magis, D., Yan, D., & von Davier A.A. (2017). Computerized adaptive and multistage testing With R: Using packages catr and mstr. Cham, Switzerland: SPRINGER INTERNATIONAL PU.

Mccabe, E., Gross, D. P., & Bulut, O. (2018). Procedures to develop a computerized adaptive test to assess patient-reported physical functioning. *Quality of Life Research, 27*(9), 2393-2402. doi:10.1007/s11136-018-1898-0

Metwally, F. L., Al-Qahtani, S. M. (2016*) Learning Difficulties for Innovators and Talented People* (In Arabic). Cairo: The Anglo-Egyptian Library.

Mikhael, A. (2005). *Psychological Measurement* (In Arabic). Damascus: Damascus University Publications.

Mizumoto, A., Sasao, Y., & Webb, S. A. (2017). Developing and evaluating a computerized adaptive testing version of the Word Part Levels Test. *Language Testing, 36*(1), 101-123. doi:10.1177/0265532217725776

Muhammad, M. Q., & Kazem, A. M. (2015) Comparison between the paper test and the computerized test for estimating the verbal reasoning ability (In Arabic). *Journal of Scientific Research in Education*, *1* (16), 443-464.

Muslim, F. M. (2011). *Suggested method for paragraph selection and ability estimation in computerized adaptive tests* (Unpublished Doctoral Dissertation) (In Arabic). College of Education, Yarmouk University, Jordan.

Odeh, A. S., & Obaidat, O. S. (2013). The effectiveness of a computerized adaptive test in estimating mental ability using Raven's Matrices (In Arabic). *Educational Sciences Studies, 40* (2), 1602-1621.

Özyurt, H. & Özyurt, Ö. (2015). Ability level estimation of students on probability unit via computerized adaptive testing. *Eurasian Journal of Educational Research*, 58, 27-44.http://dx.doi.org/10.14689/ejer.2015.58 .7

Paek, I. (2015). An investigation of the impact of guessing on coefficient α

and reliability. *Applied Psychological Measurement*, *39*(4), 264–277. doi:10.1177/0146621614559516

Pierson, E. E., Kilmer, L. M., Rothlisberg, B. A., & Mcintosh, D. E. (2011). Use of brief intelligence tests in the identification of giftedness. *Journal of Psychoeducational Assessment, 30*(1), 10–24. doi: 10.1177/0734282911 428193

Rababaa, H. A. (2012). *The effect of the number of nominated paragraphs for submission from an item bank on the effectiveness of the Kanarbari-Zara method for controlling the presentation of multi-stage paragraphs in computerized adaptive tests* (Unpublished Doctoral Dissertation) (In Arabic). College of Education, Yarmouk University, Jordan.

Rezaie, M. & Golshan, M. (2015). Computer Adaptive Test (CAT): Advantages and limitations. *International Journal of Educational Investigations*, *2*(5), 128-137.

Schweizer, K. (2010). Some guidelines concerning the modeling of traits and abilities in test construction. *European Journal of Psychological Assessment, 26*(1), 1-2.doi: 10.1027/1015-5759/a000001

Stafford, R. E., Runyon, C. R., Casabianca, J. M., & Dodd, B. G. (2018). Comparing computer adaptive testing stopping rules under the generalized partial-credit model. *Behavior Research Methods, 51*(3), 1305–1320. doi: 10.3758/s13428-018-1068-x

Thompson, Nathan A., & Weiss, David A. (2011). A Framework for the development of computerized adaptive tests. *Practical Assessment, Research & Evaluation, 16*(1). Available online: http://pareonline.net/getvn.asp?v=16&n=1.

Uno, A., Shinya, N., Haruhara, N., & Kaneko, M. (2005). Ravens Coloured Progressive Matrices in Japanese children: As a screening intelligence test for children with learning disorder and acquired childhood aphasia. *Japan Journal of logop- Phonates, 49,* 185-189.

Wang, C., & Lu, H. (2018). Mediating effects of individuals' ability levels on the relationship of reflective-impulsive cognitive style and item response time in CAT. *Educational Technology & Society, 21* (4), 89–99.

Weiss, D. J., & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement, 21*, 361-375.

Wellisch, M., & Brown, J. (2012). An integrated identification and intervention model for intellectually gifted children. *Journal of Advanced Academics, 23*(2), 145–167.

Yulianto, A. (2007, July). *The Effectiveness of Computerized Adaptive Testing on Advanced Progressive Matrices*. Paper presented at International Meeting on Psychometric Society (IMPS) 2007, Tokyo, Japan.