[١١]

الخصائص الإحصائية لفقرات مقياس رافن للقدرات العقلية العامة في ضوء النموذج الثنائي والثلاثي المعلم– دراسة مقارنة

أ.د. وفاء محمود عبد الرازق
أستاذ علم النفس المشارك
كلية التربية– جامعة الملك سعود

الخصائص الإحصائية لفقرات مقياس رافن للقدرات العقلية العامة في ضوء النموذج الثنائي والثلاثي المعلم– دراسة مقارنة

أ.د. وفاء محمود عبد الرازق ★

ملخص البحث:

ملخص الدراسة:

استهدفت الدراسة مقارنة تقديرات الصعوبة والتمييز المشتقة من بيانات مفردات اختبار المصفوفات المتتابعة (العادي) لقياس الذكاء والقدرة العقلية "لجون رافن والناتجة من تحليلات النموذجين (الثنائي– الثلاثي) البارمتر، وقد استخدمت الدراسة في التحليلات الخاصة بها عينة من طلاب خمس مدارس ثانوية تابعة لمناطق غرب وجنوب مدينة الرياض، بلغ حجمها (٦١١) طالبة من جميع الصفوف الدراسية (الأول– حتى الثالث)، وقد استخدمت الدراسة الخطأ المعياري للقياس في المعالجات الإحصائية الخاصة باختبار فرضياتها، وقد توصلت الدراسة إلى وجود اختلاف جوهري في تدريجي الصعوبة والتمييز لفقرات اختبار المصفوفات المتتابعة "لرافن" كما توصلت الدراسة إلى أن النموذج ثنائي البارمتر أفضل من النموذج الثلاثي البارمتر في تقديمه لأعلى وأدق تقديرات لمتغيري الصعوبة والتمييز، وأوصت الدراسة في النهاية بضرورة استخدام النموذج الثنائي البارمتر في التحليلات الخاصة بإعداد الاختبارات المختلفة.

كلمات مفتاحية: اختبار المصفوفات المتتابعة– النموذج الثنائي البارمتر– النموذج الثلاثي البارمتر– الخطأ المعياري للقياس– الصعوبة– التمييز.

★ أستاذ علم النفس المشارك– كلية التربية– جامعة الملك سعود.

Study Summary:

The study aimed to compare the difficulty and discrimination estimates derived from John Raven's (regular) matrices test data and the resulting analysis of the two models (Duo- Triple) Albarmitr, has been used to study in their own analysis

a sample of five high school students belonging to the areas west and south of the city of Waller additional, reached its size (611) A student from all grades (I- 3), The study used the standard error of measurement in the statistical treatments for the testing of hypotheses. The study found that there is a fundamental difference in the gradual difficulty and discrimination of the successive matrices of the "Raven". The study also found that the binary model is better than the tripartite model in presenting the highest and most accurate estimates for variables Difficulty and discrimination, Ost need to use Albarmitr binary model tests for the preparation of various analysis

Keywords: Testing successive matrices- The double-parameter model- Triple model Albarmitr- Standard error of measurement- difficulty- discrimination

المقدمة والاطار النظري:

تعتبر نظرية الاستجابة للمفردة أحد الاتجاهات الحديثة للقياس ولقد حققت النماذج الرياضية اللوغاريتمية المرتبطة بها طفرة هائلة ومتطورة في مجال القياس النفسي والسلوكي، ولقد أكدت العديد من الدراسات الأجنبية والعربية تفوق هذه النظرية في تحقيق الموضوعية المامولة منها في هذا المجال إذا ماقورنت بالنظرية الكلاسيكية،(De Champlain, 2010: 109)، (Njriu &Romanoski,2007) كما أن هذه النظرية تعتمد على افتراضات وشروط أقوى من الافتراضات والشروط ذات الصلة بالنظرية الكلاسيكية (Uttaro & leHman,1994:41) .(Rotou,2002:131)

كما تبين من خلال العديد من الدراسات العربية والأجنبية المهتمة بهذا المجال فاعلية استخدام نظرية الاستجابة للمفردة (IRT) والنماذج الرياضية المرتبطة بها في بناء وإعداد مختلف أنواع أدوات القياس التي أجري استخدامها لأغراض جمع انواع مختلفة من البيانات إذا ماقورنت بالنظرية التقليدية للقياس (Loken , Kelly,2011) كما ان الدراسات التي استخدمت نفس النظرية في تقديرات بارمترات المفردات الاختبارية المختلفة قد أكدت أيضاً تميز هذه الادوات من حيث تمتعها بخصائص الموضوعية التي تتسم بها تلك الأدوات المستخدمة في مجال القياس الفيزيقي (2010) Wauters, Van den Noortgate، & Van Der Linden Jason , Edward, Walter (2009)، Fox, Entink (2010)، محمد (٢٠١٠)، Tavakol, Alinia, , Maracy, Mohammad Shafiei, Edelen , Bryce. Reeve (2007)، (2009)Sedaghati, , Foroughi, Linn (1990:115)، Zickar (1998:104)، Rouse,et al. (1999, 282)، Wainer (1989:154)،

ونتيجة لما حققته هذه النظرية وكذلك النماذج الرياضية المرتبطة بها في تحقيق الموضوعية في مجال قياس السلوك الانساني، وفي بناء وإعداد اختبارات حققت تطورا وتقدما هائلاً في شتى ميادين القياس, فقد اتجهت العديد من الدراسات الأجنبية والعربية إلى بناء وتطوير اختبارات جديدة أو إعادة تطوير بعض الاختبارات العالمية المشهورة وشائعة الاستخدام في ضوء النظرية الحديثة للقياس والتحليلات

الإحصائية المتعلقة بأحد النماذج الرياضية المرتبطة بها بهدف تحقيق المزيد من الموضوعية فيما تسفر عنه من قياسات الزيد (٢٠٠٧)، Edelen, ،Reeve (2007) Bryce (2005)، يماني(٢٠٠٥)، Stark, Chernyshenko ،Wiberg (2004) Gumpel, et. ،McHorney(2002) ،& Drasgow Raudenbush (2003) al. (1998)، Molenaa(1996)، Ackerman(1994).

ولقد اشتملت التحليلات الإحصائية الخاصة بالنماذج الرياضية المرتبطة بهذه النظرية الحديثة بعض المؤشرات الإحصائية التي تستخدم في التحقق من دقة وموضوعية نتائج القياس بالاختبار وبنوده المختلفة ومن بين هذه المؤشرات مؤشر الصعوبة ومؤشر التمييز ومؤشر التخمين والذي يمكن تقديرها لكل مفردة من مفردات الاختبار وكذلك للاختبار ككل، (Hambleton& Swaminathan, 1985), (Chernyshenko, 2001)، وتختلف أساليب وصيغ التقديرات الخاصة بهذه المؤشرات باختلاف نموذج التحليل المستخدم، حيث يعتمد كل نموذج من هذه النماذج على متغيرات تختلف عن المتغيرات التي تعتمد عليها النماذج الأخرى، مما قد يؤدي إلى نتائج متباينة لتقديرات قيم تلك المؤشرات، ومما قد يؤدي في النهاية إلى اختلاف دقة تقدير القدرات التي يمكن استخلاصها من الاستجابات على مفردات الاختبار وفقا لاختلاف نموذج التحليل المستخدم (أحادي- ثنائي- ثلاثي) البارمتر، ونظرا لاختلاف المؤشرات الإحصائية التي يعتمد عليها كل نموذج من النماذج الثلاث في تقديرات القدرة الخاصة بالأفراد، فقد دعت الحاجة إلى ضرورة إجراء هذه الدراسة التي تستهدف عقد مقارنة بين تقديرات الصعوبة والتمييز المشتقة من تحليل البيانات الخاصة بالمفردات الاختبارية (باستخدام اختبار المصفوفات المتتابعة لجون رافن العادي) والتي يمكن التوصل إليها باستخدام نموذجين شائعين من حيث الاستخدام في هذا المجال وهما النموذجين(الثنائي- الثلاثي)البارمتر، وذلك بهدف التعرف على أي منهما يمكن أن يعطي أفضل تقدير للخصائص الإحصائية لفقرات الاختبار، وبالتالي يمكن من خلال النتائج المتوقعة من هذه الدراسة تحديد أي من هذه النماذج يمكن الإعتماد عليها مستقبلاً وفقا لكفاءتها في تقديم افضل وأدق معلومات عن قدرات الأفراد في حال خضوعها لإجراءات وعمليات القياس.

نظرية الاستجابة للمفردة Item Response Theory

نظراً لنواحي القصور التي تعاني منها أساليب القياس التي يتم إعدادها وفقاً للنظرية التقليدية للقياس والتي بدت بصورة واضحة من خلال الانتقادات التي وجهت إليها (Linden, Hambleton:2000) كانت هناك محاولات مستمرة من قبل المهتمين بمجال القياس النفسي والسلوكي أسفرت عن بعض الاتجاهات الحديثة للقياس، والتي تمثلت في ظهور ما يسمى بنظرية الاستجابة للمفردة، والنماذج الرياضية المرتبطة بها، والتي حققت إلى حد بعيد الموضوعية المنشودة والتي تتصف بها أدوات القياس المستخدمة في المجال الفيزيقي. (van der Linden, Entink, Fox, Jean-Paul, Rinke, 2010)، (ياسين،٢٠٠٤). وتفترض هذه النظرية وجود سمات أو خصائص يطلق عليها السمات أو القدرات تكمن خلف أداء الفرد على الاختبار، حيث يمكن التنبؤ بقدرة الفرد من خلال هذا الأداء على الاختبار ودرجاته على الخصائص والسمات المقاسة من الاختبار (Steven, Andrew & Mark , 2005) وحيث أن هذه السمات لا يمكن ملاحظتها أو قياسها بشكل مباشر فيطلق عليها السمات الكامنة (Hambelton & et al.:1993,7):

تقدير بارامترات المفردات:

قبل تطبيق الاختبار الذي يتم إعداده وفقاً لأحد النماذج الرياضية المرتبطة بنظرية الاستجابة للمفردة من الضروري تقدير بارامترات كل مفردة داخل هذا الاختبار وفقا للنموذج الذي تم اختياره، وذلك من خلال تطبيق الاختبارات على عينة كبيرة نسبيا من الأفراد، ثم اختبار النموذج، تمهيدا لاشتقاق البارامترات المختلفة وفيما يلي بيان لهذه البارامترات:

(أ) بارامترات التمييز Discrimination parameters:

وهي مؤشرات المفردة التي تفيد بقدرتها على التمييز أو التفريق بين المستويات المختلفة من القدرة المقاسة لدى الممتحنين Azevedo,C.L.N., (Urry ,1977) , (2009)) وتقدر بارمترات التمييز في التحليلات الخاصة بنماذج التحليل المرتبطة بنظرية الاستجابة للمفردة بظل الزاوية التي يصنعها المماس عند

نقطة منتصف المنحنى المميز للمفردة على محور القدرة Wang &Vispoel (1998: 116,).

(ب) بارامترات الصعوبة: Difficulty parameters

وتشير إلى مواقع المفردات على متصل وتدريج الصفة المقاسة من الاختبار وتحدد امتلاكها لمستويات محددة من هذه الصفة , Jason, Edward& Walter (2009)، وأكد الباحثون على أهمية أن تتوزع قيم بارامترات صعوبة المفردة بصورة واسعة..(Ponsoda, V. Wise, S., Olea, J. & Revuelta, J. ,1997)، De McBride & Martine (Ayala, R., Dodd, B., & Koch, W. ,1992). (1983,)، (De Ayala, R., Dodd, B. & Koch, W. ,1990)،

(ج) بارامترات التخمين: Guessing parameters

وهي المؤشرات التي تشير إلى توصل الممتحن إلى الإجابة الصحيحة عن المفردات التي لم تمكنه قدرته في الإجابة عنها عن طريق التخمين العشوائي (Azevedo,C.L.N., 2009)(Urry ,1977) Ponsoda& et al. 1997: (213).

ونظرا للاعتبارات السابقة والتي تتعلق بحساسية تقديرات القدرة لكل من بارمترات الصعوبة والتمييز، والتخمين فإن الباحثة الحالية ترى أن الاعتماد على اختبار المصفوفات العادي لرافن في دراستها الحالية يلبي الهدف الرئيس لهذه الدراسة نظرا لاعتباره من بين أفضل المقاييس لقياس العامل العام (فؤاد أبو حطب، ١٩٩٦، ٣٦٧) وهو اختبار متحرر من أثر الثقافة فهو لايحتاج في استخدامه اللغة او التعليم (Raven,1977,1) واحتوائه على مدى واسع من القدرة المقاسة (الذكاء العام) وكذلك تفاوت مفردات الاختبار من حيث مستوى صعوبتها وقوة تمييزها لعدد (٦٠) مفردة اختبارية هي قوام هذا الاختبار (متيرد، ٢٠٠٠، ١٢١)، وبالتالي تكون هناك إمكانية لإبراز الفروق بين نماذج التحليل المستهدفة من الدراسة فيما يتعلق بتقديرات قدرات الأفراد.

وتقدر قدرات الأفراد، باستخدام التحليلات الرياضية الخاصة بالنماذج الرياضية المختلفة المرتبطة بنظرية الاستجابة للمفردة، Barnes & Wise (1991:143,)، حيث تحدد هذه النماذج العلاقة بين أداء الفرد الملاحظ على الاختبار وبين السمات الكامنة أو القدرات التي تكمن خلف هذا الأداء، وتتلخص الفكرة الرئيسة لنموذج الاستجابة للمفردة في محاولة اشتقاق قياسات أو قيم تقديرية للسمة (القدرة) المشتملة في مجموعة مفردات الاختبار من خلال تدريج لهذه السمة أو القدرة وإعطاء قيم لقدرات الأفراد على هذا التدريج (الحكماني، ٢٠٠٧).

ونماذج الاستجابة للمفردة هي نماذج أو دوال رياضية، على أساس رياضي احتمالي، حيث تتحدد العلاقة بين أداء الفرد على الاختبار والقدرة التي تكمن خلف هذا الأداء وفقاً لأسس نظرية الاحتمالات، وتقدر البارمترات الخاصة بالمفردات الاختبارية وقدرات الأفراد باستخدام نماذج نظرية الاستجابة للمفردة المبينة فيما يلي:

نماذج الاستجابة للمفردة:

هناك عدد من النماذج في نظرية الاستجابة للمفردة IRT models ؛ غير أن هناك ثلاثة نماذج شائعة الاستخدام في البيانات المصححة ثنائيا وهي المقدرة بالدرجة (1) للاستجابة الصواب، و(0) للاستجابة الخاطئة) وسوف يتم إلقاء الضوء على النموذجين المستهدفين بهذه الدراسة (Hambelton& Swaminathan,1985:46) (Bryant, 2005) (علام، ٢٠٠٥):

٢- النموذج ثنائي البارامتر: Two-Parameter Logistic Model

يشترط هذا النموذج أن تكون مفردات الاختبار متفاوتة من حيث الصعوبة ومن حيث التمييز أيضا؛ ولا يوجد بها مجال للتخمين (De Ayala,2009:99-106)، وبالتالي تضمنت الصيغة الرياضية لهذا النموذج متغيري الصعوبة (b) والتمييز (a):

$$P(\theta) = \frac{e^{Dai\,(\theta - b)}}{1 + e^{Dai\,(\theta - b)}}$$

حيث تتضح في هذا النموذج أن المنحنيات المميزة للمفردات تختلف في الجزء الذي تلتقي فيه بمحور القدرة أي أن المفردات تختلف في صعوبتها كما تختلف في الميل slope فميل بعض المنحنيات يتزايد بدرجة أكبر من غيره من المنحنيات، أي أن المفردات التي تمثلها هذه المنحنيات تميز بين الأفراد بدرجة أكبر من غيره من المنحنيات وهذا النموذج يسمى نموذج برينبوم Hambelton) Birnbauum ,1985:46):

ويلاحظ من الصيغة الرياضية الخاصة بالنموذج أنها تتشابه مع الصيغة الرياضية والخاصة بالنموذج أحادي البارمتر إلا أنها تختلف عنها في إضافة عامل التدريج scaling factor D الذي يحول المنحنى اللوغاريتمي إلى تطابق تقريبي مع النموذج الطبيعي Normal Ogive Model وقيمته حوالي (1.7) أما (ai) فهو معامل تمييز المفردة ويتراوح مداه من (−2 إلى +2) وحدة معيارية على افتراض أن توزيع القدرة يتراوح على الفترة المغلقة من (−3 إلى + 3)

وقد توصلت دراسة Van Der Maas, Han L. J.(2011,) والتي اثبت دقة النموذج ثنائي البارمتر عن النماذج الأخرى في اتخاذ القرارات بشأن الفروق بين الأفراد فيما يتعلق بعلم النفس المعرفي مقابل نظرية القياس، كما أن هذا النموذج يتشابه إلى حد ما في تقديراته لقدرات الأفراد للنموذج الأحادي البارمتر، كما ان هناك بعض الدراسات مثل دراسة .Daniel,(2010,) والتي توصلت إلى أن النموذج ثنائي البارمتر أكثر نماذج المفردات دقة في تقدير الخصائص الإحصائية لفقرات اختبارات القدرات العقلية، كما أن هذا النموذج يعطي تقديرات للقدرة تزيد جوهريا عن النموذج الثلاثي البارمتر المشتقة، وتوصلت كذلك دراسة .Raykov,T.,) (Dimitrov,D ,Asparouhov,T.,2010 إلى تفوق النموذج ثنائي البارمتر في الوصول إلى أعلى تقدير للثبات إذا ماقون بالنموذج أحادي البارمتر للبيانات لاختبارات التحصيل ثنائية التقسيم كما أعطى النموذج أخطاء معيارية للتقدير أقل من نظائرها بالنموذج أحادي البارمتر، كما أكدت الدراسة عدم وجود فروق جوهرية بين تقديرات القدرة المشتقة من هذا النموذج وتقديرات القدرة المشتقة من النموذج الثلاث البارمتر. كما توصلت دراسة (Courville, 2004)والتي أجريت على عينة من بيانات اختبار الاستعداد المدرسي بالولايات المتحدة الأمريكية، إلى تفوق النموذج

ثنائي البارمتر على النموذج أحادي وثلاثي البارمتر في دقة تقديرات القدرة، كما توصلت دراسة (Rizavi, S ; Way, W. D.; Lu, Ying; P. M.; S, M., 2004)، التي استهدفت مقارنة نماذج الاستجابة للمفردة في دقة تقديرات القدرة للطلاب المتقدمين لكلية الطب اعتمادا على اختبارات الموائمة التي تقيس القدرة على الاستدلال اللفظي، في ضوء طول الاختبار، أن النموذج ثنائي البارمتر أفضل من النموذج ثلاثي البارمتر وأعطى نتائج إيجابية مع الاختبارات الأقل من حيث الطول، وقد توصلت دراسة (Jaio,2003) التي استهدفت تحديد معدلات الخطأ في تقديرات نماذج الاستجابة للمفردة الثلاثة (الأحادي والثنائي والثلاثي) البارمتر، وفقا للاستجابات ثنائية التقسيم أن النموذج الثنائي كان افضل النماذج الثلاث وأقلهم من حيث معدلات الخطأ.

٣- النموذج ثلاثي البارامتر: Three-Parameter Logistic Model

يفترض أن بيانات الاختبار تتفاوت في صعوبتها وفي قوتها التمييزية وكذلك في احتمالية التخمين لإجاباتها.(De Ayala,2009:123-130) وتفترض نظرية الاستجابة للمفردة أن هناك دالة رياضية ترتبط باحتمالية الاستجابة الصحيحة على المفردة، وتسمى هذه الدالة الرياضية دالة الاستجابة للمفردة item response function والتي تعني ببساطة احتمالية P(θ) الاستجابة الصحيحة على المفردة. وذلك على افتراض أن P(θ) تزيد بزيادة القدرة (θ). وتمثل الدالة في النموذج ثلاثي البارامتر بالمعادلة التالية Edward, 2010():

$$P(\theta) = c + \frac{1-c}{1+e^{-1.7a(\theta-b)}}$$

حيث ترمز a وb وc للبارامترات المميزة للمفردة كما يلي: فيمثل البارامتر a القوة التمييزية للمفردة discriminating power أما b فهي صعوبة المفردة item difficulty والبارامتر c يمثل الخط التقاربي الأسفل lower asymptote للمنحنى المميز للمفردة وهو بارامتر التخمين guessing parameter أو كما أطلق عليه (Lord ,1980: 12-13) مستوى شبه الصدفة Pseudo-chance level parameter وهذا البارامتر يعرض لاحتمالية الفرد ذي القدرة المنخفضة أن يجتاز

المفردة بصورة صحيحة، وعندما لا يجاب على المفردة بصورة صحيحة بالتخمين تكون c =0. (Shih , Kuo & Hsin ,2010)

وقد أثبت هذا النموذج فعالية ودقة في مجال تقدير الخصائص الإحصائية للمفردات الاختبارية وفي مجال تقديرات القدرة للأفراد، فقد توصلت العديد من الدراسات التي استهدفت المقارنة بين نظرية الاستجابة للمفردة (باستخدام النموذج الثلاثي البارمتر) وتقديرات النظرية الكلاسيكية للاختبارات مثل دراسة (Aimee,et,al,2008) التي توصلت إلى أهمية استخدام النموذج اللوغاريتمي ثلاثي البارمتر نظرا لدقة هذا النموذج إذا ماقورن بكل من النموذجين الأحادي والثلاثي البارمتر في انتقاء مفردات اختبار القبول لكليات الطب بالجامعات الأمريكية،كما أثبتت دراسة العكايلة (٢٠٠٧) أن متوسط تقديرات الخطأ المعياري للقياس المقدر باستخدام النموذج الثلاثي البارمتر أدنى من متوسط الخطأ المعياري للقياس المقدر بالنظرية الكلاسيكية، مما يشير إلى الدقة العالية لتقدير النموذج الثلاثي البارمتر بالمقارنة بدقة تقديرات الطرق الكلاسيكية، كما أثبتت دراسة (DeMars,2001) أن هناك أفضلية لاستخدام النموذج الثلاثي البارمتر لمعرفة الفروقات الجماعية المبنية على نظرية الاستجابة للمفردة إذا ماقورن بكل من النموذجين (الأحادي والثنائي)، كما توصلت دراسة الوليلي (٢٠٠٢) التي استهدفت تحديد دقة نماذج الاستجابة للمفردة الثلاث (الأحادي والثنائي والثلاثي البارمتر) في انتقاء مفردات اختبار مرجعي المحك في الجبر لدى عينة من طلاب الصف الثالث الثانوي إلى أن النموذج الثلاثي البارمتر هو الأكثر دقة في انتقاء البيانات مقارنة بالنموذجين الآخرين.

مشكلة الدراسة:

يتضح من خلال العرض السابق للأدبيات الخاصة بنظرية الاستجابة للمفردة والنماذج اللوغاريتمية الثلاثة المرتبطة بها أن دقة تقديرات الإحصائية لبارمترات المفردات اختلفت من نموذج لآخر، فقد أثبتت دراستي كل من (Fan,2009)، (Robert,2010) أن نموذج راش (الأحادي البارمتر) الاكثر دقة في تقدير كل من القدرة والصعوبة مقارنة بالنموذجين الآخرين الثنائي والثلاثي

مجلة الطفولة والتربية – العدد الثاني والخمسون – الجزء الثالث – السنة الرابعة عشرة – أكتوبر ٢٠٢٢

البارمتر. في حين أثبتت Van Der Maas, Han L. J.(2011,) تفوق النموذج الثنائي البارمتر على النماذج الأخرى في اتخاذ القرارات فيما يتعلق بالفروق الفردية للأفراد،كما أثبتت دراستي (Courville,2004)، (Jaio,2003) أن النموذج الثنائي كان افضل النماذج الثلاث وأقلهم من حيث معدلات الخطأ.كما أثبتت أيضا دراسة (Rizavi, et.al., 2004)، أن النموذج الثنائي كان أفضل في تقديرات القدرة مع الاختبارات ذات الطول الأقل إذا قورن ذلك بالنموذج الثلاثي البارمتر، بينما أثبتت دراسات (العكايلة، ٢٠٠٧) (DeMars,2001) و(الوليلي، ٢٠٠٢) أن النموذج الثلاثي البارمتر هو النموذج الأكثر دقة في انتقاء البيانات مقارنة بالنموذجين الآخريين، مما يشير إلى تعارض النتائج فما يتعلق بدقة تقديرات معالم المفردات الاختبارية وكذلك معالم الأفراد (متمثلة في تقديرات القدرة) باختلاف النموذج المستخدم، وعلى الرغم ان وجهة الدراسة الحالية لاتتجه نحو حسم هذا الخلاف فيما يتعلق بدقة تقدير قدرات الأفراد إلا أن الدراسة الحالية تتجه نحو الكشف عن دقة تقديرات متغيرات (الصعوبة والتمييز) وحيث ان نتائج الدراسات المشار إليهنا آنفا أشارت لوجود اختلافات في تقديراتها للقدرة وفقا للنموذج المستخدم، أي نموذج من النماذج المرتبطة بنظرية (IRT)، وهذا مما أثار عدد من التساؤلات لدى الباحثة الحالية تتعلق بمدى اختلاف هذه النماذج في تقديراتها لمتغيريي صعوبة وتمييز فقرات الاختبار من خلال نموذج الاستجابة المستخدم في ضوء استخدام أسلوب مختلف عن الأساليب التي استخدمت بالدراسات السابقة (وهو اسلوب الخطا المعياري لتقدير القدرة) في تقدير دقة هذين المتغيرين، كما تعلقت تلك التساؤلات بالنموذج الذي يمكن ان يحقق اعلى دقة تقديرات صعوبة وتمييز فقرات الاختبار، هذا ماتحاول ان تجيب عنه الدراسة الحالية، وبالتالي يمكن أن تتلخص مشكلة الدراسة في الإجابة على الأسئلة التالية:

- هل تختلف تقديرات صعوبة وتمييز فقرات الاختبار باختلاف نموذج التحليل المستخدم (النموذج الثنائي والثلاثي البارمتر)؟
- هل تختلف الأخطاء المعيارية لتقدير الصعوبة والتمييز لفقرات الاختبار باختلاف نموذج التحليل المستخدم؟

أهمية الدراسة:

تبدو أهمية الدراسة من خلال النقاط التالية:

أولاً: الأهمية النظرية والتطبيقية:

الدراسة تهتم بتحليل اختبار المصفوفات المتتابعة (العادي) لقياس الذكاء والقدرات العقلية باستخدام بعض النماذج الرياضية اللوغارتمية المرتبطة بنظرية الاستجابة للمفردة وهذا له أهميته النظرية والعلمية في مجال القياس العقلي بصفة خاصة وفي مجال القياس السلوكي بصفة عامة.

تحاول الدراسة التعرف على أي من نماذج التحليل الثلاثة المستهدفة من الدراسة التي يمكن أن تعطي أفضل وأدق تقدير للخصائص الإحصائية لفقرات الاختبار المستخدم بالدراسة، وهذا له قيمة علمية في مجال التطبيقات التي تستخدم بها تلك النماذج وبصفة خاصة في مجالات القياس المختلفة والمتنوعة.

أهداف الدراسة:

١- الكشف عن مدى تأثر الخصائص الإحصائية لفقرات اختبار المصفوفات المتتابعة لقياس الذكاء والقدرة العقلية بنموذج التحليل المستخدم.

٢- الكشف عن مدى اختلاف الأخطاء المعيارية لتقديرات الصعوبة والتمييز الخاصة بفقرات اختبار المصفوفات المتتابعة بنموذج التحليل المستخدم.

٣- الكشف عن دقة النموذج المستخدم في تقدير متغيري الصعوبة والتمييز لفقرات الاختبار المستخدم بالدراسة.

محددات الدراسة:

- اقتصرت الدراسة على بيانات اختبار المصفوفات المتتابعة (العادي) لجون رافن.
- اقتصرت الدراسة على تطبيق الأداة الخاصة بها على الطالبات دون الطلاب.
- اقتصرت الدراسة في تطبيق الأداة الخاصة بها على طالبات المرحلة الثانوية من التعليم العام دون الخاص.
- اقتصرت الدراسة على التحليلات بنموذجين من نماذج مرتبطة بنظرية الاستجابة للمفردة (النموذج الثنائي البارمتر – النموذج الثلاثي البارمتر).

- اقتصرت التحليلات الخاصة بمفردات اختبار الدراسة وتقديرات القدرة لعينة البحث على التحليلات الخاصة ببرنامج BILOG-MG3 وكذلك برنامج التحليلات الإحصائية SPSS الإصدار الخامس والعشرون.

المفاهيم الأساسية للدراسة الحالية:

مؤشرات دقة القياس:

حيث ان الدراسة الحالية تهتم بقيمة دوال معلومات مفردات الاختبار وأهمية دور هذه الدوال في دقة تقدير القدرة في الاختبارات التي يتم إعدادها في ضوء نظرية الاستجابة للمفردة؛ فقد تبين من خلال مسح الأدبيات وجود عدد من المؤشرات التي تستخدم لقياس دقة تقدير القدرة، وأصبحت هذه المؤشرات بمثابة الدلالات التي يرجع إليها لتفضيل أسلوب لتقدير القدرة عن أسلوب آخر. ومؤشرات دقة القياس measurement precision indices يمكن أن تنقسم إلى نوعين من المؤشراتSamajima(2010,):

مؤشرات عامة حيث تعطي قيمة مختصرة عن دقة التقدير للاختبار ككل:

وتتمثل في تقديرات الثبات ومؤشر التقارب والدقة والكفاءة.

مؤشرات شرطية:

- وتتمثل في الجذر التربيعي لمتوسط مربعات الخطأ Root Mean Square Error (RMSE)
- الخطأ المعياري Standard Error of Estimation (SEE)

وسوف تقتصر الدراسة على المؤشرات الشرطية وبيانها فيما يلي:

مؤشرات الدقة المشروطة:

(أ) الجذر التربيعي لمتوسط مربعات الخطأ:

يعتبر الجذر التربيعي لمتوسط مربعات الخطأ Root Mean Square Error (RMSE) هو مقياس الخطأ الكلي للتقدير، (Wang & Vispoel 1998)، ويوضح بالرسم من خلال محورين المحور الأفقي الذي يمثل القدرة والمحور الرأسي من 0 إلى 1، ثم ترسم نقاط عند كل مستوى قدرة التي تمثل القيمة بموازاة المحور الأفقي من خلال المحور الرأسي، وكلما انخفضت القيمة واقتربت من

الصفر دل هذا دقة أعلى.(Vispoel ,1993:130)،(Weiss ,1982: 483)، (Wang & Hanson ,1999: 271).

(ب) الخطأ المعياري:

يشير (Urry, 1977)أن الخطأ المعياري لتقدير القدرة Standard Error of Estimation (SEE). هو الانحراف المعياري لتقديرات القدرة حول خط الانحدار عند مستوى ثابت من القدرة الحقيقية، وقد استخدم الخطأ المعياري في دقة تقدير العديد من البارمترات في بعض الدراسات مثل دراسة ((Waugh , 2004)، والتي اجريت على عينة من طلاب الجامعة باستراليا والتي استهدفت المقارنة بين دقة الخصائص السيكومترية لاختبار مداخل الدراسة من خلال التحليلات الخاصة بالنظرية التقليدية ونموذج راش والتي تبين من خلال هذا المؤشر أن تقديرات الصعوبة لفقرات الاختبار جاءت أدق من تقديرات الصعوبة الناتجة من تحليلات النظرية التقليدية للقياس، كما استخدم هذا المؤشر في دراسة (أبو هاشم، ٢٠٠٦) في دراسة استهدفت مقارنة بين النظرية التقليدية ونموذج راش فى اختيار فقرات مقياس مداخل الدراسة لدى طلاب الجامعة، والتي اجريت على عينة من طلاب كلية التربية بجامعة الزقازيق في التنبؤ بالمستوى التحصيلي للطلاب من خلال اختبار مداخل الدراسة التى اختيار فقراته في ضوء نموذج راش (الأحادي البارمتر)، كما استخدم هذا المؤشر أيضا في دراسة (العكايلة، ٢٠٠٧) والتي اجريت على عينة من طلاب الصف التاسع الابتدائي بمدارس مديريات التربية بعمان، والتي استهدفت مقارنة تقدير اخطاء القياس لكل من النظرية التقليدية والحديثة للقياس،مستخدمة في ذلك الخطأ المعياري للتقدير، وانتهت بأن الخطا المعياري للتقدير باستخدام نماذج الاستجابة للمفردة جاءت أقل من الخطا المعياري للتقدير باستخدام النظرية التقليدية ويمكن حساب الخطأ المعياري بالمعادلة الآتية من (Wang & Vispoel, 1998:119):

$$SE\ (\hat{\theta}) = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N}\left(\hat{\theta}_j - \frac{\sum_{K=1}^{N}\hat{\theta}_k}{N}\right)^2}$$

فروض الدراسة:

١- لا تختلف قيم تقديرات الصعوبة وتقديرات التمييز لفقررات اختبار المصفوفات المتتابعة باختلاف نموذج التحليل المستخدم.

٢- لا تختلف الأخطاء المعيارية لتقديرات الصعوبة وتقديرات التمييز لفقررات اختبار المصفوفات المتتابعة باختلاف نموذج التحليل المستخدم.

الطريقة والإجراءات:

عينة الدراسة:

اعتمدت الدراسة في إجراءات اختبار الفرضيات الخاصة بها على عينة عشوائية طبقية من طالبات المرحلة الثانوية بلغ تعدادها (٦١١) طالبة من جميع الصفوف الدراسية بالمرحلة الثانوية (الأول- حتى الثالث) أجري سحبها من خمسة من المدارس الثانوية بالمدارس الثانوية (غرب وجنوب الرياض)

أداة الدراسة:

استخدمت الدراسة اختبار المصفوفات المتتابعة (العادي) "لجون رافن"، وهو أحد الاختبارات العالمية المشهورة في قياس (الذكاء) القدرة العقلية العامة والمتحررة من أثر الثقافة، ويتكون الاختبار في صورته العادية من خمسة مجموعات من الأسئلة (أ، ب، ج، د، هـ) حيث تحتوي كل مجموعة على اثنتي عشر مشكلة (مفردة) وبذلك يتكون الاختبار في جملته من (60) مفردة مرتبة في سياق متدرج من حيث الصعوبة، وتتكون كل مفردة من مربع كبير- مقدمة السؤال- يحتوي على خلفية من أشكال معينة قد تكون نقوش أو عدد من الرسومات التي تربطها علاقة محددة، ويتضمن هذا المربع الكبير في أقصى يساره السفلي على جزء صغير- على هيئة مربع- قد نزع من الشكل الكلي الأساسي المعروض بهذه المقدمة، ثم تلي هذه المقدمة ستة أو ثمانية بدائل تشبة المربع السفلي المنزوع من المربع الأساسي الذي يمثل مقدمة السؤال، حبث يحتوي كل بديل من هذه البدائل على شكل من الأشكال قد يتسق مع خلفية الأشكال الواردة بمقدمة السؤال او لا يتسق، وتتضمن هذه البدائل بديل واحد فقط يمثل الإجابة الصواب ويمثل الجزء الصغير المنزوع من الشكل الوارد بالمقدمة، ويجب على المفحوص اختيار هذا الجزء من البدائل الستة أو الثمانية

الموجودة أسفل مقدمة السؤال. والبيانات التي يمكن الحصول عليها من تطبيق الاختبار هي بيانات ثنائية التقسيم حيث تقدر الاستجابة الصحيحة بالدرجة (١) والخاطئة بالدرجة (صفر)، لكل مفردة من مفردات الاختبار (فؤاد أبو حطب،١٩٩٦) (Raven, J.,1989) (صفوت فرج،٢٠٠٧).

وقد وقع اختيار الباحثة لهذا الاختبار نظرا لإمكانية استخدامه مع مدى واسع من الفئات العمرية المختلفة من (6 سنوات إلى 65 سنة) (معوض، ٢٠٠٧: ٨٠) وقد أجري التحقق من الشروط والخصائص السيكومترية للاختبار في خلال العديد من الدراسات والأبحاث العربية والأجنبية التي استخدمت هذا الاختبار (أبو حطب، ١٩٧٨)، (القرشي،١٩٨٧)،(عليان، ١٩٨٩)،(زمزمي،١٩٩٩)، (متيرد، ٢٠٠٠)، (Evans, L ,1980). (Raven,j.,1989) (Valencia, R ,1984). (Carlson, J.S. & Jensen, M ,1981). كما تحققت الدراسة أيضا من ثبات الاختبار باستخدام معادلة كيودار – ريتشاردسون (صيغة 21) وبلغت قيمته (0.95) وهي قيمة تشير إلى ثبات مرتفع.

كما استخدمت الباحثة المجموع الكلي للطالبات في السنة الدراسية السابقة كمحك خارجي لصدق الاختبار على عينة استطلاعية قوامها (200) من طالبات الصفوف الثلاثة من المرحلة الثانوية بواقع عدد [(70) من الصف الثالث الثانوي +(65) من الصف الثاني، (65) طالبة من الصف الأول]، وقد بلغ قيمة معامل ارتباط بيرسون بين الاختبار والمحك (0.90) وهي قيمة مرتفعة تؤكد تمتع المقياس بمستوى عال من الصدق وبالتالي كانت هناك ثقة في دقة البيانات التي تم الحصول عليها باستخدام هذا الاختبار.

الإجراءات:

١- أجري تطبيق الاختبار على عينة الدراسة.

٢- تم تحليل استجابات عينة الطالبات لجميع بنود الاختبار باستخدام نموذجي التحليل (ثنائي- ثلاثي) البارمتر).وذلك باستخدام برنامج BILOG-MG3.

٣- إيجاد تقديرات الصعوبة والتمييز لفقرات الاختبار لكل نموذج على حدة.

٤- إيجاد الأخطاء المعيارية لتقديرات الصعوبة والتمييز لفقرات الاختبار المشتقة من تحليلات كل نموذج على حدة.

٥- إيجاد مطلق الفرروق بين تقديرات الصعوبة المتناظرة لفقرات الاختبار والمشتقة من تحليلات كلا النموذجين.

٦- إيجاد مجموع الخطأين المعياريين لتقديرات الصعوبة المتناظرة لكل فقرة من فقرات الاختبار والمشتقة من تحليلات كلا النموذجين.

٧- تحديد دلالة الفروق بين تقديرات الصعوبة المتناظرة لفقرات الاختبار والمشتقة من تحليلات كلا النموذجين من خلال القاعدتين التاليتين:

- "إذا كان الفرق بين تقديري الصعوبة المتناظرين للفقرة الاختبارية أكبر من أو يساوي مجموع الخطأين المعيارين لهما اعتبر هذا الفرق دالاً لتلك الفقرة أما إذا كان الفرق بين تقديري الصعوبة المتناظرين للفقرة الاختبارية أقل من مجموع الخطأين المعيارين لهما اعتبر هذا الفرق غير دالاً ".
- إذا كان عدد الفروق الدالة يتجاوز نسبة (٠.٠٥) من العدد الكلي لفقرات الاختبار أعتبر ذلك فرقا جوهريا بين تدريجي الصعوبة المشتقين من تحليلات النموذجي (الثنائي والثلاثي) البارمتر أما إذا كان عدد الفروق الدالة لايتجاوز نسبة (٠.٠٥) من العدد الكلي لفقرات الاختبار أعتبر ذلك فرقا غير جوهريا

٨- تكرار الخطوات أرقام(٥إلى٩) مع المتغير البحثي الآخر وهو متغير التمييز.

٩- إيجاد متوسط الأخطاء المعيارية لكل من تقديرات الصعوبة والتمييز (كل على حدة) لفقرات الاختبار المشتقة من التحليلات الخاصة بنموذجي التحليل (الثنائي والثلاثي) البارمتر.

١٠- أُعتبر نموذج التحليل الذي يقل متوسط أخطاؤه المعيارية المشتقة من تحليلاته لتقدير متغير (الصعوبة/التمييز) عن الأخطاء المعيارية المشتقة من تحليلات النموذج الآخر يكون هو النموذج الأكثر دقة في تقدير هذا المتغير

منهج الدراسة والأساليب الإحصائية:

اعتمدت الدراسة في إجراءاتها على المنهج الوصفي (التحليلي- المقارن) مع استخدام الأساليب الإحصائية الآتية:

- أسلوب التحليل العاملي الاستكشافي للتحقق من توافر افتراض أحادية البعد في بيانات اختبار المصفوفات المتتابعة المستخدم بالدراسة باستخدام الحزمة الإحصائية لبرنامج التحليلات الإحصائية SPSS الإصدار الخامس والعشرون.
- استخدام مؤشر الملاءمة الإحصائي مربع كاي المتوافر بتحليلات برنامج BILOG-MG3 للتحقق من توافر افتراض استقلالية المحل لمفردات اختبار المصفوفات المتتابعة.
- وكذلك استخدام تحليلات نفس البرنامج للحصول على تقديرات الصعوبة والتمييز لفقرات اختبار الدراسة.

عرض نتائج الدراسة ومناقشتها:

أولا: التحقق من توافر افتراضات نظرية الاستجابة للمفردة في بيانات اختبار الدراسة وبيانها فيما يلي:

أ- التحقق من توافر افتراض احادية البعد لبيانات اختبار المصفوفات المتتابعة:

وقد اعتمدت الدراسة على استخدام التحليل العاملي الاستكشافي للتحقق من هذا الافتراض وأسفرت التحليلات الخاصة بهذا الأسلوب عن النتائج المبينة بالجدول التالي:

جدول (1)

الجذور الكامنة للعوامل الثلاثة الأولى الناتجة عن التحليل العاملي لمفر دات اختبار المصفوفات المتتابعة

العوامل	الجذور الكامنة		
	الكلي	النسبة المئوية للتباين	النسبة المئوية للتكرار التصاعدي
١	10.288	26.146	26.146
٢	4.177	6.962	33.108
٣	3.631	6.052	39.161

يتضح من جدول التباين الكلي المفسر(السابق) أن قيمة الجذر الكامن للعامل الأول بلغت (10.288) كما بلغت قيمة الجذر الكامن للعامل الثاني (4.177)، وقد

لوحظ أن نسبة التباين المفسر بالعامل الأول من النسبة الكلية للتباين المفسر بجملة العوامل التي استخرجت من التحليل بلغت (26) تقريبا كما نسبة قيمة الجذر الكامن للعامل الأول إلى قيمة الجذر الكامن للعامل الثاني بلغت (2.5) تقريبا، أي أن النسبة تزيد عن القيمة (2) أحادية البعد لبيانات الاختبار، وهذه النتائج تشير إلى توافر افتراض أحادية البعد، حيث أشار العديد من خبراء القياس ومنهم ريكاس (Recase,cited in Hattie,1985) إلى أنه إذا فسر العامل الأول أكبر تباين، فإن ذلك يعتبر مؤشر على أحادية البعد، وحدد ريكاس أن هذه القيمة يجب أن تكون أكبر من أو تساوي (20%) كما أن مؤشر لورد(lord , 1980) يشير إلى أن المفردات تحقق خاصية أحادية البعد إذا كانت نسبة الجذر الكامن للعامل الأول إلى نسبة الجذر الكامن للعامل الثاني يجب ان تكون أكبر من (2) مما يعني أن الاختبار متشبعا بعامل واحد مهيمن.

ب- التحقق من توافر افتراض استقلالية المحل لبيانات اختبار المصفوفات المتتابعة:

ويعني هذا الافتراض عدم اعتماد بعضها على البعض اخر (أي ان الإجابة على أي مفردة اختبارية لاينبغي ان تتأثر بالإجابة على المفردات الأخرى) وقد أسفرت نتائج التحليلات الخاصة باختبار هذا الافتراض والمتضمنة ببرنامج BILOG-MG3 عن عدم تجاوز أي مفردة من مفردات الاختبار لحدود مؤشر حسن المطابقة الإحصائية (Chi-square) كما أن توافر افتراض أحادية البعد لمفردات الاختبار تعني توافر افتراض استقلالية المحل (Hambleton, Swaminathan, 1985) .(Hulin,Drasgow&Parson,1983)

ج- التحقق من توافر افتراض السرعة:

والمقصود من هذا الافتراض أن الأسئلة المتروكة في نهاية الاختبار ينبغي ان تكون قد تركت بسبب أن قدرة المفحوص لم تمكنه من الإجابة عنها بشكل صائب وليس بسبب انه لم يكن هناك وقت كاف لمحاولة الممتحن أن يجيب عنها، وقد راعت الباحثة في تطبيق اختبار الدراسة إعطاء وقتا مفتوحا للإجابة عن جميع فقرات

الاختبار لحين انتهاء كل المفحوصين من محاولة الإجابة عن جميع فقرات الاختبار.

(د) التحقق من حسن مطابقة بيانات الاختبار للنموذج (Goodness-of-Fit-Test):

لقد أسفرت نتائج المطابقة للأفراد والمفردات عن النتائج التالية:

- مطابقة المفردات Fit Item لتوقعات النماذج اللوغاريتمية (الثنائية والثلاثية البارمتر) إعتمادا اعتمادا على دليل برنامج BILOG_MG3، **(Zimowski, et.al.,2003)** وذلك بعد استبعاد المفردات غير المطابقة ثم إعادة التحليل مرة اخرى، حيث أن المفردة تكون غير مطابقة لنموذج التحليل إليها إذا كانت قيمة مؤشر المطابقة الداخلي (Chi_Squre) دالة عند مستوى (0.05)، وقد تبين من نتائج التحليل عدم وجود أية مفردة تجاوزت حدود المطابقة لأي نموذج من النماذج الثلاثة المستخدمة في التحليلات.
- مطابقة الأفراد Fit Person لتوقعات النماذج اللوغاريتمية (الثنائية والثلاثية البارمتر)، وقد تبين أن إجمالي عدد الطالبات وعددهن (18) طالبة جاء مؤشر المطابقة لقدراتهم خارج حدود المطابقة المسموح بها (في احد التحليلات الخاصة بأحد النموذجين على الأقل) وتم استبعادهن من التحليلات النهائية.

ثانياً:عرض نتائج التحليلات الخاصة باختبار فروض الدراسة:

١- عرض النتائج الخاصة باختبار الفرض الأول من فروض الدراسة:

أسفرت إجراءات التحقق الفرض الأول والذي ينص على:

"لا تختلف قيم تقديرات الصعوبة وتقديرات التمييز لفقررات اختبار المصفوفات المتتابعة باختلاف نموذج التحليل المستخدم"، والتي اعتمدت على استخدام تقديرات الصعوبة المشتقة من تحليلات نموذجي التحليل (الثنائي والثلاثي) البارمتر عن النتائج المبينة بالجداول أرقام (2، 3، 4).

جدول (2)

تقديرات الصعوبة والتمييز المشتقة من تحليلات النموذج الثنائي والثلاثي البارمتر

مفردات الاختبار	تحليلات النموذج الثلاثي البارمتر				تحليلات النموذج الثنائي البارمتر			
	تقديرات الصعوبة بوحدة اللوجيت	الخطأ المعياري لتقديرات الصعوبة	تقديرات التمييز	الخطا المعياري لتقديرات التمييز	تقديرات الصعوبة بوحدة اللوجيت	الخطأ المعياري لتقديرات الصعوبة	تقديرات التمييز	الخطا المعياري لتقديرات التمييز
ITEM0	-2.511	0.651	1.052	0.263	-3.887	0.719	0.861	0.188
ITEM0	-3.083	0.955	0.787	0.25	-6.513	2.288	0.64	0.243
ITEM0	-5.987	2.195	0.188	0.055	-6.11	1.783	0.328	0.095
ITEM0	-2.015	0.364	2.525	1.807	-5.234	1.670	0.833	0.340
ITEM0	-2.123	0.52	1.126	0.231	-2.768	0.403	1.063	0.169
ITEM0	-6.165	2.04	0.35	0.111	-6.572	2.119	0.451	0.149
ITEM0	-0.949	0.253	1.514	0.316	-1.224	0.185	1.521	0.243
ITEM0	-1.928	0.413	1.406	0.253	-2.858	0.719	1.091	0.168
ITEM0	-	6.547	0.087	0.023	-8.229	2.288	0.35	0.111
ITEM0	-1.677	0.48	0.981	0.197	-2.409	1.783	0.923	0.143
ITEM0	-0.4	0.107	2.195	0.297	-0.523	1.670	2.158	0.261
ITEM0	-0.137	0.126	2.013	0.383	-0.503	0.403	1.614	0.218
ITEM0	-3.583	0.81	1.004	0.245	-4.696	2.119	0.928	0.200
ITEM0	-3.955	0.217	0.856	0.217	-4.487	0.185	0.956	0.177
ITEM0	-	6.996	0.087	0.023	-8.951	0.356	0.359	0.124
ITEM0	-1.994	0.429	1.264	0.217	-2.566	2.681	1.19	0.168
ITEM0	-0.946	0.275	1.67	0.409	-1.36	0.363	1.602	0.274
ITEM0	-0.358	0.224	1.386	0.324	-1.043	0.089	1.04	0.165
ITEM0	-0.78	0.21	1.987	0.387	-1.382	0.091	1.464	0.174
ITEM0	-0.579	0.243	1.417	0.273	-1.228	0.816	1.112	0.139
ITEM0	-0.831	0.322	1.296	0.3	-1.581	0.752	1.036	0.138
ITEM0	-0.852	0.207	1.721	0.303	-1.089	3.195	1.732	0.242
ITEM0	0.162	0.083	2.239	0.452	-0.019	0.324	2.069	0.281
ITEM0	-17.26	4.93	0.087	0.023	-7.436	0.199	0.316	0.097
ITEM0	-1.942	0.711	0.749	0.179	-3.838	0.172	0.594	0.114
ITEM0	-1.801	0.6	0.854	0.183	-3.618	0.141	0.606	0.109
ITEM0	-0.861	0.287	1.413	0.289	-1.526	0.166	1.213	0.176
ITEM0	-1.389	0.262	1.751	0.304	-1.627	0.212	1.749	0.246
ITEM0	-0.239	0.177	1.351	0.249	-0.593	0.148	1.241	0.172
ITEM0	-0.886	0.407	1.065	0.278	-1.949	0.059	0.827	0.069
ITEM0	0.373	0.215	1.069	0.262	-0.227	2.314	0.912	0.157
ITEM0	-0.224	0.352	0.849	0.215	-1.127	0.723	0.736	0.133
ITEM0	0.24	0.16	1.274	0.224	-0.15	0.628	1.123	0.158
ITEM0	0.281	0.146	1.736	0.374	-0.441	0.211	0.962	0.117
ITEM0	0.712	0.314	0.719	0.158	-0.083	0.189	0.696	0.110
ITEM0	-1.94	0.529	1.096	0.245	-2.656	0.123	1.072	0.165

مفردات الاختبار	تحليلات النموذج الثلاثي البارمتر				تحليلات النموذج الثنائي البارمتر			
	تقديرات الصعوبة بوحدة اللوجيت	الخطأ المعياري لتقديرات الصعوبة	تقديرات التمييز	الخطا المعياري لتقديرات التمييز	تقديرات الصعوبة بوحدة اللوجيت	الخطأ المعياري لتقديرات الصعوبة	تقديرات التمييز	الخطا المعياري لتقديرات التمييز
ITEM0	-0.856	0.246	1.715	0.342	-1.433	0.310	1.43	0.188
ITEM0	-0.162	0.223	1.541	0.427	-1.025	0.122	1.063	0.160
ITEM0	-0.713	0.302	0.957	0.183	-1.191	0.243	0.923	0.142
ITEM0	-0.828	0.509	0.863	0.233	-2.15	0.099	0.726	0.122
ITEM0	-0.589	0.212	1.447	0.278	-0.994	0.116	1.284	0.191
ITEM0	-0.202	0.098	3.5	1.073	-0.625	0.148	1.955	0.285
ITEM0	-0.486	0.183	1.439	0.22	-0.854	0.360	1.256	0.145
ITEM0	-0.184	0.18	1.466	0.311	-0.62	0.174	1.297	0.188
ITEM0	-0.148	0.213	1.314	0.269	-0.844	0.166	0.981	0.121
ITEM0	0.666	0.233	0.918	0.212	0.06	0.228	0.922	0.145
ITEM0	1.943	0.332	2.772	1.712	2.885	0.366	0.649	0.142
ITEM0	-0.223	0.099	3.077	0.705	-0.553	0.166	2.207	0.285
ITEM0	0.271	0.093	2.191	0.465	-0.013	0.086	1.655	0.224
ITEM0	-0.187	0.121	1.793	0.283	-0.42	0.128	1.656	0.212
ITEM0	0.791	0.093	2.836	0.81	0.56	0.122	1.533	0.231
ITEM0	0.284	0.098	2.332	0.328	1.132	0.143	0.194	0.051
ITEM0	0.842	0.112	2.056	0.486	-0.067	0.113	1.699	0.247
ITEM0	1.032	0.173	1.372	0.337	0.611	0.608	1.6	0.242
ITEM0	1.096	0.17	1.542	0.409	0.64	0.075	1.197	0.188
ITEM0	1.569	0.31	1.092	0.316	0.715	0.173	1.176	0.187
ITEM0	4.361	1.62	0.603	0.224	1.023	0.089	0.934	0.179
ITEM0	2.78	0.784	1.034	0.413	2.542	0.089	0.46	0.115

يتضح من الجدول السابق أن هناك فروق ظاهرية بين تقديرات الصعوبة والتمييز المتناظرة لمفردات اختبار المصفوفات المتتابعة والنتجة عن التحليلات الخاصة بكل نموذج مستخدم بالدراسة، وقد تفسر هذه الفروق الجوهرية بين تقديرات الصعوبة المتناظرة لفقرات الاختبار باختلاف معالم النموذج الثنائي عن معالم النموذج الثلاثي، حيث تعتمد فلسفة تقدير صعوبة الفقرة الاختبارية في هذين النموذجين على قدرات الأفراد المقدرة وفقا للمتغيرات التي يعتمد عليها كل نموذج تحليل حيث يعتمد تقديرات القدرة على معلمي الصعوبة والتمييز في النموذج الثنائي بينما تعتمد تقديرات القدرة للأفراد على معالم الصعوبة والتمييز والتخمين للنموذج الثلاثي، مما قد يؤدي على اختلاف تقديرات صعوبة الفقرة مع اختلاف نموذج التحليل المستخدم.

جدول (3)

دلالة الفروق بين تقديرات الصعوبة المتناظرة لفقرات اختبار المصفوفات المتتابعة الناتجة من تحليلات النموذجين (الثنائي والثلاثي) البارامتر

مفردات الاختبار	مطلق الفروق بين تقديرات الصعوبة المتناظرة المشتقة من نموذجي التحليل	مجموع الخطأين المعيارين لتقديري الصعوبة المشتقين من نموذجي التحليل	دلالة الفروق	مطلق الفروق بين تقديرات التمييز المتناظرة المشتقة من نموذجي التحليل	مجموع الخطأين المعيارين لتقديري التمييز المشتقين من نموذجي التحليل	دلالة الفروق
ITEM0001	3.43	3.243	*	0.147	2.531	
ITEM0002	0.123	3.978		0.14	1.878	
ITEM0003	3.219	2.034	*	1.692	2.01	
ITEM0004	0.645	0.923		0.063	0.572	
ITEM0005	0.407	4.159		0.101	2.268	
ITEM0006	0.275	0.438		0.007	0.428	
ITEM0007	0.93	0.769		0.315	0.524	
ITEM0008	15.459	9.228	*	0.263	2.792	
ITEM0009	0.732	0.843		0.058	0.506	
ITEM0010	0.123	0.196		0.037	0.35	
ITEM0011	0.366	0.217	*	0.399	0.309	*
ITEM0012	1.113	1.626		0.076	1.016	
ITEM0013	0.532	0.969		0.1	0.929	
ITEM0014	16.253	10.191	*	0.272	3.319	
ITEM0016	0.572	0.753		0.074	0.492	
ITEM0017	0.414	0.474		0.068	0.473	
ITEM0018	0.685	0.396	*	0.346	0.337	*
ITEM0019	0.602	0.351	*	0.523	0.315	*
ITEM0020	0.649	0.409	*	0.305	0.305	
ITEM0021	0.75	0.534		0.26	0.35	
ITEM0022	0.237	0.355		0.011	0.39	
ITEM0023	0.181	0.142		0.17	0.34	
ITEM0024	9.824	7.244	*	0.229	2.411	
ITEM0025	1.896	1.434	*	0.155	0.837	
ITEM0026	1.817	1.228	*	0.248	0.737	
ITEM0027	0.665	0.498	*	0.2	0.387	
ITEM0028	0.238	0.451		0.002	0.435	
ITEM0029	0.354	0.3	*	0.11	0.295	
ITEM0030	1.063	0.717		0.238	0.31	

مفردات الاختبار	مطلق الفروق بين تقديرات الصعوبة المتناظرة المشتقة من نموذجي التحليل	مجموع الخطأين المعيارين لتقديري الصعوبة المشتقين من نموذجي التحليل	دلالة الفروق	مطلق الفروق بين تقديرات التمييز المتناظرة المشتقة من نموذجي التحليل	مجموع الخطأين المعيارين لتقديري التمييز المشتقين من نموذجي التحليل	دلالة الفروق
ITEM0031	0.6	0.337	*	0.157	0.279	
ITEM0032	0.903	0.595	*	0.113	0.376	
ITEM0033	0.39	0.259	*	0.151	0.257	
ITEM0034	0.722	0.262	*	0.774	0.233	*
ITEM0035	0.795	0.462	*	0.023	0.258	
ITEM0036	0.716	0.889		0.024	0.525	
ITEM0037	0.577	0.42	*	0.285	0.362	
ITEM0038	0.863	0.389	*	0.478	0.326	*
ITEM0039	0.478	0.53		0.034	0.37	
ITEM0040	1.322	0.875	*	0.137	0.488	
ITEM0041	0.405	0.378	*	0.163	0.357	
ITEM0042	0.423	0.184	*	1.545	0.371	*
ITEM0043	0.368	0.311	*	0.183	0.273	
ITEM0044	0.436	0.302	*	0.169	0.31	
ITEM0045	0.696	0.356	*	0.333	0.264	*
ITEM0046	0.606	0.346	*	0.004	0.258	
ITEM0047	0.942	0.94		2.123	0.75	*
ITEM0048	0.33	0.174	*	0.87	0.36	*
ITEM0049	0.284	0.162	*	0.536	0.293	*
ITEM0050	0.233	0.21	*	0.137	0.301	
ITEM0051	0.231	0.182	*	1.303	0.32	*
ITEM0052	0	1.126		0	0.614	
ITEM0054	0.351	0.166	*	0.633	0.315	*
ITEM0055	0.231	0.199	*	0.456	0.329	*
ITEM0056	0.392	0.278	*	0.175	0.293	
ITEM0057	0.381	0.285	*	0.366	0.302	*
ITEM0058	0.546	0.483	*	0.158	0.352	
ITEM0059	1.819	2.243		0.143	0.738	
ITEM0060	0.539	1.28	*	0.191	0.907	

* الفرق بين تقديري الصعوبة المتناظرة المشتقين من نموذجي التحليل أكبر من أو يساوي مجموع الخطأين المعيارين لهما.

يتضح من الجدول السابق الذي يتضمن نتائج المقارنات بين تقديري الصعوبة المتناظرين لكل فقرة من فقرات اختبار المصفوفات المتتابعة المشتقة من نموذجي التحليل مع مجموع الخطأين المعيارين لهما وجود عدد (37) فرقاً دالاً.بنسبة (62%) من جملة الفروق بما يعني وجود اختلاف جوهري بين تدريجي الصعوبة المشتقين من تحليلات كل من نموذجي التحليل (الثنائي والثلاثي) البارمتر. بما يفيد عدم تبات أو اتساق تقديرات الصعوبة المشتقة من نموذجي التحليل.

كما يتضح من الجدول السابق أيضاً الذي يتضمن نتائج المقارنات بين تقديري التمييز المتناظرين لكل فقرة من فقرات اختبار المصفوفات المتتابعة المشتقة من نموذجي التحليل مع مجموع الخطأين المعيارين لهما وجود عدد (14) فرقاً دالاً بنسبة 23% من جملة الفروق، بما يعني وجود اختلاف جوهري بين تدريجي الصعوبة المشتقين من تحليلات كل من نموذجي التحليل (الثنائي والثلاثي) البارمتر. بما يفيد عدم تبات أو اتساق تقديرات التمييز المشتقة من نموذجي التحليل.

٢- عرض النتائج الخاصة باختبار الفرض الثاني من فروض الدراسة:

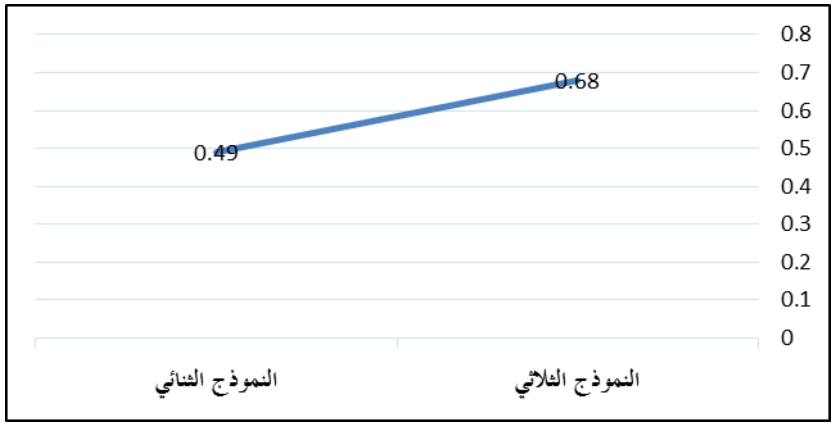
أسفرت إجراءات التحقق الفرض الثاني والذي ينص على:

"لا تختلف الأخطاء المعيارية لتقديرات الصعوبة وتقديرات التمييز لفقرات اختبار المصفوفات المتتابعة" باختلاف نموذج التحليل المستخدم " عن النتائج المبينة بالجدول التالي:

جدول رقم (4)

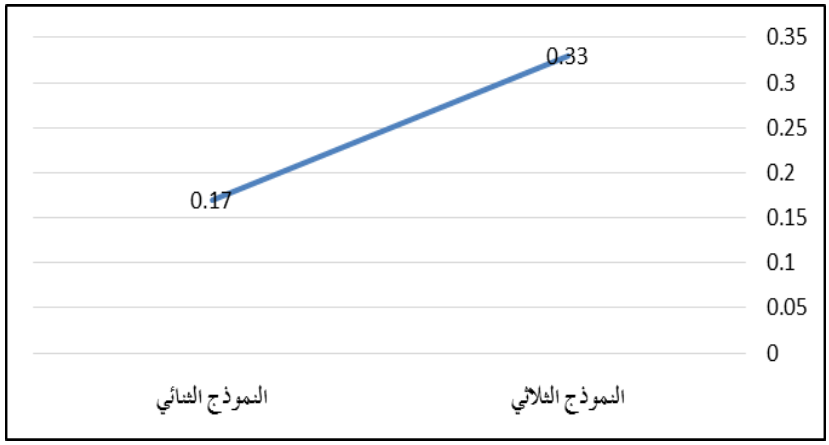
متوسطي الأخطاء المعيارية الناتجة من تحليلات الصعوبة لفقرات اختبار المصفوفات المتتابعة المشتقة من نموذجي التحليل (الثنائي والثلاثي) البارمتر

مفردات الاختبار (٦٠ مفردة)	النموذج الثنائي البارمتر	النموذج الثلاثي البارمتر
مجموع الأخطاء المعيارية تقديرات صعوبة فقرات الاختبار	29.369	40.724
متويط الأخطاء المعيارية لتقديرات الصعوبة	0.48948	0.6787333
مجموع الأخطاء المعيارية تقديرات التمييز لفقرات الاختبار	10.247	20.078
متويط الأخطاء المعيارية لتقديرات التمييز	0.1708	0.3346



شكل (1)

الأخطاء المعيارية الناتجة من تحليلات صعوبة فقرات الاختبار لنموذجي التحليل (الثنائي– الثلاثي) البارمتر



شكل (2)

الأخطاء المعيارية الناتجة من تحليلات التمييز لفقرات الاختبار لنموذجي التحليل (الثنائي– الثلاثي) البارمتر

يتضح من الجدول السابق رقم (5) والشكلان رقمي (1،2) أن متوسط الأخطاء المعيارية لتقديرات الصعوبة المشتقة من تحليلات النموذج الثنائي البارمتر بلغ قيمته (0.49) تقريبا وهي أقل من قيمة متوسط الأخطاء المعيارية لتقديرات الصعوبة المشتقة من تحليلات النموذج الثلاثي البارمتر التي بلغت (0.68) تقريبا، كما يتضح أن متوسط الأخطاء المعيارية لتقديرات التمييز المشتقة من تحليلات النموذج الثنائي البارمتر التي بلغت (0.17) تقريبا، وهي أقل من قيمة متوسط

الأخطاء المعيارية لتقديرات التمييز المشتقة من تحليلات النموذج الثلاثي البارمتر التي بلغت (0.33) تقريبا، بما يعني دقة النموذج الثنائي البارمتر إذا ما قورن بالنموذج الثلاثي البارمتر عند تقديرات الصعوبة الخاصة بفقرات الاختبار

مناقشة نتائج فروض الدراسة:

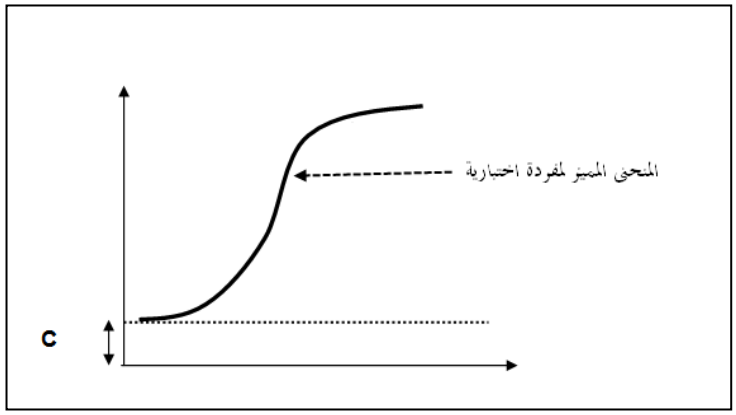
لقد اتسقت النتائج التي توصلت إليه الدراسة الحالية من حيث نتائج الفرض الأول الذي ينص على " لا تختلف قيم تقديرات الصعوبة وتقديرات التمييز لفقرات اختبار المصفوفات المتتابعة باختلاف نموذج التحليل المستخدم" وكذلك الفرض الثاني الذي ينص على "لا تختلف الأخطاء المعيارية لتقديرات الصعوبة وتقديرات التمييز لفقرات اختبار المصفوفات المتتابعة" التي اسفرت عنها النتائج السابقة مع ما توصلت إليه بعض الدراسات التي أجريت في هذا الإطار حيث أظهرت دراسة Van Der Maas, Han L. J.(,2011) دقة النموذج ثنائي البارمتر عن النماذج الأخرى في اتخاذ القرارات بشأن الفروق بين الأفراد في بعض مجالات التحصيل، وهذا ما أكدته الدراسة الحالية، كما ان هناك بعض الدراسات مثل دراسة .,Daniel(,2010) والتي توصلت إلى أن النموذج ثنائي البارمتر أكثر نماذج المفردات دقة في تقدير الخصائص الإحصائية للفقرات الاختبارية، وهذا ما تأكد بالفعل مع النتائج التي تم التوصل إليها من خلال الدراسة الحالية، كما اتسقت هذه النتيجة مع ما توصلت دراسة (Courville,2004)والتي أجريت على عينة من بيانات اختبار الاستعداد المدرسي بالولايات المتحدة الأمريكية، إلى تفوق النموذج ثنائي البارمتر عن الثلاثي البارمتر في دقة تقديرات القدرة واختلاف هذه التقديرات جوهريا عند مقارنة هذين النموذجين، في حيث أظهرت هذه النتيجة تناقضا واضحا مع النتائج التي توصلت إليها دراسة (Raykov,T.,. Dimitrov,D ,Asparouhov,T.,2010) التي أشارت إلى عدم وجود فروق جوهرية بين تقديرات القدرة المشتقة من هذا النموذج الثنائي وتقديرات القدرة المشتقة من النموذج الثلاثي البارمتر.، كما اتسقت هذه النتيجة أيضا مع ما أشارت إليه دراسة (Rizavi, S .; Way, W. D.; Lu, Ying; P. M.; S, M., 2004)، بأن النموذج ثنائي البارمتر أفضل من النموذج ثلاثي البارمتر وجاءت نتائجه أكثر إيجابية، كما اتسقت هذه النتيجة مع ما توصلت

دراسة (Jaio,2003) التي أكدت أن النموذج الثنائي كان أفضل النماذج الثلاث من حيث حصول المفحوصين على تقديرات أدق من القدرة. كما يتضح من النتيجة السابقة أن متوسط الأخطاء المعيارية لتقديرات الصعوبة والتمييز الناتجة عن التحليل باستخدام النموذج الثنائي البارمتر جاء اقل بصورة جوهرية من متوسطي تقديرات الصعوبة والتمييز التي اسفرت عنها التحليلات الخاصة بالنموذج الثلاثي البارمتر مما يشير إلى ان تقديرات المعالم الخاصة بهذا النموذج كانت أكثر دقة من التقديرات التي أسفرت عنها التحليلات الخاصة بالنموذج الثلاثي. وهذا ما يبدو واضحا من التمثيل البياني الذي يشير على أن متوسط الأخطاء المعيارية لتقديرات الصعوبة الناتجة عن التحليلات الخاصة بالنموذج الثنائي البارمتر جاء أقل بصورة واضحة عن متوسط الأخطاء المعيارية لتقديرات الصعوبة الناتجة عن التحليل باستخدام النموذج الثلاثي البارمتر.

وتتسق هذه النتيجة أيضا مع النتائج التي انتهت إليها بعض الدراسات، مثل دراسة (Raykov,T., et,al.,2010) التي انتهت إلى أن النموذج الثنائي البارمتر أعطى أخطاء معيارية للتقدير أقل من نظائرها بالنموذج الأحادي البارمتر، كما أن هذه النتيجة التي تفيد بان النموذج الثنائي البارمتر أسفرت تحليلاته عن أقل أخطاء معيارية لقياس تقديرات القدرة التي تشير إلى النموذج ثنائي البارمتر أكثر النماذج دقة في تقدير القدرة، وقد اتسقت مع النتائج التي توصلت إليها دراسة Van Der& .Han(,2011) التي أثبتت تفوق النموذج الثنائي البارمتر على النماذج الأخرى من حيث دقة تقديرات القدرة وكذلك دراسة (Daniel,,2010)التي أثبتت تفوق النموذج الثنائي البارمتر على النماذج الأخرى في دقة تقديرات القدرة، ودراسة (Courville,2004) التي توصلت إلى دقة النموذج الثنائي البارمتر إذا ماقورن بالنموذج الأحادي والثلاثي، وهذه النتيجة تتسق كذلك مع ما توصلت إليه دراسة (Jaio,2003) التي أشارت إلى أن النموذج الثنائي كان افضل النماذج الثلاث وأقلهم من حيث معدلات الخطأ، كما تتسق هذه النتائج مع ما توصلت إليه دراسة (Aimee,et,al,2008) التي توصلت إلى دقة هذا النموذج الثنائي البارمتر إذا ماقورن بكل من النموذجين الأحادي وكذلك الثلاثي البارمتر.، كما اتسقت هذه النتيجة ايضا مع ما توصلت إليه دراسة (Rizavi, S ; Way, W. D.; Lu, .

(Ying; P. M.; S, M., 2004، التي استهدفت مقارنة نماذج الاستجابة للمفردة في دقة تقديرات القدرة للطلاب المتقدمين لكلية الطب اعتمادا على اختبارات الموائمة التي تقيس القدرة على الاستدلال اللفظي، في ضوء طول الاختبار، أن النموذج ثنائي البارمتر أفضل من النموذج ثلاثي البارمتر وأعطى نتائج أكثر إيجابية، وقد توصلت دراسة (Jaio,2003) التي استهدفت تحديد معدلات الخطأ في تقديرات نماذج الاستجابة للمفردة الثلاثة (الأحادي والثنائي والثلاثي) البارمتر، وفقا للاستجابات ثنائية التقسيم أن النموذج الثنائي كان افضل النماذج الثلاث من حيث دقة تقديرات القدرة، وقد تناقضت النتيجة الحالية مع أظهرته دراسة (DeMars,2001) من أن هناك أفضلية لاستخدام النموذج الثلاثي البارمتر لمعرفة الفروقات الجماعية المبنية على نظرية الاستجابة للمفردة إذا ماقورن بكل من النموذجين (الأحادي والثنائي)، كما توصلت دراسة الوليلي (٢٠٠٢) التي استهدفت تحديد دقة نماذج الاستجابة للمفردة الثلاث (الأحادي والثنائي والثلاثي البارمتر) في انتقاء مفردات اختبار مرجعي المحك في الجبر لدى عينة من طلاب الصف الثالث الثانوي إلى أن النموذج الثلاثي البارمتر هو الأكثر دقة في انتقاء البيانات مقارنة بالنموذجين الآخريين.

وترى الباحثة أن النتيجة التي تتعلق بتفوق النموذج الثنائي البارمتر على النموذج الثلاثي البارمتر من حيث إعطائه أقل خطأ معياري لتقديرات الصعوبة والتمييز قد تعود إلى متغير التخمين الذي يستخدم في الحسابات الخاصة تقديرات القدرة بتحليلات النموذج الثلاثي البارمتر والذي يجرى تقديره من خلال الخط المقارب أسفل المنحنى Lower asymptote والذي يشير ارتفاعة عن نقطة الأصل (c) على مقدار التخمين على المفردة– شكل رقم (3)– (Lord ,1980: 12-13) وقد ترى الباحثة أن هذا الارتفاع يمكن تفسيره بسهولة المفردة الاختبارية لدى أضعف ممتحن أجاب عن هذه المفردة وبالتالي فإن ارتفاع (c) عن القيمة صفر (المقابلة للخط المقارب– المتقطع– أسفل المنحنى المميز للمفردة) قد تعزي إلى سهولة المفردة ولا تعزي لمستوى التخمين لدى الممتحن غير القادر على الإجابة الصواب، مما قد يؤثر في النهاية على دقة تقديرات القدرة التي يمكن الحصول عليها باستخدام النموذج الثلاثي البارمتر.



شكل (3)

قيمة التخمين (c) لمفردة اختبارية

الخلاصة:

توجد اختلافات جوهرية بين تقديرات متغيري الصعوبة والتمييز لفقرات اختبار المصفوفات المتتابعة لرافن كما تبين من خلال نتائج الدراسة الحالية أن النموذج الثنائي البارامتر أفضل من النموذج الثلاثي في تقديمه لأدق تقديرات إحصائية عن النموذج الثلاثي البارمتر.

دراسات مقترحة:

في ضوء النتائج التي تم التوصل إليها من خلال إجراءات الدراسة الحالية يمكن التوصية بإجراء الدراسات التالية:

١- دراسة للكشف عن دلالة الفروق بين النماذج الثلاثة (الأحادي والثنائي والثلاثي) البارمتر في دقة تقدير تقديرات الصعوبة للمفردات الاختبارية.

٢- دراسة للكشف عن مدى اختلاف قيم دوال معلومات الاختبار باختلاف نموذج التحليل المستخدم (أحادي، ثنائي، ثلاثي) البارمتر.

٣- دراسة تأثير اختلاف حجم عينة الأفراد على دقة تقدير المشتق من التحليلات الخاصة بنموذج التحليل المستخدم (أحادي، ثنائي، ثلاثي) البارمتر.

التوصيات:

في ضوء النتائج التي انتهت إليها الدراسة الحالية يمكن اقتراح التوصيات التالية:

١- يجب الاعتماد على النموذج الثنائي البارمتر في التحليلات الخاصة بالمفردات الاختبارية نظرا لدقة تقديراته لمتغيري القدرة والتمييز بالمقارنة بالنموذج الثلاثي البارمتر.

٢- يجب الاعتماد على النموذج الثنائي البارمتر في التحليلات الخاصة بإنشاء بنوك الأسئلة نظرا لدقة هذا النموذج بالمقارنة بالنموذج الثلاثي البارمتر

المراجع:

- أبو حطب، فؤاد (١٩٩٦). القدرات العقلية، القاهرة، مكتبة الأنجلو، ط٥.
- أبو حطب، فؤاد وآخرون (١٩٧٨) .تقنين اختبار المصفوفات المتتابعة على البيئة السعودية :المنطقة الغربية، مكة المكرمة :مركز البحوث التربوية والنفسية.
- أبو هاشم، السيد محمد، (٢٠٠٦). دراسة مقارنة بين النظرية التقليدية ونموذج راش فى اختيار فقرات مقياس مداخل الدراسة لدى طلاب الجامعة، مجلة كلية التربية، جامعة الزقازيق، العدد (٥٢).
- الأحمدي، شرف حامد (٢٠٠٨). أثر اختلاف طول الاختبار وحجم العينة على دقة واتساق تقدير بارمترات المفردة والفرد في نظرية الاستجابة للمفردة، رسالة دكتوراة غير منشورة، جامعة ام القرى، كلية التربية.
- الحكماني، رحاب سعبد، (٢٠٠٧). مقارنة بين النظرية الكلاسيكية للاختبار ونظرية الاستجابة للمفردة في تقدير قدرات الأفراد ومدى استقرار مؤشرات المفردات الاختبارية، رسالة ماجيستير غير منشورة، جامعة السلطان قابوس، كلية التربية.
- القرشي، عبد الفتاح (١٩٨٧) تقنين اختبار رآفن للمصفوفات المتتابعة، الكويت: دار القلم للنشر والتوزيع.
- زكري، علي عبد الله (٢٠٠٩). الخصائص السيكومترية لتطوير اختبار أوتيس لينون للقدرة العقلية مقدرة وفق القياس الكلاسيكي ونموذج ر اش لدى طلبة المرحلة المتوسطة بمحافظة صبيا التعليمية، رسالة دكتوراة غير منشورة، بكلية التربية ,جامعة أم القرى.
- زمزمي،عبد الرحمن معتوق (١٩٩٩) تقنين اختبار المصفوفات المتتابعة لجون رآفن علي الطلاب الصم في معاهد الأمل للمرحلة الابتدائية بالمملكة العربية السعودية، رسالة ماجيستير غير منشورة،جامعة أم القرى، كلية التربية، قسم علم النفس.
- الزيد، جواهر محمد (٢٠٠٧). فاعلية نظرية الاستجابة للمفردة في بناء الاختبارات (إعداد بنك أسئلة في مقرر علم النفس التربوي). رسالة دكتواره غير منشوره.جامعة الإمام محمد بن سعود الإسلامية، كلية العلوم الاجتماعية.
- صفوت فرج (٢٠٠٧).القياس النفسي، القاهرة، دار الفكر العربي.
- العكايلة،عبد الناصر سند (٢٠٠٧). دراسة مقارنة بين النظرية الكلاسيكية والنظرية الحديثة في القياس في كشف وتقدير الخطأ المعياري في اختبارات القدرات المعرفية لطلبة المرحلة الأساسية في الأردن، رسالة دكتوراة غير منشورة، جامعة القاهرة،مصر.
- علام، صلاح الدين محمود (٢٠٠٥). نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي، القاهرة،دار الفكر العربي.
- عليان، خليل (١٩٨٩).معايير الأداء العقلي للأفراد الأردنيين الذين تزيد أعمارهم، عن ١١ عاماً علىمصفوفة رآفن المتتابعة المتقدمة، دراسات

نفسية، المجلد ١٥، العدد ٨،ص ص: ١-١٣٢.
- متيرد، منى ربيع، (٢٠٠٠). دراسة سيكومترية حول تطوير اختبار المصفوفات المتتابعة لرافن باستخدام نموذج راش.رسالة ماجيستير غير منشورة، القاهرة، جامعة عين شمس، كلية البنات للآداب والعلوم والتربية.
- محمد،إبراهيم محمد، (٢٠١٠) استخدام نموذج التقدير الجزئي لتحليل مستويات بنية ناتج التعلم للاختبارات العنقودية ذات الفقرات المترتبة هرميا لدى عينة من طلاب كلية التربية بالمنيا، رسالة دكتوراه غير منشورة مودعة بمكتبة كلية التربية، جامعة المنيا.
- معوض،خليل ميخائيل.(٢٠٠٧). القدرات العقلية.مصر.مركز الإسكندرية للكتاب. ط٤.
- منتصر، شادية عبد العزيز (٢٠٠٢). التصور البصري المجسم،لدى عينة من طالبات الجامعة "قياسه وتنميته، رسالة دكتوراه غير منشورة، كلية البنات للعلوم والآداب والتربية، جامعة عين شمس.
- الوليلي , أسماعيل حسن (٢٠٠٢).دراسة سيكومترية مقارنة لبعض نماذج الاستجابة للمفردة في انتقاء مفردات الاختبارات مرجعية المحك , رسالة دكتوراة غير منشورة كلية التربية,جامعة الأزهر.
- ياسين،عمر صالح (٢٠٠٤). الخصائص السيكومترية لاختبار محكي المرجع في الكيمياء، جامعة ام القرى.طلاب الصف الأول الثانوي العلمي مقدرة وفق النظريتين الكلاسيكة والحديثة للقياس، رسالة دكتوارة غير منشورة، كلية العلوم التربوية، جامعة عمان العربية.
- يماني، أماني عبد الباري (٢٠٠٥). تقنين اختبار أوتيس- لينون للقدرة العقلية المستوى المتقدم الصورة (j) على عينة من طلاب وطالبات المرحلة الثانوية في مدينة مكة المكرمة بالمملكة العربية السعودية، رسالة دكتوراة غير منشورة، مكة المكرمة،جامعة ام القرى.
- **Adedoyin, O. O. (2010):Using IRT Approach to Detect Gender Biased Items in Public Examinations: A Case Study from the Botswana Junior Certificate Examination in Mathematics, Educational Research and Reviews, 5(7):385-399.**
- **Aimee,M,(2008). A comparison of (CAT) scores based on based on different Measurement Models for testlets, Educational and psychological measurement,31(3):371-398.**
- **Azevedo, Caio L.N , Bolfarine, Heleno: Andrade, Dalton F.(2011):Bayesian inference for a skew-normal IRT model under the centred parameterization. Computational Statistics & Data Analysis; Jan, 55(1):353-365.**

- Azevedo,C.L.N. (2009): Some Observations on the Identification and Interpretation of the 3PL IRT Model, Applied Psychological Measurement , 33(7): 89–114.
- Bjorner JB, Kosinski M,Ware JE(2004).Computerized adaptive testing and item banking. In: Fayers PM, Hays RD, editors. Assessing quality of life. Oxford: Oxford University Press.
- Bryant, D. U. (2005). A Note on Item Information in Any Direction for the Multidimensional Three-Parameter Logistic Model. Psychometrika. 70(1):123-144
- Carlson, J.S. & Jensen, M (1981). Beliability of the Raven Progresssive matrices test: Age and ethnic group comparisons , Journal of Consulting and clinical Psychology, v 49, n3, pp: 320- 322
- Courville,T.G.(2004). an empirical comparison of item response theory and classical test theory item/ person statistics unpublished dissertation, office of graduate studies of texas a&w university.
- Daniel, Robert C.; Embretson, Susan E.(2010). Designing Cognitive Complexity in Mathematical Problem-Solving Items. Applied Psychological Measurement, 34(5):348-364.
- De Ayala, R., Dodd, B. & Koch, W. (1990). A simulation and comparison of flexi level and Bayesian computerized adaptive testing. Journal of Educational Measurement, 27, pp.227-239.
- De Ayala, R., Dodd, B., & Koch, W. (1992). A comparison of partial credit and graded response model in computerized adaptive testing. Applied Psychological Measurement, 5,pp.17-34.
- De Ayala,R.J.(2009).The Theory and Practice of Item Response Theory.NY.the Guilford press.
- De Champlain,A(2009): A primer on classical test theory and item response theory for assessments in medical education,Blackwell Publishing Ltd, Medical Education ; 44(1):109–117.
- DeMars,C.(2001). Group differences based on IRTscores:Does the model mater? Educational and psychological measurement,61(1):60-70.

- **Edward, H. (2010): Interpretation of the Three-Parameter Testlet Response Model and Information Function. Applied Psychological Measurement;,. 34 (7):467-482.**
- **Embretson SE, Reise SP (2000). Item Response Theory for Psychologists.London: Lawrence Erlbaum Associates.**
- **Evans, L (1980). Wisc performance scale and Progressive matrices with Deaf children , British: Journal of Educational Psychology ,v50, pt 3, pp: 216- 222)**
- **Fan,a.(2009).Item Response Theory Models:an Empirical Comparison of Their Item Statistics. Educational and psychological measurement, 44 (2):.256-270.**
- **Fumiko Samajima (2010). Estimation of Reliability Using the item Information Function, Applied Psychological Measurement feb., Vol. 18 No. 3 PP.229-244.**
- **Gessaroli ME & De Champlain A. (2005). Test dimensionality: Assessment of. In: Everitt BS, Howell DC, Eds.Encyclopedia of Statistics in Behavioral Science. Chichester:John Wiley & Sons**
- **Hattie,J.:Methodology Review(1985).Assessing Unidimension-ality of Tests and Items Applied Psycho-logical Measurement;, 9 (2):139-164.**
- **Hulin,C.Drasgow,F.& Parson,K.(1983):Item Response Theory: Pplications to Psychological Measure-ment. Hoomewood,Illinois:Dow Jones-Irwin.**
- **Jason L. Meyers, G. Edward Miller , Walter D. Way ,(2009). Item Position and Item Difficulty Change In An IRT-Based Common Item Equating Design , Applied Measurement In Education, 22(3): 38–60,**
- **Linden WJ, Hambleton RK.(2000).Handbook Of modern item response theory.Berlin: Springer.**
- **Loken,e., Kelly, L. (2010). Estimation Of A Four-Parameter Item Response Theory Model, British Journal of Mathematical and Statistical Psychology, 63(2): 509–525.**
- **Lord,f.(1980).Applicatios of Item Response Theory To Practical Testing Proplems.hillsdale,n.j:lawrenve Erlbaum.**

- M. Rose, J.B. Bjorner, J. Becker , J.F. Friesc, J.E. Ware. (2008). Evaluation Of A Preliminary Physical Function Item Bank Supported The Expected Advantages Of The Patient-Reported Outcomes Measurement Information System (Promis) , Journal of Clinical Epidemiology ,61(4): 17- 33.
- Maria Orlando Edelen , Bryce B. Reeve R(2007):Applying Item Response Theory (IRT) Modeling To Questionnaire Development, Evaluation, And Refinement,Qual Life Res 16:5–18.
- Massof, Robert W.(2011):Understanding Rasch and Item Response Theory Models: Applications to the Estimation and Validation of Interval Latent Trait Measures from Responses to Rating Scale Questionn-aires. Ophthalmic Epidemiology; Feb, 18 (1):1-19.
- Noel Y, Dauvier B. (2007).A Beta Logistic Item Response Model For Continuous Bounded Responses. Applied Psychological Measurement,29 (3): 184-203.
- Ponsoda, V. Wise, S., Olea, J. & Revuelta, J. (1997). An investigation of self- adaptive testing in a Spanish high school population. Educational and Psychological Measure-ment, 57, 210-221.
- Raven,j.(1989).The Raven Progressive Matrices,A Review Of National Norming Studies And Ethnic And Socio- Economic Variation Within The United States.jornal of educational measurement, 26(1),:1-16.
- Reckase, M.D.(1999).The Past And Future Of Multidimensional Item Response Theory. Applied Psycho-logical Measurement; 21(3):25-36.
- Rizavi,Saba; Way,Walter D.; Lu,Ying; Pitoniak, Mary; Steffen, Manfred.(2004).Evaluating 1-, 2- And 3-Parameter Logistic Models Using Model-Based And Empirically-Based Simula-tions Under Homogeneous And Heterogeneous Set Conditions. Paper presented at the Annual Meeting of National Council on Measurement in Education (Chicago, IL, Apr 10, 2003). ERIC. ED490464.p35.

- Robert, J and Bock, d.(2003): Item Analysis And Test Scoring With Binary Logistic Models, Scientific Software Inc USA.
- Robert,B.C.(2010).ComparisonsBetween Item Response Theory Models , counseling psychologist,college park,vol.21,no.3,pp234-260.
- Shafiei, Bijan, Tavakol, Samira, Alinia, Leyla, Maracy, Mohammad Reza , Sedaghati, Leyla, Foroughi, Roghayah. (2009):Developing A Screening Inventory Reading Test (Irt) For The Isfahanian Students Of The First To Fifth Grade. (English), Audiology; 17 (2):53-60.
- Shih-Ching Lo , Kuo-Chang Wang and Hsin-Li Chang.(2010): Equal Area Logistic Estimation for Item Response Theory. , Educational Research and Reviews, 7 (4): 233-254.
- Stark, S. Chernyshenko, O. S., Drasgow, F. (2005). "An IRT Approach to Constructing and Scoring Pairwise Preference Items Involving Stimuli on Different, ERIC. ED440146. p89.
- Steven P. Reise, Andrew T. Ainsworth, and Mark G. Haviland (2005):Item Response Theory, Fundamentals, Applications, And Promise in Psychological Research, American Psychological Society, 14(2): (95-101)
- Stocking, M. (1984). Two Simulated Feasibility Study in Computerized Adaptive Testing. (Research Rep. 84-15). Princeton NJ: Educational Testing Service
- Tenko Raykov, Dimiter M. Dimitrov, Tihomir Asparouhov. (2010). Evaluation of Scale Reliability With Binary Measures Using Latent Variable Modeling. Applied Psychological Measurement, 33 (4):242-274.
- Urry, V.(1977).Tailored testingA successful application of latent trait theory, Journal of Educational Measurement, 14, 181-196.
- Valencia, R (1984). Reliability of the Raven Progressive matrices for Anglo and for mexican-American Chiclren , Psychology in the schools ,v21, n1, pp: 49- 52.

- **Van Der Linden, Wim J. Entink, Rinke H, Klein, Fox, Jean-Paul(2010):IRT Parameter Estimation With Response Times As Collateral Information. Applied Psychological Measurement; Jul, 34 (5):327-347.**
- **Van Der Maas, Han L. J.(2011): Cognitive Psychology Meets Psychometric Theory: On the Relation Between Process Models for Decision Making and Latent Variable Models for Individual Differences. Psychological Review; Apr2011, Vol. 118 Issue 2, p339-356.**
- **Walter O, Becker J, Fliege H, jorner JB, Kosinski M, Klapp BF, et al. (2005). Developmental steps for a computer adaptive test for anxiety (A-CAT). Diagnostica;51:88e100;2014–21.**
- **Wang, T. & Vispoel, W. (1998). Properties of ability estimation methods in computerized adaptive testing. Journal of Educational Measurement, 35, 109-135.**
- **Waugh,R.F. (2002). Approaches to Studying Students in Higher Education: A Rasch Measurement Model Analysis. British Journal of Educational Psychology ,70(1): 120-144.**
- **Wauters, K., Desmet, P., Van den Noortgate, W. (2010).Adaptive item-based learning environments based on the item response theory: possibilities and challenges. Journal of Computer Assisted Learning; Dec, 26 (6):549-562.**
- **Wiberg, M.(2004). Classical Test Theory vs. Item Response Theory , An Evaluation of the Theory Test in the Swedish Driving- License Test.UMEA University , Educational and psychological measurement 50(2): 1103-2685.**
- **Zimowski, M. ,Muraki,E.,. Mislevy,R., Bock ,R.(2007): Bilog-MG 3 for Windows, Chicago:Scientific Software Inc.**
- **Zimowski, M., Muraki, E., Mislevy, R. J., & Bock, R. D. (2003). BILOG-MG 3: Item.**