مجلة جامعة الشارقة للعلوم الإنسانية والاجتماعية

اسم المقال: طريقة مطورة لتفسير أنماط الاستجابات غير المطابقة للفرد في الاختبارات ومقارنة خمسة مؤشرات مطابقة الفرد
اسم الكاتب: راشد سيف المحرزي، أمل ثابت عودة
رابط ثابت: https://political-encyclopedia.org/index.php/library/9040
تاريخ الاسترداد: 2026/11/01 20:48 +03

الموسوعة السياسيّة هي مبادرة أكاديمية غير هادفة للربح، تساعد الباحثين والطلاب على الوصول واستخدام وبناء مجموعات أوسع من المحتوى العلمي العربي في مجال علم السياسة واستخدامها في الأرشيف الرقمي الموثوق به لإغناء المحتوى العربي على الإنترنت.
لمزيد من المعلومات حول الموسوعة السياسيّة – Encyclopedia Political، يرجى التواصل على info@political-encyclopedia.org

استخدامكم لأرشيف مكتبة الموسوعة السياسيّة – Encyclopedia Political يعني موافقتك على شروط وأحكام الاستخدام المتاحة على الموقع https://political-encyclopedia.org/terms-of-use

تم الحصول على هذا المقال من موقع مجلة جامعة الشارقة للعلوم الإنسانية والاجتماعية ورفده في مكتبة الموسوعة السياسيّة مستوفياً شروط حقوق الملكية الفكريّة ومتطلبات رخصة المشاع الإبداعي التي ينضوي المقال تحتها.

جامعة الشارقة
UNIVERSITY OF SHARJAH

مجلة جامعة الشارقة

مجلة علمية محكمة

للعلوم الإنسانية والاجتماعية



عدد A

المجلد 17، العدد 1
شوال 1441 هـ / يونيو 2020م

الترقيم الدولي المعياري للدوريات 1996-2339

يونيو 2020م **مجلة جامعة الشارقة للعلوم الإنسانية والاجتماعية** المجلد 17 العدد 1 (A)
https//:doi.org/10.36394/jhss/17/1A/13

طريقة مطورة لتفسير أنماط الاستجابات غير المطابقة للفرد في الاختبارات ومقارنة خمسة مؤشرات مطابقة الفرد

راشد سيف المحرزي[1]

أمل ثابت عودة[2]

تاريخ الاستلام: 11-12-2018 **تاريخ القبول: 28-03-2019**

ملخص البحث:

هدفت الدراسة الحالية إلى الكشف عن أنماط الاستجابة غير المطابقة باستخدام خمسة مؤشرات لمطابقة الفرد (المحرزي الموزون ، المحرزي غير الموزون ، درازجو ، رايت الموزون ، رايت غير الموزون) باستخدام بيانات اختبار القدرة العددية في مقياس الخليج للقدرات العقلية المتعددة. وقد بلغت عينة الدراسة (4206) طلاب من جميع دول الخليج. واستخدمت الدراسة الأساليب الإحصائية المناسبة للتحقق من افتراضات النموذج ثلاثي المعالم، وملاءمة مفردات اختبار القدرة العددية للنموذج. كما تم التحقق من صدق نتائج المؤشرات في تصنيف استجابات الأفراد من حيث مطابقتها أو عدم مطابقتها للنموذج في ضوء تفسيرات المحكمين. وقد أشارت نتائج الدراسة إلى أن المؤشرين WSR و Lz كانا أكثر المؤشرات صدقًا ودقة. وتم تصنيف استجابات الطلاب غير المطابقة إلى ستة أنماط، هي: (التكاسل، والتباطؤ، والإبداع الاستثنائي، والغش، والتخمين، والخطأ في نقل الإجابة). وجاء الغش أكثر الأنماط انتشارًا، وتمت مناقشة النتائج في ضوء الإطار النظري والدراسات السابقة، وتقديم بعض التوصيات والمقترحات المتعلقة بموضوع الدراسة.

الكلمات الدالة: مؤشرات مطابقة الفرد، أنماط الاستجابة غير المطابقة، نظرية الاستجابة للمفردة، مقياس الخليج للقدرات العقلية المتعددة.

(1) كلية التربية - جامعة السلطان قابوس (مسقط - سلطنة عمان)
mehrzi@squ.edu.om

(2) وزارة التربية والتعليم (مسقط - سلطنة عمان)

المقدمة:

يولي واضعو الاختبارات ومستخدموها اهتمامًا متزايدًا بتطبيق أفضل الطرق للحصول على معلومات دقيقة عن المفحوصين، مما لدقة هذه المعلومات من تأثير في تعزيز صدق نتائج الاختبار وصدق تفسيراتها. ومن بين الطرق الشائعة في الحصول على معلومات دقيقة عن أداء الفرد في الاختبار التركيز على أنماط استجابته عبر المفردات داخل الاختبار بدلاً من الاكتفاء بالأداء الكلي في الاختبار. فعلى سبيل المثال، تساعد دراسة أنماط استجابات الفرد عبر مفردات الاختبار المكون من مفردات الاختيار من متعدد في الحكم على ما إذا كانت درجة الفرد في الاختبار نتيجة لامتلاكه للمتطلبات اللازمة للمحتوى المقاس في الاختبار، أو نتيجة لاستخدمه تقنيات مختلفة في الاستجابة ليست لها علاقة بمحتوى مادة الاختبار، مثل التخمين العشوائي أو الغش. كما قد تساعد دراسة أنماط استجابات الفرد عبر مفردات الاختبار في الكشف عن درجة تأثير دافعية الفرد خلال تقديم الاختبار على أدائه ونتيجته. وتزيد دراسة أنماط استجابات الفرد من لتعزيز صدق نتائج الطلبة في الاختبار.

وتقدم مؤشرات مطابقة الفرد Person Fit Statistic (PFS) خدمات كبيرة في مجال الاستفادة من دراسة أنماط استجابات (Response Pattern) الأفراد في تفسير أداء الفرد في السمة المقاسة؛ إذ إنها تساعد في تصنيف أنماط استجابة الفرد بأنها مطابقة أو غير مطابقة، من أجل تحقيق الصدق والعدالة في استخدام وتفسير علامات الاختبار. وتساعد هذه المؤشرات في اتخاذ القرار المناسب حول مطابقة نمط استجابات الفرد من خلال فحص صحة فرضية صفرية حول مطابقة استجابة الفرد. فإذا تم تصنيف نمط الاستجابة على أنها مخالفة/غير مطابقة (Aberrant)، فيعد ذلك مؤشرا على عدم مصداقية قياس درجة الفرد في الاختبار لمستوى الفرد في السمة المقاسة بالاختبار، أو اعتبار الاختبار المستخدم غير ملائم لقياس أداء الفرد، أو يعتبر الاختبار ضعيفا في قياس قدرة المفحوص بدقة (Deng & Torre, 2008).

وتتوفر العديد من مؤشرات مطابقة الفرد في الأدبيات التربوية والنفسية (Tendeiro & Meijer, 2014; Tendeiro, Meijer & Niessen, 2016). ويمكن تقسيم مؤشرات مطابقة الفرد إلى قسمين: المؤشرات القائمة على أداء المجموعة (group-based person fit statistics) والمؤشرات القائمة على النماذج (model-based person fit statistics). وتقارن المؤشرات القائمة على أداء المجموعة نمط استجابة الفرد بالتوقعات القائمة من أداء المجموعة التي تقدم الاختبار. فإذا توافق نمط استجابة الفرد مع توقعات أداء المجموعة الاختبارية دل ذلك على مطابقة نمط استجابة الفرد. ويعد مؤشرات مطابقة الفرد لهارنيستش وتاتسوكا (Harnisch & Tatsouka, 1983) ومؤشر درازجو (Drasgow, Levine, & William, 1985) ومؤشرات المحرزي (Almehrizi, 2010) من مؤشرات مطابقة الفرد القائمة على النماذج.

بينما تقارن المؤشرات القائمة على النماذج نمط استجابة الفرد بالتوقعات القائمة على أحد النماذج الرياضية في نظرية الاستجابة على المفردة (Item Response Theory). فإذا توافق نمط استجابة الفرد مع توقعات النموذج الرياضي دل ذلك على مطابقة نمط استجابة الفرد. ويعد مؤشرات مطابقة الفرد لرايت (Wright & Stone, 1979) ومؤشر درازجو (Drasgow, Levine, & William, 1985) ومؤشرات المحرزي (Almehrizi, 2010) من مؤشرات مطابقة الفرد القائمة على النماذج.

وتعتبر الأسباب وراء ظهور أنماط مخالفة في استجابات الفرد عوامل دخيلة على القدرة المقاسة وتؤثر بصورة منتظمة ودالة على أداء الفرد، وتؤدي إلى قياس غير دقيق للقدرة. ويذكر حمادنة (2011) عددًا من هذه الأسباب المتعلقة بخصائص الفرد نفسه مثل الغش (قيام المفحوص ذي القدرة المتدنية بنقل الإجابة الصحيحة على مفردة صعبة من مفحوص ذي قدرة عالية أو يستخدم مصادر غير مسموح بها في الاختبار)، وقلق الاختبار (الحالة التي يكون فيها المفحوص في حالة اضطراب نتيجة الخوف من الفشل والذي قد يؤدي إلى تقليل التركيز أثناء تأدية الاختبار مما يجعل المفحوص يجيب على المفردات الصعبة إجابة صحيحة، ويفقد الإجابة الصحيحة على المفردات السهلة)، وفقدان الدافعية، والتباطؤ في أداء الاختبار (الحالة التي يكون فيها أداء المفحوص بطيئا أثناء تأدية الاختبار ولا ينتقل إلى المفردة التالية حتى يدقق بأقصى درجة ممكنة على صحة إجابته على المفردة السابقة، مما يفقده الوقت الكافي بنهاية الاختبار للإجابة الصحيحة حتى عن المفدرات السهلة).

وناقشت العديد من الدراسات في مؤشرات مطابقة الفرد منذ بداية الاهتمام بمثل هذه الدراسات أسبابا أو أنماطا أخرى لسلوك الاستجابة غير المطابقة. فناقش رايت وستون (Wright & Stone, 1979) سلوك التكاسل (الحالة التي يواجه فيها المفحوص مجموعة من المشاكل في بداية الاختبار، ويفقد التركيز في بداية الاختبار وعندما ينتقل الى المفردات المتبقية في الاختبار ويتكيف مع الاختبار لا يقوم بمراجعة إجاباته على المفردات في بداية الاختبار)، والتخمين (الإجابة الصحيحة عن مفردة صعبة بشكل عشوائي من قبل المفحوصين ذوي القدرة المتدنية). وناقش هولين ودرازجو وبارسونس (Hulin, Drasgow & Parsons, 1983) سلوك الخطأ في نقل الإجابة إلى ورقة الإجابة في الاختبار من نوع اختيار من متعدد.

وأضاف كاراباتسوس (Karabatsos, 1984; 1996; 2003) نمط الإبداع الاستثنائي وهو قيام المفحوص بتفسير المفردات السهلة بطريقة إبداعية غير تقليدية مما يجعله يجيب عن المفردات السهلة بشكل خاطئ، وسوء فهم المفردة (الإجابة عن المفردة بشكل خاطئ بناء على معلومات خاطئة مكتسبة). وأضاف الجراح (2009) المرض كأحد أسباب ظهور أنماط استجابة غير مطابقة فالمرض قد يؤدي إلى تقليل التركيز أثناء تأدية الاختبار مما قد

يؤدي بالمفحوص أن يجيب عن المفردات السهلة بشكل خاطئ. واعتبر (Schmitt, Chan, Sacco McFarland & Jennings, 1999) الخصائص الديموغرافية للفرد، مثل: الجنس، والعرق، والتحيز الثقافي أسبابا تؤدي أيضا إلى توليد أنماط إجابة شاذة.

وأضافت عدد من الدراسات (Harnisch & Tatsouka, 1983; Harnisch & Linn, 1981)، الاختلاف في أساليب التدريس كسبب لظهور أنماط استجابة غير مطابقة وذلك من خلال اكتساب الطلبة لمعلومات خاطئة حول موضوع معين بسبب عدم استخدام الطريقة المناسبة لتدريس موضوع الدرس أو عدم إعطاء الطلبة الفرصة المناسبة للتدرب. كما أن دخول عامل سرعة الاختبار قد يؤدي إلى حدوث أنماط غير مطابقة بأن يقوم المفحوص بالإجابة عن المفردات بشكل عشوائي بحيث يمكن أن يجيب عن المفردات الصعبة وبينما لا يجيب عن المفردات السهلة (Meijer, 1996).

واعتبر ميجر (Meijer, 1996) أن تعدد القدرات التي يقيسها الاختبار (تعدد الأبعاد) يمكن أن يكون سببا لظهور استجابات غير مطابقة، فإذا كان المقياس يتضمن قدرتين، وتوزعت المفردات على القدرتين بحيث كانت أغلب المفردات السهلة تقيس القدرة الأولى، وأغلب المفردات الصعبة تقيس القدرة الثانية، فإن استجابة الفرد المتمكن في القدرة الثانية استجابة صحيحة على المفردات الصعبة قد تعد نمطا غير مطابق.

وقد فحصت العديد من الدراسات السابقة أداء مؤشرات مطابقة الفرد باستخدام بيانات مولدة وبيانات حقيقية. فقام دودين (Dodeen, 2003) باستخدام مؤشر مطابقة الفرد لدرازجو لتحليل اختبارين أحدهما يقيس مستوى الطلاب في اللغة العربية (Arabic Placement Test; APT) والآخر يقيس مستوى الطلاب في اللغة الإنجليزية (English placement Test; EPT) لتحديد قبول الطلبة في إحدى كليات جامعة الإمارات العربية المتحدة وفق النموذج الثلاثي المعلمة . وقد توصلت الدراسة إلى أن التوزيع التكراري لمؤشر لكلا الاختبارين كان قريبا جدا من التوزيع الطبيعي المعياري؛ إذ كان المتوسط الحسابي للمؤشر 0.17، والانحراف المعياري 0.99، لاختبار اللغة الإنجليزية، وكان المتوسط الحسابي للمؤشر 0.02، والانحراف المعياري 0.92 لاختبار اللغة العربية. وأظهرت النتائج أن نسبة عدم المطابقة للأفراد كانت قليلة جدًا؛ إذ كانت النسبة 4.8%، في اختبار اللغة الإنجليزية، و3.1% في اختبار اللغة العربية. كما أظهرت الدراسة عدم وجود ارتباط إحصائي بين مؤشر مطابقة الفرد ومستوى القدرة؛ إذ كان معامل الارتباط مساويا 0.002، لاختبار اللغة الإنجليزية، و0.13 لاختبار اللغة العربية.

وقارن جراح (2009) بين ثلاثة مؤشرات لمطابقة الفرد: المؤشر (WT) للمطابقة الداخلية لرايت وماسترز (Wright & Masters, 1982) والمؤشر (L_Z)، والمؤشر ($\zeta 2$)

لتاتسوكا (Tatsouka, 1996)، وقد استخدم بيانات اختبار تمس الدولي (TIMSS, 2003) الدولي لمبحث الرياضيات لعينة من لطلبة الأردنيين . أظهرت نتائج التحليل حول نسب الكشف عن الأنماط غير المطابقة عبر نماذج الاستجابة للمفردة وعبر أطوال الاختبارات المختلفة أن المؤشر ($\zeta 2$) كشف عن أعلى نسبة من الأنماط غير المطابقة ثم مؤشر (L_Z)، ثم المؤشر (WT) . كان ترتيب المؤشرات حسب قوتها في تحديد الأنماط غير المطابقة بناء على تحقق خاصية عدم التأثر بالقدرة ضمن النموذج الأحادي والثنائي وعبر الأطوال المختلفة للاختبار هو ($\zeta 2$) ثم (L_Z) ثم (WT). وأظهر المؤشر ($\zeta 2$) ميلا نحو خصائص التوزيع الطبيعي عبر مستويات القدرة كافة وعبر نماذج استجابة المفردة وعبر الأطوال المختلفة للاختبارات. بينما أظهر المؤشران (L_Z, WT) (تشابها من حيث إن لهما التواءً سالبًا وتفلطحًا كبيرًا عبر المستويات المختلفة من القدرة وعبر نماذج استجابة المفردة الأحادي والثنائي والثلاثي وعبر الأطوال المختلفة للاختبارات على الرغم من اختلاف الأساس الرياضي لها. وقد أوصت الدراسة باستخدام المؤشر ($\zeta 2$) مع النماذج الثلاثة لاستجابة المفردة واستخدام المؤشر (L_Z) مع النموذج ثلاثي المعالم.

مشكلة الدراسة:

تستخدم مؤشرات مطابقة الفرد في تقديم أدلة على دقة وصدق تقييم الاختبار للسمة المقاسة، من خلال تحليل أنماط استجابات الأفراد. وتختلف مؤشرات مطابقة الفرد فيما بينها في قدرتها على كشف جميع أنماط الاستجابات غير المطابقة للفرد عند استخدامها مع نفس عينة الأفراد في الاختبار الواحد. وتختلف أيضا في دقتها في الكشف الصحيح لأنماط استجابات الأفراد، فمن الخطأ التوقع بأن المؤشر الذي يظهر أكبر نسبة من الأفراد ذوي استجابات غير مطابقة على أنه المؤشر الأفضل. فمن الضروري التأكد من صدق تصنيف كل مؤشر من مؤشرات مطابقة الفرد باستخدام محك خارجي لتحديد المؤشر الأكثر دقة في تصنيف الأفراد ذوي الاستجابات غير المطابقة. فالتصنيف الثنائي لنمط استجابة الفرد (مطابق/غير مطابق) يحتمل الوقوع في خطأين يتم الحرص على تجنبهما أو التقليل منهما لتفادي النتائج المترتبة على الوقوع بأي منها. ويتمثل الخطأ الأول في تصنيف النمط المطابق لاستجابات الفرد على أنه غير مطابق، ويسمى الخطأ السالب. ويتمثل الخطأ الثاني في تصنيف النمط غير المطابق لاستجابات الفرد على أنه مطابق ويسمى الخطأ الموجب. ويزيد خفض الخطأين من صحة قرارات تصنيف مطابقة استجابات الفرد.

وما يعاب على أغلب مؤشرات مطابقة الفرد العامة أنها تكتفي بتحديد مدى مطابقة استجابات الأفراد مع الاستجابات المتوقعة وفقا للنماذج المستخدمة في نظرية الاستجابة للمفردة والكشف عن أنماط الاستجابة غير المطابقة. بينما يغلب على أغلبها عدم القدرة

على الكشف عن أسباب ظهور هذه الأنماط غير المطابقة (جراح، 2009). فهنالك العديد من أنماط الاستجابة الشاذة أو غير المطابقة التي يمكن أن تكشف عنه مؤشرات مطابقة الفرد، إلا أنه قد لا يمكن تفسير نمط استجابة الفرد الذي تكشف عنه مؤشرات مطابقة الفرد. وحقيقية الأمر تبقى عملية تحديد أسباب حدوث الأنماط غير المطابقة من الأمور الصعبة وذلك لتدخل عدة عوامل تؤثر في إجابة المفحوص عن مفردات الاختبار (Meijer, 1996). وتتطلب عملية تفسير نمط الاستجابة غير المطابقة توفر مؤشرات مطابقة الفرد قادرة على تقديم تفسيرات من خلال ربط الاستجابات على المفردات مع خصائص المفردات. كما يساعد الاستعانة بمحكمين على محاولة تفسير استجابات الفرد الذي تكشف عنه مؤشرات مطابقة الفرد من خلال السمة التي يقيسها الاختبار.

وتهدف الدراسة الحالية إلى تطبيق أسلوب مطور لتفسير أنماط الاستجابة غير المطابقة التي تكشف عنها مؤشرات مطابقة الفرد الخمسة باستخدام بيانات حقيقية في اختبار القدرة العددية المستوى الأول لمقياس الخليج للقدرات العقلية المتعددة للعام 2010، للمفاضلة بين صدق الأحكام التي يتم تصديرها من هذه المؤشرات الخمسة. كما تهدف الدراسة إلى الكشف عن أكثر أنماط الاستجابة غير المطابقة شيوعا في اختبار القدرة العددية وإلى المفاضلة بين مؤشرات مطابقة الفرد الخمسة في قدرتها على كشف أنماط استجابة الفرد وتوافقها مع تصنيف المحكمين. وتتلخص مشكلة الدراسة في المحاولة للإجابة عن السؤالين التاليين:

1. ما مدى صدق تصنيف مؤشرات مطابقة الفرد الخمسة في الكشف عن الأفراد ذوي الاستجابات غير المطابقة في اختبار القدرة العددية لمقياس الخليج للقدرات العقلية المتعددة؟
2. ما أنماط الاستجابة غير المطابقة التي كشف عنها كل مؤشر من مؤشرات مطابقة الفرد الخمسة في اختبار القدرة العددية لمقياس الخليج القدرات العقلية المتعددة؟

أهمية الدراسة:

اكتسب هذا النوع من الدراسات أهمية كبيرة من الناحية النظرية عند العديد من الباحثين في مجال القياس والتقويم وتحديدًا في دراسة مؤشرات مطابقة الفرد، ونذكر من هذه الدراسات على سبيل المثال دراسة جراح (2009) والتي هدفت إلى مقارنة ثلاثة مؤشرات لمطابقة الفرد، في حين جاءت الدراسة الحالية لتقارن بين خمسة مؤشرات لمطابقة الفرد (مؤشري رايت، مؤشر درازجو، مؤشري المحرزي) من حيث فعاليتها في الكشف عن الأفراد ذوي الاستجابات غير المطابقة، وهذا من شأنه أن يساعد القائمين على

عملية التقويم في تحديد المؤشر الأفضل في الكشف عن أنماط الاستجابة غير المطابقة. وكما تكتسب الدراسة أهميتها من الناحية العملية من خلال استخدام أسلوب مطور وسهل لتحكيم أنماط الاستجابة غير المطابقة والحكم على فعالية ومصداقية نتائج المؤشرات، وهو أمر ذو أهمية كبيرة يحتاجه القائمون على عملية التقويم التربوي، مما يمكن متخذي القرار لاتخاذ القرارات المناسبة فيما يتعلق بنتائج هؤلاء الأفراد.

مصطلحات الدراسة:

مؤشر مطابقة الفرد: مؤشر إحصائي يحدد المدى أو البعد بين البيانات الفعلية (استجابة المفحوصين) والقيم المتوقعة من خلال النموذج المستخدم، وتقارن قيم مؤشر مطابقة الفرد مع قيمة حرجة واحدة لتحديد أنماط الإجابة المطابقة وأنماط الإجابة غير المطابقة (Karabatsos, 2003).

نماذج الاستجابة أحادية البعد ثنائية الاستجابة (Dichotomous Univariate IRT Models): نماذج احصائية تستخدم في تحليل المفردات التي يجيب المفحوص عنها إجابة صحيحة أو إجابة خطأ. فقد يتم تدريج المفردات في ضوء مستوى صعوبتها، كما في النموذج أحادي المعالم أو نموذج راش اللوغاريتمي، وقد يتم تدريج المفردات في ضوء مستوى صعوبتها وقدرتها على التمييز بين المستويات المختلفة من القدرة، كما في النموذج ثنائي المعالم، وقد يتم هذا التدريج في ضوء مستوى صعوبة المفردة وقدرتها على التمييز، ومستوى التخمين، كما في النموذج ثلاثي المعالم (علام، 2005).

حدود الدراسة:

- تتحدد الدراسة ببيانات العينة المستخدمة باختبار القدرة العددية من مقياس الخليج للقدرات العقلية المتعدد (جماس 2010).
- تتحدد الدراسة باستخدام النموذج ثلاثي المعالم من نماذج نظرية الاستجابة للمفردة.
- تعتمد دقة نتائج الدراسة على دقة البرمجيات المستخدمة في تقدير معالم الفقرات والمفحوصين والبرمجيات المستخدمة في تقدير مؤشرات الكشف عن عدم المطابقة.
- تقتصر الدراسة الحالية على ستة أنماط لاستجابات الفرد غير المطابقة (التخمين، والتكاسل، التباطؤ، والغش، والإبداع الاستثنائي، والخطأ في نقل الإجابة).
- تتحدد نتائج الدراسة بصدق تقديرات المحكمين في تصنيف أنماط الاستجابات غير المطابقة للفرد.

الطريقة والاجراءات:

منهج الدراسة:

استخدمت الدراسة المنهج الوصفي في تحقيق أهداف هذه الدراسة في المقارنة بين مؤشرات مطابقة الفرد وقدرتها على كشف أنماط الاستجابة غير المطابقة، باستخدام النموذج ثلاثي المعالم في نظرية الاستجابة للمفردة وباستخدام بيانات أرشيفية من اختبار القدرة العددية في مقياس الخليج للقدرات العقلية المتعددة لعام 2010.

مجتمع الدراسة:

تكون مجتمع الدراسة من جميع الطلاب والطالبات المسجلين بمدارس التعليم العام الحكومية من الصف الأول الابتدائي والثاني الابتدائي بالعام الدراسي 2009 / 2010 في دول مجلس التعاون الخليجي الست: الإمارات العربية المتحدة، والبحرين، والمملكة العربية السعودية، وسلطنة عمان، وقطر، والكويت (الزيات وآخرون، 2011).

عينة الدراسة:

تكونت عينة الدراسة من بيانات أرشيفية لطلبة الصف الأول الابتدائي والثاني الابتدائي الذين تقدموا لأداء المستوى الأول لاختبار القدرة العددية في مقياس الخليج للقدرات العقلية المتعددة. ويوضح جدول (1) توزيع عينة الدراسة الحالية حسب الصف الدراسي والجنس. إذ بلغ عدد طلبة الصف الأول الابتدائي (1055) من الذكور و(1088)من الإناث، في دول الخليج العربي مجتمعه. كما بلغ عدد طلبة الصف الثاني الابتدائي (1070)من الذكور و(993) من الإناث، وكان عدد الطلبة المشاركين من مملكة البحرين (780)، ومن السعودية (663)، ومن دولة الكويت (705)، ومن قطر (684)، ومن الإمارات (715)، ومن سلطنة عمان (659)، وبلغ حجم عينة الدراسة (4206) طالبا وطالبة ممن طبق عليهم اختبار القدرة العددية.

جدول (1): توزيع عينة الدراسة لمقياس الخليج للقدرة العددية المستوى الأول وفق الصف والجنس بدول الخليج.

الدولة	الصف الأول		الصف الثاني		المجموع
	ذكر	أنثى	ذكر	أنثى	
البحرين	194	200	191	195	780
السعودية	214	200	131	118	663

الكويت	156	204	162	183	705
قطر	143	167	182	192	684
الإمارات	145	159	234	177	715
عمان	203	158	170	128	659
المجموع	1055	1088	1070	993	4206

(الزيات وآخرون، 2011)

أداة الدراسة:

استخدمت الدراسة بيانات أرشيفية من اختبار القدرة العددية المستوى الأول لمقياس الخليج للقدرات العقلية المتعددة بدول الخليج، والذي تم إعداده وتقنينه من قبل فريق من الباحثين (الزيات وآخرون، 2011). ويتكون الاختبار من 30 مفردة، تتوزع على اختبارين فرعيين: اختبار العد (15 مفردة)، واختبار حل المسائل الحسابية (15 مفردة). وقد روعي في بناء واختيار مفردات الاختبار توظيف عمليتي الإدراك والاستدلال. ويتكون الاختبار من كراسة أسئلة وورقة إجابة. ويغطي المستوى الأول من اختبار القدرة العددية الصفين الأول والثاني الابتدائي، بحيث يمتد المدى العمري من 5 سنوات وتسعة شهور إلى سبع سنوات وثمانية شهور (5:9 - 7:8)، ويتم تطبيق الاختبار بشكل فردي (الزيات وآخرون، 2011).

صدق وثبات اختبار القدرة العددية:

نظرا لاستخدام الدراسة لبيانات أرشيفية لمقياس الخليج للقدرات العقلية فقد تم استعراض مؤشرات الصدق والثبات التي ذكرت في الدليل الفني للمقياس. جاء معامل الارتباط بين الاختبارين الفرعيين في اختبار القدرة العددية للمستوى الأول 0,426، وهو ما يشير إلى ارتباط الاختبارات الفرعية في داخل القدرة العددية. كما تم تحليل الارتباطات البينية للاختبارات الفرعية للقدرة العددية تحليلا عاملياً باستخدام التحليل العاملي التوكيدي من الدرجة الثانية باستخدام برنامج Lisrel 8.53. وقد جاءت قيم التشبعات عالية للاختبارات الفرعية مع القدرة العددية التي تنتمي اليها، وقد جاء اختبار العد بتشبع مقداره 0,64 مع القدرة العددية واختبار المسائل الحسابية بتشبع مقداره 0,67 مع القدرة العددية في المستوى الأول للمقياس (الزيات وآخرون، 2011). كما بلغت قيم معاملات ثبات الفا كرونباخ للدرجات الخام للقدرة العددية بين 0.823 - 0.869 (الزيات وآخرون، 2011).

مؤشرات مطابقة الفرد الخمسة:

تستخدم الدراسة خمسة مؤشرات مطابقة الفرد مبنية على نظرية استجابة المفردة:

أ. مؤشري لرايت وستون (Wright & Stone, 1979):

استخدم رايت (Wright, 1977) الدرجة لاقتراح نسختين من مؤشر متوسط المربعات: وهي مؤشر مربع مطابقة الفرد الموزون وغير الموزون، وهما

1) مؤشر مربع مطابقة الفرد غير الموزون:

$$UMS_j = \frac{1}{n-1}\sum_{i=1}^{n} Z_{ij}^2 = \frac{1}{n-1}\sum_{i=1}^{n} \frac{(y_{ij}-p_{ij})^2}{p_{ij}q_{ij}} \qquad (1)$$

2) مؤشر مربع مطابقة الفرد الموزون:

$$WMS_j = \frac{\sum_{i=1}^{n}(p_{ij}q_{ij}){z_{ij}}^2}{\sum_{i=1}^{n} p_{ij}q_{ij}} = \frac{\sum_{i=1}^{n}(y_{ij}-p_{ij})^2}{\sum_{i=1}^{n} p_{ij}q_{ij}} \qquad (2)$$

واقترح كل من رايت وستون (Wright & Stone, 1979) تحويلا لكل من مؤشري مربع المتوسط الموزون وغير الموزون لإزالة أثر حجم العينة كالتالي:

$$UT = (In(UMS_j) + UMS_j - 1)\sqrt{\frac{n-1}{8}} \qquad (3)$$

$$WT = \left(\sqrt[3]{WMS_j} - 1\right)\left({}^{3}/_{r}\right) + \left({}^{r}/_{3}\right) \qquad (4)$$

حيث r: هو الانحراف المعياري لمؤشر

$$r = \frac{\sqrt{\sum_{i=1}^{n} p_{ij}q_{ij}(p_{ij}-q_{ij})^2}}{\sum_{i=1}^{n} p_{ij}q_{ij}} \qquad (5)$$

ب. مؤشر لوغاريتم الارجحية المعياري Standardized Log-Likelihood. Statistic L_Z :

قام درازجو وليفين ووليم (Drasgow, Levine & William, 1985) بتطوير صورة معيارية للمؤشر اعتمادًا على الوسط الحسابي والتباين للمؤشر L_Z . ويفترض ان المؤشر L_Z أقل اعتمادًا على مستويات القدرة، وله توزيع طبيعي معياري عبر مستويات مختلفة

مـن القـدرة، ويعطـى مؤشـر لوغاريتـم الأرجحيـة المعيـاري بالمعادلـة:

$$L_z = \frac{L_o - E(L_o)}{\sqrt{Var(L_o)}} \quad (6)$$

$$E(L_o) = \sum_{i=1}^{n} p_{ij}(\Theta) Inp_{ij}(\Theta) + q_{ij}(\Theta) In(q_{ij}(\Theta)) \quad (7)$$

$$Var(L_o) = \sum_{i=1}^{n} p_{ij}(\Theta) q_{ij}(\Theta) \left[In\left(\frac{p_{ij}(\Theta)}{q_{ij}(\Theta)}\right) \right]^2 \quad (8)$$

ج. مؤشري المحرزي Lz لمطابقة الفرد (Almehrizi, 2010):

يوظـف مؤشـر المحـرزي لمطابقـة الفـرد منهـج البواقـي لتقييـم مطابقـة اسـتجابة الفـرد. وقـد جـاءت نسـختان لمؤشـر المحـرزي المعتمـد علـى البواقـي لمطابقـة الفـرد المـوزون وغيـر المـوزون:

1) مؤشر مطابقة الفرد للمحرزي غير الموزون:

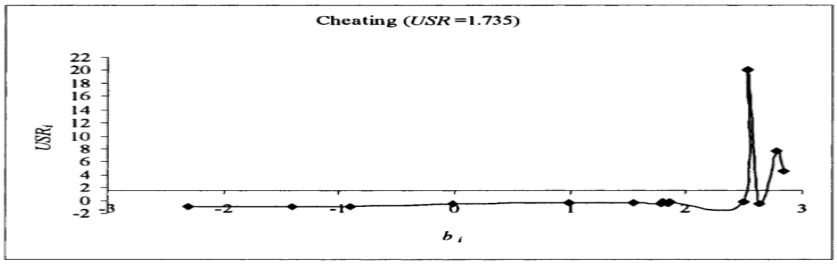
$$USR = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{(p_{ij} - y_{ij})(p_{ij} - q_{ij})}{\sqrt{p_{ij} q_{ij} (p_{ij} - q_{ij})^2}} \quad (9)$$

2) مؤشر مطابقة الفرد المحرزي الموزون

$$WSR = \frac{\sum_{i=1}^{n} (p_{ij} - y_{ij})(p_{ij} - q_{ij})}{\sqrt{\sum_{i=1}^{n} p_{ij} q_{ij} (p_{ij} - q_{ij})^2}} \quad (10)$$

أسلوب مقترح لتفسير أنماط الاستجابة غير المطابقة:

تقـدم الدراسـة الحاليـة أسـلوبًا لتفسـير أنمـاط الاسـتجابة غيـر المطابقـة التـي تكشـف عنهـا مؤشـرات مطابقـة الفـرد مـن خـلال تطويـر طريقـة توظيـف مؤشـر مطابقـة الفـرد علـى مسـتوى المفـردة الواحـدة الـذي قدمـه المحـرزي (Almehrizi, 2010) والتـي تقـوم علـى تصميـم رسـم انتشـاري لقيمـة هـذا المؤشـر عبـر مفـردات الاختبـار مـن أجـل الحكـم علـى صـدق المؤشـرات وتفسـير أنمـاط الاسـتجابة غيـر المطابقـة كمـا فـي شـكل (1).

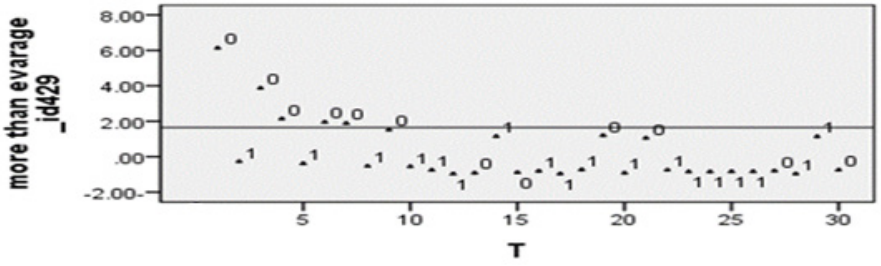


شكل (1): رسم انتشاري لقيمة مؤشر مطابقة الفرد للمحرزي للمفردة الواحدة عبر مفردات الاختبار لطالب ذي قدرة =1- (almehrizi, 2010).

ويعرض هذا الشكل علاقة بين معلمة صعوبة المفردات وقيمة مؤشر USR_{ij} (Almehrizi, 2010) التالية:

$$USR_{ij} = \frac{(y_{ij}-p_{ij})^2-p_{ij}q_{ij}}{\sqrt{p_{ij}q_{ij}(p_{ij}-q_{ij})^2}}, \qquad i = 1،2، \dots ،n, j = 1،2، \dots ،N \qquad (11)$$

وما يعاب على هذه الطريقة عدم الاحتفاظ بترتيب المفردات في الورقة الامتحانية وعدم معرفة استجابة الطالب للمفردات والتي لها أهمية كبيرة في فهم استجابات الطالب والمساعدة في تفسير استجاباته. وتقدم هذه الدراسة تطويرا لطريقة عرض المعلومات عن استجابات الطالب على المفردات في رسم واحد يحتوي على معلومات منظمة تساعد في فهم وتفسير استجابات الطالب. ويظهر الشكل (2) مكونات الرسم الانتشاري بعد التطوير لمؤشر مطابقة الفرد على مستوى المفردة الواحدة لاستجابات فرد واحد في 30 مفردة.



شكل (2): شكل انتشاري لقيم مؤشر مطابقة الفرد على مستوى المفردة الواحدة للمحرزي (Almehrizi, 2010) لاستجابات طالب على 30 مفردة

يتكون الشكل الانتشاري من المعلومات التالية:

1. المحور السيني: ويوضع فيه أرقام مفردات الاختبار كما ورد ترتيبها في الورقة الاختبارية وعددها 30 مفردة.
2. المحور الصادي: ويوضع فيه تدريج لقيم مؤشر مطابقة الفرد على مستوى المفردة الواحدة للمحرزي (Almehrizi, 2010).
3. نقاط في وسط الشكل: تعبر كل نقطة عن القيمة المحسوبة لمؤشر مطابقة الفرد على مستوى المفردة الواحدة عبر مفردات الاختبار.
4. الرقم (0 / 1): يتم وضع الرقم 0 أو 1 عند كل نقطة بحيث تمثل الدرجة التي حصل عليها الطالب على إجابته في المفردة (1) إجابة صحيحة، (0) إجابة خاطئة.
5. قدرة الطالب: يتم وضع وصف لقدرة الفرد في الاختبار على يسار المحور الصادي، ويمكن استخدام التقسيم السداسي لقدرة الطالب كما يلي: منخفضة جدا (-3 > $\boldsymbol{\Theta}$ > -2)، قدرة منخفضة (-2 > $\boldsymbol{\Theta}$ > -1)، قدرة أقل من المتوسط (-1 > $\boldsymbol{\Theta}$ > 0)، قدرة أكبر من المتوسط (0 > $\boldsymbol{\Theta}$ > 1)، قدرة مرتفعة (1 > $\boldsymbol{\Theta}$ > 2)، قدرة مرتفعة جدا (2 > $\boldsymbol{\Theta}$ > 3).
6. خط أفقي عند القيمة 1.645 في المحور الصادي: يتم وضع خط أفقي موزاي للمحور السيني عند القيمة 1.645 في المحور الصادي وتمثل الدرجة الزائية المقابلة لمساحة 0.05 تحت منحنى التوزيع الطبيعي الاعتدالي، وتمثل القيمة الحرجة للحكم على مطابقة استجابة الفرد على كل مفردة.

ويتم قراءة الشكل لتفسير أنماط استجابة الفرد من خلال التركيز على موقع النقاط التي تمثل قيمة مؤشر مطابقة الفرد بالنسبة إلى الخط الأفقي للقيمة المرجعية مع الأخذ بعين الاعتبار الدرجة التي حصل عليها الطالب في المفردة (0 / 1) وتسلسل رقم المفردة في الورقة الاختبارية. فتشير النقاط التي تمثل قيمة مؤشر مطابقة الفرد والتي تقع أعلى أو قريبة جدا من الخط الأفقي للقيمة المرجعية على عدم مطابقة استجابة الفرد على هذه المفردات (أي ان درجة الطالب على المفردة لا تتناسب مع صعوبة المفردة وقدرة الطالب). وفي المقابل تشير النقاط التي تقع أسفل وبعيدة عن الخط الأفقي للقيمة المرجعية على مطابقة استجابة الفرد على تلك المفردات. ويشير العدد الأكبر من النقاط التي تقع أعلى أو قريبة من الخط الأفقي للقيمة المرجعية إلى زيادة فرصة الحكم على نمط استجابة الفرد على أنها غير مطابقة.

ومن أجل تفسير عدم مطابقة نمط استجابة الفرد يتم التركيز على الدرجة 1 أو 0 عند كل نقطة للمفردات ذات الاستجابة غير المطابقة، بحيث يتم قراءتها كالتالي: تشير الدرجة 1 للمفردات ذات الاستجابة غير المطابقة إلى أن الطالب حصل على إجابة صحيحة على مفردة صعبة بالنسبة إلى قدرته وتشير الدرجة 0 للمفردات ذات الاستجابة غير المطابقة إلى أن الطالب حصل على إجابة خاطئة على مفردة سهلة بالنسبة إلى قدرته.

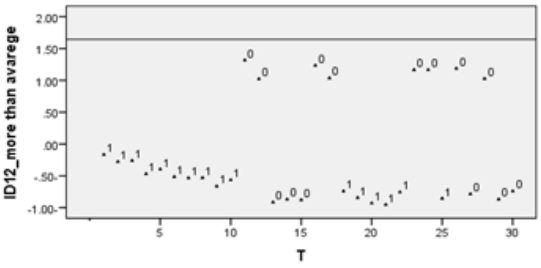
ويمكن توظيف مستوى قدرة الفرد الواردة في الشكل الانتشاري في تفسير أنماط الاستجابة غير المطابقة، فظهور نفس النمط عند الفرد يختلف تفسيره فيما إذا كانت قدرته عالية او متوسطة أو منخفضة.

من خلال دراسة الشكل الانتشاري لكل طالب من خلال الموجهات السابقة، يتم اصدار حكم على مدى مطابقة استجابات الفرد على مفردات الاختبار بالكامل لتحديد مطابقة استجابته وكذلك المساعدة في تفسير الاستجابات الملاحظة عند تصنيف عدم مطابقة استجابات الفرد. وفيما يلي توضيح لنمط الاستجابات المطابقة ولبعض أنماط الاستجابات غير المطابقة مع أمثلة عليها. وسيتم تقديم تعريف نظري للنمط ويعقبه تعريف اجرائي حسب الاسلوب المطور مع تصور للشكل الانتشاري الذي يمثل النمط.

أولاً: نمط الاستجابات المطابقة: توافق استجابات الطالب على مفردات الاختبار مع الاستجابة المتوقعة حسب مستوى صعوبتها في ضوء قدرة الفرد في الاختبار.

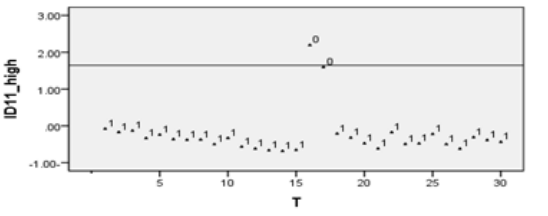
التعريف الاجرائي: تقع جميع أو معظم قيم مؤشر مطابقة الفرد على مستوى المفردة للمحرزي (Almehrizi, 2010) أسفل وبعيدة عن القيمة المرجعية (1.645).

مثال1: جميع قيم مؤشر مطابقة الفرد على مستوى المفردة للمحرزي (Almehrizi, 2010) للمفردات أصغر من القيمة المرجعية وكلها أسفل عن الخط الأفقي للقيمة المرجعية، وشكل (3) يوضح ذلك.

أ- مطابق ب- غير مطابق: • التكاسل • التباطؤ • الإبداع الاستثنائي • الغش • التخمين • خطأ في نقل الإجابة	

شكل (3): استجابة طالب مطابقة لصعوبة مفردات الاختبار

مثال2: أغلب قيم مؤشر مطابقة الفرد على مستوى المفردة للمحرزي (Almehrizi, 2010) للمفردات أصغر من القيمة الحرجة مع وجود عدد قليل جدًّا من المفردات تقع أعلى عن الخط الأفقي للقيمة الحرجة مع عدم إمكانية تفسير لها، وشكل (4) يوضح ذلك.

أ- مطابق ب- غير مطابق: • التكاسل • التباطؤ • الإبداع الاستثنائي • الغش • التخمين • خطأ في نقل الإجابة	

شكل (4): استجابة طالب مطابقة لصعوبة مفردات الاختبار

ثانياً الأنماط غير المطابقة:

تم تطبيق الأسلوب المطور لتفسير أنماط الاستجابة غير المطابقة على ستة أنماط مشمولة بالدراسة وهي (التكاسل، والتباطؤ، والإبداع الاستثنائي، والتخمين، والغش، والخطأ في نقل الإجابة).

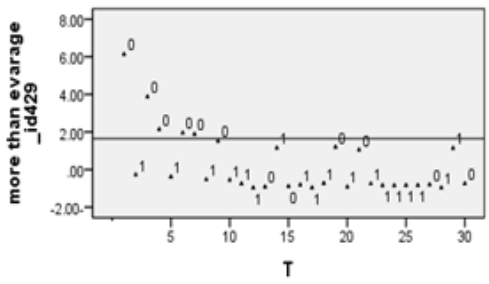
1. التكاسل: وهي الحالة التي يواجه فيها المفحوص ضعفًا في التركيز في بداية الاختبار عند الإجابة عن المفردات الأولى للاختبار وعندما يتحسن تركيزه في باقي الأسئلة لا يقوم بمراجعة تلك المفردات الأولى بسبب الكسل، مما يجعله يظهر استجابات خاطئة وغير متناسبة مع صعوبة هذه المفردات.

التعريف الاجرائي: وقوع قيم مؤشر مطابقة الفرد على مستوى المفردة للمحرزي (Almehrizi, 2010) أعلى من أو قريبه من القيمة المرجعية (1.645) لمجموعة من المفردات الأولى في الاختبار وتكون استجابته خاطئة على هذه المفردات وشكل (5) يوضح ذلك.

أ- مطابق

ب- غير مطابق:

- التكاسل
- التباطؤ
- الإبداع الاستثنائي
- الغش
- التخمين
- خطأ في نقل الإجابة

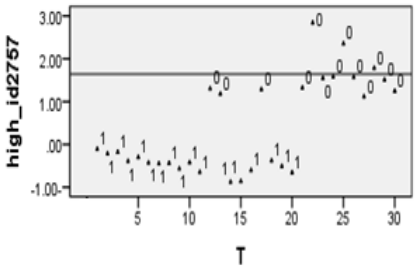


شكل (5): استجابة طالب ذي نمط «تكاسل»

يلاحظ من شكل (5) أن هذا الفرد ذو قدرة أكبر من المتوسط، ولكنه أجاب إجابة خاطئة على مجموعة من المفردات في بداية الاختبار، ثم تحسن أداؤه على بقية المفردات في الاختبار، ويظهر في الشكل استجابات غير مطابقة للمفردات الأولى للاختبار (الرقم 0 على يمين النقطة) أي أخطأ في الإجابة عن مفردات سهله ولم يقم بمراجعة حله بسبب الكسل.

2. التباطؤ في أداء الاختبار: وهي الحالة التي يكون فيها أداء الطالب بطيئا أثناء تأدية الاختبار ويرفض أن ينتقل إلى المفردة التالية حتى يتأكد بأقصى درجة ممكنة بأنه أجاب المفردة الحالية بشكل صحيح، مما يجعله يفقد وقت الاختبار قبل الوصول إلى الأسئلة الأخيرة، الأمر الذي يؤدي إلى تولد استجابات غير مطابقة على المفردات في نهاية الاختبار.

التعريف الإجرائي: وقوع قيم مؤشر مطابقة الفرد على مستوى المفردة للمحرزي (Almehrizi, 2010) أعلى من أو قريبه من القيمة المرجعية (1.645) للمفردات في نهاية الاختبار، وتكون إجاباته خاطئة على هذه المفردات وشكل (6) يوضح ذلك.

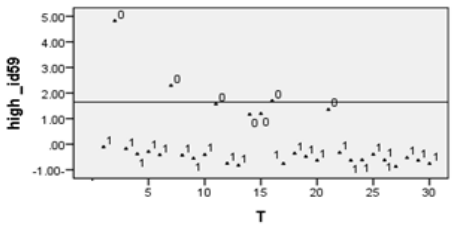
أ- مطابق ب- غير مطابق: • التكاسل • التباطؤ • الإبداع الاستثنائي • الغش • التخمين • خطأ في نقل الإجابة	

شكل (6): استجابة طالب ذي نمط «تباطؤ»

يلاحظ من شكل (6) أن هذا الفرد حريص في أدائه على مفردات الاختبار، فهو لا ينتقل إلى المفردة الأخرى إلا إذا تأكد من إجابته إجابه صحيحة على المفردة التي يقوم بحلها، ونتيجةً لهذا السلوك فإنه سيقوم بالإجابة فقط عن المفردات الأولى بشكل صحيح في الوقت المخصص للاختبار، في حين أن استجاباته على المفردات في نهاية الاختبار خاطئة. ويظهر في الشكل استجابات غير مطابقة للمفردات في نهاية للاختبار (الرقم 0 على يمين النقطة) أي أخطأ في الإجابة عن مفردات سهله ولم يتمكن من الإجابة عليها بشكل صحيح بسبب بطئه في الحل ضمن الوقت المخصص للاختبار.

3. الإبداع الاستثنائي: وهو قيام طالب ذي قدرة مرتفعة بتفسير المفردات السهلة بطريقة إبداعية غير تقليدية مما يجعله يجيب عن المفردات السهلة بشكل خاطئ.

التعريف الإجرائي: وقوع قيم مؤشر مطابقة الفرد على مستوى المفردة للمحرزي (Almehrizi, 2010) أعلى من أو قريبه من القيمة الحرجة (1.645) للمفردات السهلة وتكون إجابته خاطئة عليها أيا كان ترتيب هذه المفردات، لطالب ذي قدرة مرتفعة وشكل (7) يوضح ذلك.

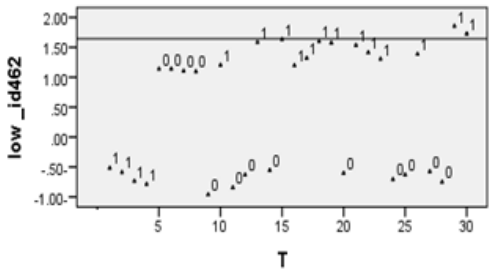
أ- مطابق ب- غير مطابق: • التكاسل • التباطؤ • الإبداع الاستثنائي • الغش • التخمين • خطأ في نقل الإجابة	

شكل (7): استجابة طالب ذي نمط «إبداع استثنائي»

يلاحظ من شكل (7) أن هذا الفرد يمتلك قدرة مرتفعة جداً إلا أنه قد بدأ يشكك في إجابته «هل من الممكن أن تكون سهلة إلى هذا الحدا!!!». وكنتيجة لذلك سيقوم بإعادة تفسير لهذه المفردات وسيقوم بالإجابة عنها بشكل خاطئ بسبب تفسيراته الخاطئة والبعيدة عن الحل الصحيح، أما بالنسبة لباقي المفردات الأصعب، فعندها يزول الشك عنه، ومن ثم يجيب عنها بشكل صحيح. ويظهر في الشكل استجابات غير مطابقة لمفردات مختلفة الترتيب في الاختبار، إذ يظهر الرقم (0) على يمين النقطة أي أن الطالب أخطأ في الإجابة عن مفردات سهله وأصاب في باقي مفردات الاختبار.

4. الغش: هي الحالة التي يجيب فيها الطالب إجابات صحيحة على المفردات الصعبة أياً كان موقعها في الاختبار، ويحصل هذا السلوك عندما يحصل المفحوص على إجابات صحيحة لمفردات الاختبار بشكل غير عادل علما بأنه غير قادر على الإجابة عنها بشكل صحيح، تشيع هذه الحالة مع الطلبة ذوي القدرة المنخفضة.

التعريف الإجرائي: وقوع قيم مؤشر مطابقة الفرد على مستوى المفردة للمحرزي (Almehrizi, 2010) أعلى من أو قريبه من القيمة المرجعية (1.645) وتكون إجابته صحيحة أيا كان موقع هذه المفردات في الاختبار، وشكل (8) يوضح ذلك.

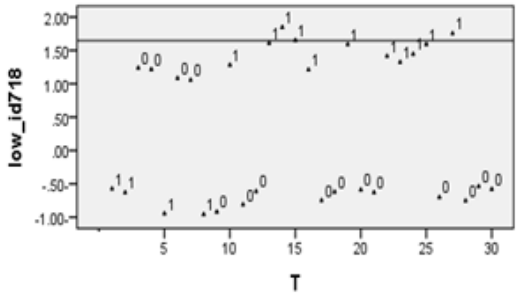
أ- مطابق ب- غير مطابق: • التكاسل • التباطؤ • الإبداع الاستثنائي • الغش • التخمين • خطأ في نقل الإجابة	

شكل (8): استجابة طالب ذي نمط « الغش»

يلاحظ من شكل (8) أن هذا الفرد لديه قدرة منخفضة، ولكن احتمال إجابته إجابة صحيحة عالٍ على المفردات الأولى، ولأن قدرته منخفضة للإجابة عن الأسئلة المتبقية فإنه سيعمل على نقل الإجابات الصحيحة لأصعب المفردات من زميله المجاور له الذي يمتلك إجابة صحيحة على هذه المفردات. ويظهر في الشكل العديد من الاستجابات غير المطابقة لمفردات مختلفة الترتيب في الاختبار، إذ يظهر الرقم 1 على يمين النقطة؛ أي أن الطالب أصاب في الإجابة عن مفردات صعبة، بالرغم من عدم قدرته على حلها كونه طالبًا ضعيفًا، وهذا يعني أنه يمكن أن يكون قد قام بنقل الإجابة من زميلٍ مجاور له في الامتحان.

5. التخمين: قيام الطالب بالإجابة العشوائية على مفردات الاختبار، وحصوله على إجابة صحيحة على مجموعة من المفردات الصعبة في الاختبار، وإجابة خاطئة على مفردات سهلة أيا كان ترتيبها في الورقة الاختبارية، بحيث يظهر عدم مطابقة في الاستجابة على مجموعة المفردات التي أجاب عنها إجابة صحيحة وتلك التي أجاب عنها إجابة خاطئة، وتظهر هذه الحالة بشكل أكبر عند الطلبة ذوي القدرة المنخفضة.

التعريف الإجرائي: وقوع قيم مؤشر مطابقة الفرد على مستوى المفردة للمحرزي (Almehrizi, 2010) أعلى من أو قريبة من القيمة المرجعية (1.645) في عدد من المفردات التي أجاب عنها إجابة صحيحة، وهي مفردة صعبة، وأخرى أجاب عنها إجابة خاطئة، وهي مفردة سهلة وتتوزع هذه المفردات في الورقة الاختبارية بغض النظر عن ترتيبها وشكل (9) يوضح ذلك

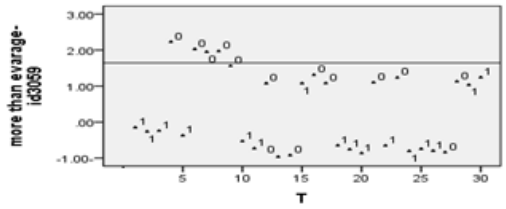
أ- مطابق ب- غير مطابق: • التكاسل • التباطؤ • الإبداع الاستثنائي • الغش • **التخمين** • خطأ في نقل الإجابة	

شكل (9): استجابة طالب ذي نمط « التخمين»

يلاحظ من شكل (9) أن هذا الفرد، لديه قدرة منخفضة، حصل على العديد من النقاط القريبة من القيمة المرجعية والتي تدل على استجابات غير مطابقة، حيث أخطأ في الإجابة على مفردة سهلة وأصاب في الإجابة عن مفردة صعبة بشكل عشوائي في الاختبار مما يدل على لجوء هذا الطالب للتخمين في حل الاختبار.

6. الخطأ في نقل الإجابة: أن يخطئ الطالب في نقل إجابة سؤال معين في ورقة الإجابة مما يؤدي إلى أخطاء في نقل الإجابة في الأسئلة التالية، وتظهر هذه الحالة في نقطة معينة في الاختبار، وتستمر إلى نهايته، وتظهر هذه الحالة بشكل أكبر عند الأفراد ذوي القدرة المرتفعة.

التعريف الإجرائي: وقوع قيم مؤشر مطابقة الفرد على مستوى المفردة للمحرزي (Almehrizi, 2010) أعلى من أو قريبه من القيمة المرجعية (1.645) وتكون إجابة الطالب على هذه المفردات خليط من الإجابات الصحيحة والخاطئة التي لا تتوافق مع صعوبة الأسئلة، وتظهر هذه القيم المرتفعة عند مفردة محددة في بداية الاختبار وتستمر في الظهور عند معظم القيم وفق ترتيبها في ورقة الاختبار، وشكل (10) يوضح ذلك.

أ- مطابق ب- غير مطابق: • التكاسل • التباطؤ • الإبداع الاستثنائي • الغش • التخمين • خطأ في نقل الإجابة	

شكل (10): استجابة طالب ذي نمط « خطأ في نقل الإجابة»

يلاحظ من شكل (10) أن هذا الفرد، لديه قدرة مرتفعة، ومن ثم سينجح بشكل كبير في الإجابة بشكل صحيح على معظم المفردات، ومع ذلك قد لا يعرف الإجابة عن المفردة 17، وبالتالي سيقفز عنها وقد يضع إجابة المفردة 18 محل المفردة 17، وإجابة المفردة 19محل المفردة 18، وهكذا...، في ورقة الإجابة وهذا ما يجعل نتيجته خاطئة على هذه المفردات، كما أن قدرة الفرد المرتفعة يجعل هذه الإجابات الخاطئة تظهر بشكل أكبر.

إجراءات تطبيق الدراسة:

أولاً- التحقق من افتراضات نظرية الاستجابة للمفردة في اختبار القدرة العددية للمستوى الأول لمقياس الخليج للقدرات العقلية المتعددة:

1. افتراض أحادية السمة: تم إجراء التحليل العاملي الاستكشافي على استجابات عينة الاختبار (ن=4206)، بطريقة المكونات الأساسية (Principle Component)، ويشير جدول (2) إلى نتائج التحليل العاملي للتحقق من أحادية السمة.

جدول (2): نتائج التحليل العاملي لمفردات اختبار القدرة العددية المستوى الأول لمقياس الخليج للقدرات العقلية المتعددة.

عدد العوامل الكامنة	الجذر الكامن	نسبة التباين المفسر
1	4.82	16.07
2	2.40	7.99

3	1.41	4.69
4	1.36	4.54
5	1.13	3.80

يشير جدول (2) إلى أن هناك خمسة عوامل حصلت على جذر كامن أكبر من الواحد الصحيح، ويلاحظ أن العامل الأول يفسر أعلى نسبة مئوية من التباين؛ إذ بلغت نسبة التباين المفسر للعامل الأول (16.07)، وهو ضعف ما يفسره العامل الذي يليه (7.99) وهذا يعتبر مؤشر على السيطرة النسبية للعامل الأول مقارنة بالعوامل المتبقية المؤثرة على استجابات الأفراد على مفردات الاختبار كما أشار إليه جلورفيلد (Glorfeld, 1995).

2. الاستقلال الموضعي للمفردات: تم استخدام مؤشر Q3 (Yen, 1992)، من خلال تقدير معالم المفردات ومستويات القدرة استنادًا للنموذج ثلاثي المعالم، ومن ثم حساب الدرجة المتوقعة للفرد في كل مفردة من مفردات الاختبار، وحساب البواقي (Residuals) بطرح درجة الفرد المتوقعة من درجته التي حصل عليها بالفعل في كل مفردة. ويتم بعد ذلك حساب قيمة معامل الارتباط بين درجات البواقي لكل مفردتين من مفردات الاختبار. وجاء المتوسط الحسابي والانحراف المعياري للقيمة المطلقة لمعاملات ارتباط بيرسون بين درجات البواقي المعيارية بين كل مفردتين بقيمة 0.03 لكل منهما، مما تشير إلى تحقق افتراض الاستقلال الموضعي للمفردات في اختبار القدرة العددية.

ثانيًا- التحقق من ملائمة النموذج ثلاثي المعالم لاختبار القدرة العددية لدول الخليج العربي:

تم التحقق من ملائمة المفردات، وذلك باستخدام مؤشر البواقي المعيارية، وقيمة مربع كاي، وكذلك مؤشر القيمة القصوى باستخدام البرنامج (Multi-log 7.03) وبرنامج (Excel). تم فحص مفردات اختبار القدرة العددية باستخدام مؤشر البواقي المعيارية؛ إذ تبين أن جميع المفردات مطابقة وملائمة للنموذج ثلاثي المعالم، فقد تراوح متوسط مؤشر البواقي المعيارية بين (0.42 - 0.97) وجميع هذه القيم غير دالة إحصائيا. كما بلغ المتوسط الكلي لمتوسط متوسطات المفردات لمؤشرات البواقي المعيارية (0.80)، وهو غير دال إحصائيا.

ثالثًا- تقديرات معالم مفردات اختبار القدرة العددية:

استخدمت الدراسة برنامج Multi-log 7.03 في تقدير معالم المفردات (الصعوبة، والتمييز، والتخمين) كما تم تقدير معالم القدرة (Θ) للمفحوصين في العينة الكلية (ن=4206).

كما تم حساب المتوسطات الحسابية (م) والانحرافات المعيارية (ع) لكل معلمة في العينة الكلية. وتراوحت قيم معلمة التمييز للمفردات بين (0.55 - 2.96). كما تراوحت قيم معلمة صعوبة المفردات بين (2.09 - 1.40). وتراوحت قيمة تخمين المفردات بين 0 - 0.38.

رابعًا- تم حساب مؤشرات مطابقة الفرد الخمسة (رايت الموزون، رايت غير الموزون، درازجو، المحرزي الموزون، والمحرزي غير الموزون) وتحديد مطابقة استجابات أفراد الاختبار.

خامسًا- تمت الإجابة عن السؤال الأول باتباع الإجراءات التالية:

أ. تم تجهيز رسومات انتشارية استخدم فيها الأسلوب المطور لتفسير أنماط الاستجابة غير المطابقة باستخدام مؤشرات مطابقة الفرد الخمسة.

ب. تم تجهيز استمارة تحكيم تحتوي على الرسم الانتشاري لكل فرد وأمام كل رسم تم وضع سبعة اختيارات لتصنيف نمط استجابات الفرد: 1) مطابق، 2) التكاسل، 3) التباطؤ، 4) الإبداع الاستثنائي، 5) التخمين،، 6) الخطأ في نقل الإجابة. ويجدر التنبيه أنه يمكن أن يتم تفسير عدم مطابقة نمط استجابة الفرد إلى أكثر من نمط من الأنماط الستة. فقد يعود عدم مطابقة استجابة الفرد إلى أكثر من سبب وعامل في آن واحد. فيمكن أن يقوم الطالب بالغش والتخمين مثلا في أسئلة مختلفة في الاختبار الواحد.

ج. تم استخدام الباحثين أنفسهم في القيام بمهمة تصنيف أنماط استجابة الأفراد باستخدام الطريقة المطورة. لم يتم استخدام محكمين آخرين لضمان صدق نتائج عملية التحكيم فهم على فهم كبير بطريقة التصنيف المطورة. وصدق نتيجة التحكيم ذات أهمية كبيرة، إذ إنها تعتبر محكًا خارجيًا لفحص صدق تصنيف مؤشرات مطابقة الفرد الخمسة في الدراسة. ولتحقيق درجات عالية من صدق التصنيف؛ ولضبط عملية التحيز المحتملة في عملية التصنيف تم القيام بالإجراءات التالية:

- تم تدريب الباحثين على استخدام استمارة تصنيف أنماط الاستجابة.
- تم خلط الأفراد في استمارة التصنيف بشكل عشوائي.
- تم التأكد من عدم وجود أي إشارة عن هوية الطالب في استمارة التصنيف.
- تم التأكد من عدم وجود أية معلومة عن تصنيف نمط استجابات كل فرد في استمارة تصنيف أنمط الاستجابة وفقا لأيٍّ من مؤشرات مطابقة الفرد الخمسة.

- تمت عملية تصنيف أنماط استجابات الأفراد بشكل مستقل بين المحكمين. وبعد الانتهاء من قيام كل محكم بعملية تصنيف جميع الأفراد في الاستمارة، يتم مقارنة التصنيفين لكل فرد. إذا كان التصنيف متشابه بين المحكمين يتم الأخذ بالتصنيف، وفي حالة اختلافهما يتم عقد حوار بين المحكمين لمعرفة أسباب اختلاف تصنيفهما ويتم بعد ذلك الاتفاق على تصنيف موحد لنمط استجابة ذلك الفرد.

د. تم في البداية التأكد من ثبات نتائج عملية التحكيم باستخدام استمارة تصنيف أنماط الاستجابة. تم اختيار عينة عشوائية من الأفراد كمجموعة استطلاعية تكونت من (3) طلاب من ذوي استجابات مطابقة، و(12) طالباً من ذوي الاستجابة غير المطابقة. تم خلط الرسومات الانتشارية لهذه العينة وقام المحكمان بتصنيفها بشكل مستقل. تم حساب معامل الاتفاق لكابا بين تصنيف أنماط الاستجابة من المحكمين، وقد جاءت قيمة معامل الاتفاق لكابا (0.83) وهي نسبة ثبات جيدة.

ه. بعد ذلك، باستخدام استمارة تصنيف أنماط الاستجابة، قام المحكمان بتصنيف المجموعة الكلية من الطلاب والتي تكونت من (178) طالب من ذوي الاستجابة غير المطابقة التي كشفت عنها أي من المؤشرات الخمسة، علما بأن هناك أفرادًا تم تصنيفهم بعدم مطابقة أنماط استجابتهم من مؤشر واحد فقط أو أكثر من مؤشر. كما تم اختيار 22 طالب بشكل عشوائي من ذوي أنماط الاستجابة المطابقة. وقد تم خلط الرسومات الانتشارية لهؤلاء الطلبة (22+178=200).

المعالجة الإحصائية للبيانات:

استخدمت الدراسة لتحليل البيانات كلا من برنامج (Multi-log 7.03) للحصول على معالم المفردات وقدرات الأفراد وكلا من البرامج الإحصائية SPSS، (Excel) للمعالجة الإحصائية:

1. للإجابة عن السؤال الأول تم استخدام معامل الاتفاق لكابا للتحقق من صدق نتائج مؤشرات مطابقة الفرد الخمسة في تصنيف استجابات الأفراد من حيث مطابقتها أو عدم مطابقتها للنموذج في اختبار القدرة العددية بمقياس الخليج للقدرات العقلية المتعددة مع حكم المحكمين على مطابقة الفرد. يتم حساب معامل اتفاق كابا بين تصنيف كل مؤشر مع التصنيف الموحد للمحكمين (مطابق/غير مطابق) لأنماط الاستجابة لعينة 200 طالب.

2. للإجابة عن السؤال الثاني، تم حساب التكرارات والنسب المئوية لتفسيرات أنماط الاستجابات للأفراد المصنفة بأنها غير مطابقة وفقا لكل مؤشر من المؤشرات

الخمسة، أي للأفراد المصنفين من كل مؤشر من المؤشرات الخمسة بعد مطابقة استجابتهم، وتم حصر عدد الأفراد الذين تم تصنيفهم من المحكمين وفقا لتصنيفات الأنماط السبعة في استمارة تصنيف أنماط الاستجابة. بعد ذلك تم استخدام اختبار مربع كاي لفحص الفرضية الصفرية بتساوي توزيع الأفراد لكل مؤشر على تصنيف أنماط استجابات الأفراد.

نتائج الدراسة:

السؤال **الأول:** «ما مدى صدق تصنيف مؤشرات مطابقة الفرد الخمسة في الكشف عن الأفراد ذوي الاستجابات غير المطابقة في اختبار القدرة العددية بمقياس الخليج للقدرات العقلية المتعددة؟»؟

يوضح جدول (3) عدد ونسبة الأفراد الذين تم تصنيف أنماط استجابتهم في اختبار القدرة العددية بأنها غير مطابقة وشاذة وفقا لكل مؤشر من مؤشرات مطابقة الفرد الخمسة. يلاحظ أن مؤشر رايت الموزون أظهر أكبر عدد من الأفراد ذوي الاستجابات الشاذة (164 فرد) يليه مؤشر درازجو (81 فرد) يليه مؤشر المحرزي الموزون (79 فرد) يليه مؤشر المحرزي غير الموزون (71 فرد) يليه مؤشر رايت غير الموزون (9 أفراد). وبلغ إجمالي عدد الأفراد الذين تم تصنيفهم بعدم مطابقة استجاباتهم من جميع المؤشرات الخمسة (404) أفراد، وفي حالات استبعاد الحالات المتكررة، فيتقلص العدد إلى (178) فردًا مما نسبته 4.46% من إجمالي الطلبة في اختبار القدرة العددية.

جدول (3): عدد ونسبة الأفراد المصنفين ذوي استجابات غير مطابقة وفقاً لمؤشرات مطابقة الفرد الخمسة في اختبار القدرة العددية (ن=4206).

اسم المؤشر	العدد	النسبة المئوية
المحرزي الموزون	79	1.88%
المحرزي غير الموزون	71	1.60%
درازجو	81	1.90%
رايت الموزون	164	3.89%
رايت غير الموزون	9	0.21%
الإجمالي مع تكرار الحالات	404	9.61%
الإجمالي بدون تكرار الحالات	178	4.46%

وللحكم على صدق تصنيف كل مؤشر من المؤشرات الخمسة لمدى مطابقة استجابات الفرد، تم مقارنة كل مؤشر مع تصنيف المحكمين لأنماط استجابة الأفراد (مطابق/غير مطابق) باستخدام المجموعة الكلية المكونة من 200 فردا (178 + 22) في استمارة تصنيف أنماط الاستجابة. يوضح جدول (4) قيمة معامل الاتفاق لكابا والدلالة الاحصائية له لكل مؤشر من مؤشرات مطابقة الفرد الخمسة.

جدول (4): معامل الاتفاق لكابا في الحكم على صدق تصنيف كل مؤشر من مؤشرات مطابقة الفرد الخمسة مع حكم المحكمين في مطابقة استجابات الأفراد.

المؤشر	معامل كابا	الدلالة الإحصائية
المحرزي الموزون	0.72	0.000
المحرزي غير الموزون	0.38	0.000
درازجو	0.68	0.000
رايت الموزون	0.33	0.000
رايت غير الموزون	0.09	0.002

تظهر النتائج أن قيمة معامل كابا لمؤشر المحرزي الموزون أكبر القيم؛ إذ بلغت (0.72)، مما يشير إلى أنه الأكثر صدقا وتوافقا مع تصنيف المحكمين لأنماط استجابة الفرد من بين مؤشرات مطابقة الفرد الخمسة. جاء بعد ذلك مؤشر درازجو بمعامل اتفاق كابا بلغ (0.681)، ويليه مؤشر المحرزي غير موزون بمعامل اتفاق كابا بلغ (0.38)، ومن ثم مؤشر رايت الموزون بمعامل اتفاق كابا بلغ (0.33). في حين أن مؤشر رايت غير الموزون أظهر أصغر قيمة لمعامل اتفاق كابا؛ إذ بلغت (0.08). ويظهر أن مؤشري رايت الموزون وغير الموزون أقل المؤشرات صدقا من بين مؤشرات مطابقة الفرد الخمسة في تصنيف مطابقة استجابات الأفراد برغم من أن مؤشر رايت الموزون أظهر أكبر عدد من الأفراد ذوي استجابات غير مطابقة.

السؤال **الثاني:** «ما أنماط الاستجابة غير المطابقة التي كشف عنها كل مؤشر من مؤشرات مطابقة الفرد الخمسة في اختبار القدرة العددية لمقياس الخليج للقدرات العقلية المتعددة؟

يظهر جدول (5) عدد ونسب نتائج تصنيف المحكمين للأفراد المصنفين بعدم مطابقة استجابتهم من كل مؤشر من مؤشرات مطابقة الفرد الخمسة. وقبل استعراض نتائج السؤال الثاني، أظهرت نتائج تصنيف المحكمين لأنماط استجابة الأفراد أن هناك عددًا من الأفراد

تم تصنيف استجابتهم إلى أكثر من نمط من الأنماط الستة المفسرة لعدم مطابقة الاستجابة. فمثلا تم تصنيف (79) فردا ذوي أنماط الاستجابة غير المطابقة حسب مؤشر المحرزي الموزون إلى (97) نمطًا، أي أن هناك أفرادًا تم تصنيف نمط استجابتهم إلى أكثر من تفسير لعدم استجابتهم.

يظهر جدول (5) بشكل عام وجود تباين بين مؤشرات مطابقة الفرد الخمسة في أسباب وأنماط استجابات الأفراد التي كشف عنها المحكمان باستخدام استمارة تصنيف مطابقة استجابات الأفراد. كما تظهر النتائج الدلالة الإحصائية لاختبار مربع كاي في جميع المؤشرات الخمسة ما عدا مؤشر رايت غير الموزون، والتي تشير إلى رفض الفرضية الصفرية بتساوي توزيع الأفراد على تصنيف أنماط استجابات الأفراد والإشارة إلى وجود غلبة لبعض أنماط الاستجابات غير المطابقة أكثر من غيرها.

ففي مؤشر المحرزي الموزون، جاء التخمين أكثر الأنماط انتشارًا في استجابات (13) طالبًا وبنسبة 13.4%، ومن ثم نمط التكاسل في استجابات (12) طالبًا وبنسبة 12.4%، ويليه التباطؤ في استجابات (6) طلاب وبنسبة 6.2%، والإبداع الاستثنائي في استجابات (5) طلاب وبنسبة 5.2%، وأخيرا جاء الخطأ في نقل الإجابة أقل الأنماط انتشارًا في استجابات (4) طلاب وبنسبة انتشار 4.1%، كما تم تصنيف استجابات (5) طلاب من ذوي الاستجابة غير المطابقة على أنها استجابات مطابقة وبنسبة 5.2% .

ويتشابه مؤشر المحرزي غير الموزون مع مؤشر المحرزي غير الموزون في ترتيب نسب انتشار أنماط الاستجابات غير المطابقة مع اختلافهما في النسب المئوية، إذ جاء كل من التخمين والتكاسل في استجابات (9) طلاب في كل منهما وبنسبة 10.7%، ومن ثم التباطؤ في استجابات (6) طلاب وبنسبة انتشار بلغت 7.1%، ومن ثم جاء الإبداع الاستثنائي في استجابات (5) طلاب وبنسبة 6%، وأخيرا جاء الخطأ في نقل الإجابة في استجابة طالب واحد، وبنسبة انتشار بلغت 1.2%، كما تم تصنيف استجابات (8) طلاب من ذوي الاستجابات غير المطابقة على أنها استجابات مطابقة وبنسبة 17.9%.

بينما اختلف مؤشر درازجو قليلا عن مؤشري المحرزي، إذ جاء التكاسل أكثر الأنماط انتشارًا في استجابات (14) طالبًا وبنسبة 13.9%، ومن ثم نمط التخمين في استجابات (12) طالبًا وبنسبة 11.9%، ويليه الإبداع الاستثنائي في استجابات (7) طلاب وبنسبة 6.9%، والتباطؤ في استجابات (5) طلاب وبنسبة 5%، وأخيرا جاء الخطأ في نقل الإجابة أقل الأنماط انتشارًا في استجابات (4) طلاب وبنسبة انتشار 4%، كما تم تصنيف استجابات 8 طلاب من ذوي الاستجابة غير المطابقة على أنها استجابات مطابقة وبنسبة 7.9%.

جدول (5): عدد ونسب تصنيف المحكمين للأفراد المصنفين بعدم مطابقة استجابتهم من كل مؤشر من مؤشرات مطابقة الفرد الخمسة واختبار مربع كاي.

القيمة الاحتمالية		*0.000		*0.000		*0.000		*0.000		*0.308	
درجة الحرية		6		6		6		6		3	
مربع كاي		128.08		80,17		113,51		158,77		3,60	
مجموع الأنماط	97	100%	84	100%	101	100%	187	100%	10	100%	
خطأ في نقل الإجابة	4	4.1%	1	1,2%	4	4%	5	2,7%	-	-	
التخمين	13	13.4%	9	10,7%	12	11,9%	15	8%	-	-	
الغش	52	53.6%	39	46.4%	50	50.5%	65	34.8%	5	40%	
الإبداع الاستثنائي	5	5.2%	5	6%	7	6,9%	16	8,6%	1	10%	
تباطؤ	6	6.2%	6	7.1%	5	5%	5	2,7%	4	40%	
تكاسل	12	12.4%	9	10.7%	14	13,9%	16	8,6%	1	10%	
مطابق	5	5.2%	15	17.9%	8	7,9%	65	34,8%	-	-	
النمط	العدد	النسبة	العدد	النسبة	العدد	النسبة	العدد	النسبة	العدد	النسبة	
العدد	79		71		81		164		9		
المؤشر											

في مؤشر رايت الموزون، جاء كل من الإبداع الاستثنائي والتكاسل أكثر الأنماط انتشارا في استجابات (16) طالباً في كل منهما وبنسبة 8.6%، ومن ثم التخمين في

استجابات (15) طالبًا وبنسبة انتشار بلغت 8%، وأخيرا جاء التباطؤ والخطأ في نقل الإجابة في استجابة (5) طلاب وبنسبة انتشار بلغت 2.7%. كما تم تصنيف استجابات (65) طالبًا من ذوي الاستجابة غير المطابقة على أنها استجابات مطابقة وبنسبة 34.8%.

بينما لم تكن هناك فروق دالة إحصائيا في نسب انتشار أنماط الاستجابات غير المطابقة في مؤشر رايت غير الموزون، إذ كانت قيمة اختبار مربع كاي (3.6) وبقيمة احتمالية (0.31) غير دالة إحصائيا عند مستوى دلالة 0.05.

ومن خلال النظر إلى النتائج في المؤشرات الأربعة التي كانت دالة إحصائيًا، نجد أن الغش أكثر الأنماط انتشارًا بين أنماط الاستجابات غير المطابقة يليه كل من التكاسل والتخمين والتباطؤ والإبداع الاستثنائي، ونمط الخطأ في نقل الإجابة، والذي ظهر بنسب قليلة مقارنة بأنماط الاستجابة غير المطابقة الأخرى. وأشارت النتائج أيضًا إلى أن النسبة الأكبر في تفسير المحكمين لاستجابات الأفراد غير المطابقة كاستجابات مطابقة في مؤشر رايت الموزون، في حين كانت النسبة الأقل في عدد حالات الأفراد التي فسرها المحكمون على أنها مطابقة وهي غير مطابقة في مؤشر المحرزي الموزون، وهذه النتيجة تشير إلى أن أداء مؤشر المحرزي الموزون في الكشف عن مطابقة استجابات الأفراد أفضل من مؤشر رايت الموزون.

المناقشة:

هدفت الدراسة إلى تقديم طريقة مطورة لتفسير أنماط الاستجابة غير المطابقة في الاختبارات بعد تطبيق مؤشرات مطابقة الفرد العامة وتصنيفها على أنها غير مطابقة. كما سعت الدراسة إلى استخدام هذه الطريقة لفحص صدق تصنيف خمسة مؤشرات مطابقة الفرد شائعة الاستخدام باستخدام بيانات اختبار القدرة العددية في مقياس الخليج للقدرات العقلية.

أظهرت النتائج تفوق ثلاثة مؤشرات، وهي مؤشر المحرزي الموزون ومؤشر درازجو ومؤشر المحرزي غير الموزون، إذ حصلت على مؤشرات صدق عالية من خلال معامل اتفاق كابا مع تصنيف الخبراء لمطابقة/عدم مطابقة استجابات الفرد بعد تطبيق الطريقة المطورة لتفسير أنماط الاستجابة. وتتفق هذه النتيجة مع دراسة (Almehrizi, 2010) والتي ذكر فيها أن هذه المؤشرات الثلاثة حصلت على خصائص أفضل تحت ظروف اختبارية مختلفة باستخدام بيانات مولدة، كمحافظتها على مستوى الخطأ من النوع الأول في تصنيف استجابات الأفراد وتحقيقها لمستويات قوة في الكشف عن أنماط الاستجابة غير المطابقة. وتتفق النتائج أيضًا مع دراسة لي وإوليجنيك (Li & Olijnike, 1997) والتي

أظهرت أن المؤشر، قد أظهر نسبة أعلى في الكشف عن الاستجابات غير المطابقة من مؤشر رايت الموزون .

ويقدم المحرزي (Almahrazi, 2003) تفسيرا لوجود تشابه في الأداء بين مؤشر المحرزي الموزون ومؤشر ودرازجو في الكشف عن الأفراد ذوي الاستجابة غير المطابقة من خلال التعمق في معادلة كل منهما. إن صيغة المعادلات في كلا المؤشرين متماثلة، ذلك أن كلا المؤشرين يقوم على استخدام فرق البواقي بعد وزنهما بوزن مختلف.

$$L_z = \frac{\sum_{i=1}^{n}(y_{ij} - p_{ij})\ln(p_{ij}/q_{ij})}{\sqrt{\sum_{i=1}^{n} p_{ij}q_{ij}\left[\ln(p_{ij}/q_{ij})\right]^2}}$$

أي أن الاختلاف بينهما يكمن في دالة الوزن التي يستخدمها كل مؤشر لكل مفردة على حدة. وعلى وجه التحديد دالة الوزن المستخدمة في مؤشرWSR هي الفرق بين احتمال الإجابة الصحيحة في المفردة p_i، والقيمة المتوقع لصعوبة المفردة عندما تكون المفردة والفرد لهما نفس الموقع من مقياس القدرة (0.5)، (p_i-0.5). بينما دالة الوزن المستخدمة في مؤشر هي لوغاريتم قسمة احتمال الإجابة الصحيحة في المفردة احتمال الإجابة الخاطئة في المفردة ($\ln p_i/q_i$). ويعطي كل من الوزنين قيما مختلفة لنفس احتمال الإجابة الصحيحة مع تشابهما في اتجاه التأثير على أداء المؤشرين.

وأظهرت النتائج تفوق مؤشرات المحرزي على مؤشرات رايت من حيث صدق تصنيفها لمطابقة/عدم مطابقة استجابات الأفراد. وتدعم هذه النتيجة باستخدام بيانات حقيقية المتمثلة في اختبار القدرة العددية مع النتائج التي تم الحصول عليها باستخدام بيانات مولدة في دراسة المحرزي (Almehrizi, 2010). ويمكن تفسير تفوق مؤشرات المحرزي على مؤشرات رايت نظريا كما ورد في المحرزي (Almehrizi, 2010) على النحو الآتي. تختلف مؤشرات المحرزي إجرائيًا عن مؤشرات رايت في أن مؤشرات المحرزي لمطابقة استجابة الفرد تستخدم مربع البواقي بين إجابة الطالب الملاحظة واستجابته المحتملة لكل مفردة على حده، والذي يعد متغيرًا متصلا يقبل أي قيمة بين الصفر والواحد. ومن ناحية أخرى فإن مؤشرات رايت تستخدم استجابة الفرد على المفردة الثنائية والتي تعد متغيرًا متقطعا يأخذ قيمتين فقط: إمَّا صفر أو واحد. وهنا نستطيع أن نقول إن المتغير مثار الاهتمام الذي يتم التعامل معه في مؤشرات المحرزي كمقياس لمطابقة استجابة الفرد يعتبر متغيرًا مستمرًّا، في حين يعتبر متقطعًا في مؤشرات رايت. وتقوم مؤشرات المحرزي بعد ذلك بمعايرة مربع البواقي لاستجابات الفرد على مفردات الاختبار باستخدام المتوسطات

الحسابية والانحرافات المعيارية لمربعات البواقي ليكون مقياسًا معياريًّا لمطابقة استجابة الأفراد (Almahrazi, 2003). وفي المقابل، تقوم مؤشرات رايت بمعايرة استجابة الفرد باستخدام متوسطها الحسابي وانحرافها المعياري ليكون مقياسًا معياريًّا لمطابقة استجابة الأفراد.

وقد تجعل تلك الخاصية لمؤشر مطابقة الفرد للمحرزي قادرًا على معالجة المخاوف التي تتعلق باستخدام التقريب الطبيعي للتوزيع ثنائي الحدود لاستجابات الفرد على المفردات ثنائية الاستجابة واستخدام مربع كاي لبيرسون كتوزيع لمربعات متوسطات رايت. بالإضافة إلى ذلك فمن المتوقع أن يكون مؤشر مطابقة الفرد للمحرزي أقل تأثرًا بالمشكلات الأساسية المتعلقة بحجم العينة. ولقد عمل مؤشر مطابقة الفرد للمحرزي على تحسين مؤشرات رايت في كونه يقدم المتوسطات لتحسين منهج البواقي المعيارية في سياق نظرية الاستجابة للمفردة، وهو بهذا أظهر تفوقًا على مؤشرات رايت (Almahrazi, 2003).

كما أظهرت النتائج وجود تباين بين مؤشرات مطابقة الفرد الخمسة في أسباب وأنماط استجابات الأفراد التي كشف عنها المحكمان باستخدام استمارة تصنيف مطابقة استجابات الأفراد. كما تظهر النتائج الدلالة الإحصائية لاختبار مربع كاي في جميع المؤشرات الخمسة ما عدا مؤشر رايت غير الموزون، والتي تشير إلى غلبة لبعض أنماط الاستجابات غير المطابقة أكثر من غيرها.

وقد تنوعت واختلفت نسبة الأنماط في استجابات الطلبة غير المطابقة وفي كل من مؤشرات مطابقة الفرد الخمسة في ضوء تفسير المحكمين في النموذج المقترح للتحكيم في الدراسة الحالية، كما تم تفسير أكثر من نمط في استجابة الفرد غير المطابقة. وتتفق نتيجة هذه الدراسة مع العديد من الدراسات التي أجريت على مطابقة الفرد، والتي ناقشت أنماطًا مختلفة لسلوك الاستجابات غير المطابقة، ومثال على ذلك؛ الدراسة التي ناقش فيها (Wright & Stone, 1997) سلوك التكاسل والتخمين والتباطؤ المذكور في (Meijer, 1996). كما أشار (Hulin, Drasgow & Parsons, 1983) إلى الخطأ في نقل الإجابة والإبداع الاستثنائي كأحد أنماط الاستجابة غير المطابقة.

وقد أشار ميجر (Meijer, 1996) في دراسته إلى أنه لا بد من إدراك عدم القدرة على تقديم تفسير مقبول وذا مغزى لنمط استجابة غير مطابقة. وفي معظم هذه الحالات يكون من الصعب تحديد هذا النمط، ومثال على ذلك أنماط التخمين والغش. كما أنه من الصعب تحديد ما إذا كان نمط الاستجابة غير مطابق أو مطابق نظرًا للطبيعة الاحتمالية للسلوك الكامن للاستجابة على المفردة، إذ إن نمط الاستجابة على المفردة من المحتمل ألا يعكس السلوك الكامن غير المطابق؛ ونظرًا كذلك إلى أن السلوك غير المطابق لاستجابة

الفرد يمكن أن يظهر في عدد قليل من المفردات في الاختبار.

التوصيات والمقترحات:

في ضوء نتائج الدراسة، توصي الدراسة بما يلي:

- استخدام الأسلوب المطور لتوظيف مؤشر مطابقة الفرد على مستوى المفردة الواحدة للمحرزي (Almehrizi, 2010) لتفسير أسباب أنماط الاستجابة الشاذة من قبل المتخصصين في إدارات التقويم والامتحانات بمختلف الهيئات المعنية بالاختبارات، واتخاذ إجراءات إدارية لضبط العملية الامتحانية.
- استخدام مؤشر المحرزي الموزون ومؤشر درازجو للكشف عن الأفراد ذوي أنماط استجابات غير مطابقة في الاختبارات.

في ضوء نتائج الدراسة، واستكمالًا لجوانب الدراسة الحالية، تقترح الدراسة:

- اجراء المزيد من الدراسات على بيانات حقيقية في اختبارات التحصيل الدراسي لفحص فعالية الأسلوب المقترح لتفسير أنماط الاستجابات غير المطابقة.
- اجراء مزيد من الدراسات باستخدام نظرية الاستجابة للمفردة وباستخدام نماذج رياضية مختلفة، للتحقق من صدق المؤشرات الخمسة المستخدمة في الدراسة الحالية والمقارنة بينها.
- إجراء مزيد من الدراسات في الكشف عن أنماط الاستجابة غير المطابقة في أنواع الاختبارات المختلفة في المجالات التربوية والنفسية ومقارنة هذه الأنماط وفق النوع ووفق الدولة.

قائمة المصادر والمراجع:

المراجع العربية:

حمادنة، مروان. (2011). فاعلية أسلوب تحسين مطابقة الفرد على تصحيح تقدير القدرة وتوزيعها المرجعي عند الاختلاف في حجم العينة والنموذج اللوجستي]رسالة دكتوراه غير منشورة[. جامعة اليرموك.

جراح، بندر (2009)، مقارنة مؤشرات مطابقة الفرد لنماذج استجابة المفردة باستخدام بيانات فعلية]رسالة دكتوراه غير منشورة[. جامعة اليرموك.

الزيات، فتحي، والمحرزي، راشد، وعبدالقادر، فتحي، وكمال، عبدالعزيز، والبيلي، محمد، والدوغان، عبدالله، والجاسم، علي. (٢٠١١). التقرير الفني لمقياس الخليج للقدارت العقلية المتعددة (جماس). جامعة الخليج العربية.

عطا، زايد، والشريفين، نضال. (2012). أثر اختلاف شكل توزيع القدرة على معالم المفردة ودالة المعلومات للاختبار. المجلة الأردنية في العلوم التربوية، 8(2)، 151-166.

علام، صلاح. (2005). نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي. دار الفكر العربي.

الناغي، هبة. (2010). أثر عدد مفردات الاختبار على تقدير قدرات الأفراد ودالة المعلومات لاختبار تحصيلي مرجعي المحك في الرياضيات باستخدام نماذج الاستجابة للمفردة. مجلة كلية التربية ببور سعيد، 10، 676-721. https://doi.org/10.21608/jftp.2011.40282

المراجع الأجنبية:

Almahrazi, R. S. (2003). *Investigating anew modification of the residual- based person fit index and its relationship with other indices in dichotomous item response theory* [unpublished doctoral dissertation]. University of Iowa.

Almehrizi, R. S. (2010). Comparing among new residual-fit and wright's Indices for dichotomous three -Parameter IRT model with standardized tests. *Journal of Educational & Psychological Studies*. *Sultan Qaboos University*. 4(2), 1426-. https://doi.org/10.24200/jeps.vol4iss2pp1426-

Drasgow, F., Levine, M.V., & Williams, E.A. (1985). Appropriateness measurement with polychotomous item response models and standardized indices. *British Journal of Mathematical*. https://doi.org/10.1111/j.20448317.1985-.tb00817.x

Deng, w. & Torre, J. (2008). Improving person fit Assessment by correcting the ability Estimate and its Reference Distribution. *Journal of Educational Measurement*, 45(2), 159177-. https://doi.org/10.1111/j.17453984.2008.00058-.x

Dodeen, H. (2010). The use of person-fit statistics in evaluating the accuracy of placement test. *The Educational Journal- Kuwait*, 24(95), 3754-.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory*: *Principles and applications*. Kluwer-Nijhoff. https://doi.org/10.10079-1988-017-94-978/

Hambleton, R, K., Swaminathan, H. & Rogers, H.J (1991). *Fundamental of item response theory*. CA: Sage.

Harnisch, D. L, & Linn, R.L. (1981). Analysis of item response patterns: Questionable test data and dissimilar curriculum practices. *Journal of Educational Measurement*, 18, 133146-. https://doi.org/10.1111/j.17453984.1981-.tb00848.x

Harnisch, D. L & Tatsouka. K.K. (1983). *Comparison of appropriateness indices based on item response theory*. *In R*. *Hambleton*. (*Eds*). *Application of Item Response Theory*. ERIBC.

Hulin, C. L. Drasgow, F. & Parsons, C. K. (1983). *Item response theory, application to psychological measurement*. Il: Irwin.

Glorfeld, L. W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurements*, 55, 377393-. https://doi.org/10.1177001316449505500 3002/

Karabatsos, G. (2003). Comparing the aberrant response detection performance of thirty-six person-fit statistics. *Applied Measurement in Education*, 16(4), 277298-. https://doi.org/10.1207/S15324818AME1604_2

Li, M. F. & Olijnike, S. (1997). The power of rasch person-fit statistics in detecting unusual person patterns. *Appliedpsychological Measurement*, 21, 215231-. https://doi.org/10.11770146621697021300 2/

Meijer, R. R. (1996). Person -fit research: An introduction. *Applied Measurement in Education*, 9, 38-. https://doi.org/10.1207/s15324818ame0901_2

Schmitt, N., Chan, D., Sacco, J. M., McFarland, L.A., & Jennings, D. (1999). Correlates of person fit and effect of person fit on test validity. *Applied Psychological Measurement*, 23(1), 4154-. https://doi.org/10.11770146621992203 1176/

Snijders, T. (2001). Asymptotic distribution of person-fit statistics with estimated person parameter. *Psychometrika*, 66(3), 331342-. https://doi.org/10.1007/BF02294437

Tatsouka, k, k. (1996). Use of generalized person fit indexes zetas for statistical pattern classification. *Applied Measurement in Educational*, 9, 6575-. https://doi.org/10.1207/s15324818ame0901_6

Tendeiro, J. N., Meijer, R. R. (2014). Detection of invalid test scores: The usefulness of simple nonparametric statistics. *Journal of Educational Measurement*, 51(3), 239259-. https://doi.org/10.1111/jedm.12046

Tendeiro, J. N., Meijer, R. R., & Niessen, A. S. (2016). PerFit: An R Package for Person-Fit Analysis in IRT. *Journal of Statistical Software*, 74(5), 127-. https://doi.org/10.18637/jss.v074.i05

Yen W. (1992). Item response theory (6th Ed.). *Encyclopedia of educational research*. MacMillan

Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement*, 14, 97115-. https://doi.org/10.1111/j.17453984.1977-.tb00031.x

Wright, B. D. & Stone, M. H. (1979). *Best test design*. MESA Pres.

الترجمة الصوتية لمصادر ومراجع اللغة العربية: **Romanization Arabic References:**

ḥmādnah mrwāni (2011). fā'iliyyata 'uslwbi taḥsīni muṭābaqati alfardi 'alā taṣḥīḥi taqdyri alqudrati watawzī'ihā almarji'iyyi 'inda alikhtilāafi fī ḥajmi al'aynati wa-al-namūdhaji al-lūjistiyyi risālata duktwrāhi ghayri manshūratin jāmi'ata alyarmūki

jarrāḥun bndr 2009) ،mqārinata mu'ashshirāti muṭābaqati alfardi linamādhiji istijābati almufradati bistikhdāmi bayānātin fi'liyyatin risālata duktwrāhi ghayri manshūratin jāmi'ata alyarmūki

al-zyāt ftḥy wa-al-mḥrzy rāshidun wdādr ftḥy wkmāl 'bdāl'zyz ū mḥmd wa-al-dwghān 'bdāllh wa-al-jāsm 'ly (2011). al-tqryr al-fny lmqyās al-khlyj lil-qdrāt al-'qlyah al-mt'ddah jmās jām'ah al-khlyj al-'rbyah

'aṭā zāyada wa-al-sharīfīna niḍālun (2012). 'athiri ikhtilāafa shakli tawzī'i alqudrati 'alā ma'ālimi almufradati wadālati alma'lūmāti lil-ikhtbāri almajallatu al'urduniyyatu fī al'ulūmi al-tarbawiyati 8(2)166 151- ،.

'ullāmun ṣulā'āaḥun (2005). namādhija alistijābati lil-mufradati alikhtbāriyyati 'aḥādiyata albu'di wamuta'addidati al'ab'ādi wataṭbīqātihā fī alqīāsi al-nafsiyyi wa-al-trbī dāru alfikri al'arabiyyi

al-nāghy hibatan (2010). 'athir 'udada mufradāti alikhtbāri 'alā taqdyri qudrāti al'afrādi wadālati alma'lūmāti likhtbāri taḥṣīliyyi marji'iyyi almiḥakki fī al-rīāḍiyyāti bistikhdāmi namādhiji alistijābati lil-mufradati majallatu kulliyyati al-tarbiyati bibūri sa'īdi 10721 676- ،. https://doi.org/10.21608/jftp.2011.40282

An Improved Method to Interpret the Individual's Aberrant Response Patterns in Tests and the Comparison of Five Person Fit Indices

Rashid Saif Almehrizi(1)

Amal Tabit Awadah(2)

Abstract:

The current study aimed at detecting patterns of aberrant responses through five person fit indices (WSR, USR, Lz, UT and WT) using the numerical ability test data of the Gulf Scale for Multiple Mental Abilities. The study sample consisted of 4206 students from all Gulf countries. Appropriate statistical methods were used to verify the assumptions of the three-parameter IRT model, and the item fit of the numerical ability test to the three-parameter model. The study examined the validity of classification of individuals into fitting/non-fitting for each index using the classification of two referees. WSR and Lz were the most valid indices in relation to other indices. The two referees classified the non-fitting persons into six aberrant patterns (sleeping, plodding, extremely creative examinees, cheating, guessing, alignment errors) using a new evaluation form proposed in the current study. Cheating was the most common aberrant pattern found in this data. The results were discussed in the light of the theoretical framework and previous studies, and some recommendations and suggestions related to the topic of the study were presented.

Keywords: Individual's Fit Indices, Aberrant Response Patterns, Item Response Theory, Gulf Scale for Multiple Mental Abilities.

(1) College of Education - Sultan Qaboos University (Muscat - Oman)
mehrzi@squ.edu.om

(2) Ministry of Education (Muscat - Oman)