تأثير نماذج التحليل الوغاريتمية (الأحادية والثنائية والثلاثية) البارمتر على تقديرات دوال معلومات المفردات الاختبارية وعلاقتها بدقة تقديرات القدرة (دراسة مقارنة)

د. محمد منصور محمد الشافعي

أستاذ باحث مساعد بقسم البحوث بالمركز القومي

للامتحانات والتقويم التربوي

ملخص الدراسة

استهدفت الدراسة مقارنة تقديرات دوال معلومات مفردات اختبار المصفوفات المتتابعة (العادي) "لجون رافن" وهي (أقصى قيمة للمعلومات – نقاط القدرة المقابلة لأقصى قيم للمعلومات – متوسط قيم المعلومات) الناتجة عن تحليلات ثلاثة نماذج مرتبطة بنظرية الاستجابة للمفردة (الأحادي – الثنائي – الثلاثي) البارمتر ، وعلاقة تقديرات دوال معلومات مفردات الاختبار بدقة تقديرات القدرة الناتجة عن التحليل باستخدام النماذج الثلاثة موضع المقارنة ، وقد استخدمت الدراسة في التحليلات الخاصة بها عينة من طلاب خمس مدارس ثانوية تابعة لمنطقة جنوب القاهرة التعليمية بلغ حجمها (858) طالبا من جميع الصفوف الدراسية (الأول – حتى الثالث) ، وقد استخدمت الدراسة أسلوب تحليل القياس المتكرر في المعالجات الإحصائية الخاصة باختبار فرضيات الدراسة ، وقد توصلت الدراسة إلى أن النموذج ثنائي البارامتر أفضل النماذج الثلاثة في تقديمه لأعلى تقديرات لقيم دوال معلومات فقرات الاختبار يليه النموذج ثلاثي البارمتر ثم أحادي البارمتر ، وهي نفس النتائج التي تم التوصل إليها فيما يتعلق بمتغيري الخطأ المعياري لتقدير القدرة وثبات تقدير دوال المعلومات وذلك كمؤشرين لدقة تقدير قدرات الممتحنين ، حيث تبين أن النموذج ثنائي البارمتر أدت تحليلاته إلى أقل تقدير للأخطاء المعيارية لتقديرات القدرة كما أن تقديراته لثبات قيم معلومات المفردات كانت اعلى من مثيلاتها المقدرة بالنموذجين الآخرين ، كما توصلت الدراسة لوجود علاقة بين دقة تقديرات القدرة ودوال معلومات مفردات الاختبار، حيث اتضح أنه مع إزدياد قيم (متوسط) معلومات مفردات الاختبار كان هناك ازدياد في دقة تقديرات القدرة الناتج عن التحليل باستخدام نموذج الاستجابة للمفردة.

تأثير نماذج التحليل الوغاريتمية (الأحادية والثنائية والثلاثية) البارمتر على تقديرات دوال معلومات المفردات الاختبارية وعلاقتها بدقة تقديرات القدرة (دراسة مقارنة)

د. محمد منصور محمد الشافعي

أستاذ باحث مساعد بقسم البحوث بالمركز القومي

للامتحانات والتقويم التربوي

المقدمة والخلفية النظرية

لقد حققت نظرية الاستجابة للمفردة كأحد الاتجاهات الحديثة للقياس وما ارتبط بها من نماذج رياضية لوغاريتمية طفرة هائلة ومتطورة في مجال القياس النفسي والسلوكي ، فلقد اثبتت العديد من الدراسات الأجنبية والعربية تفوق هذه النظرية في تحقيق الموضوعية المامولة منها في هذا المجال إذا ماقورنت بالنظرية الكلاسيكية ،(De Champlain, 2010: 109) ، (Romanoski& ,2007 Njriu) كما أن هذه النظرية تعتمد على افتراضات وشروط أقوى من الافتراضات والشروط ذات الصلة بالنظرية الكلاسيكية التي هيمنت على مجريات وإجراءات القياس السلوكي والتحليلات الخاصة بأدوات القياس لفترة ليست بالقصيرة من الزمان. (Uttaro & leHman,1994:41) (Rotou,2002:131)

كما أثبتت العديد من الدراسات العربية والأجنبية المعنية بهذا المجال فاعلية استخدام نظرية الاستجابة للمفردة (IRT) ونماذج 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 أثبتت كذلك تميز هذه الادوات من حيث تمتعها بخصائص الموضوعية التي تتسم بها تلك الأدوات المستخدمة في مجال القياس الفيزيقي (Wauters , Van den Noortgate,2010) ، (إبراهيم ،٢٠١٠) ، (2010) Tavakol, Shafiei, (Jason , Edward, Walter ,2009). &,Fox,EntinkVan Der Linden,

(Edelen , ، 2009)Foroughi, , Sedaghati, , Maracy,(Mohammad , Alinia, ، (zickar.1998:104) , (Rouse,et al.,1999:282) , Bryce. Reeve ,2007) (Wainer ,1989:154),(Linn,1990:115)

ونتيجة لتميز هذه النظرية وكذلك النماذج الرياضية المرتبطة بها والنجاحات التي حققتها في هذا المجال 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) Reeve (Edelen , Bryce ,2005) (Stark, Chernyshenko & Drasgow ,2007) (يماني،٢٠٠٥)، , (Wiberg , 2004) , (Molenaa,1996)(Gumpel,et.al.1998) (McHorney,2002) (Raudenbush,2003) (Ackerman,1994),

هذا وقد 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 دالة المعلومات والذي يمكن تقديره لكل مفردة من مفردات الاختبار وكذلك للاختبار ككل، حيث يستخدم هذا المؤشر للكشف عن كمية المعلومات المتصلة بالقدرة المقاسة لدى متلقي الاختبار عند مستويات مختلفة من هذه القدرة (Hambleton& swaminathan,1985),(chernyshenko,2001)، وتختلف أساليب وصيغ التقديرات الخاصة بهذا المؤشر 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 هذا المؤشر ، ومما قد يؤدي في النهاية إلى اختلاف دقة تقدير دوال المعلومات التي يمكن استخلاصها من مفردات الاختبار حول قدرات الممتحنين وفقا لاختلاف نموذج التحليل المستخدم (أحادي – ثنائي – ثلاثي) البارمتر .

ونظرا لازدياد وتنامي استخدام التطبيقات الخاصة بنظرية الاستجابة للمفردة ، في الاونة

الأخيرة على الصعيد العالمي وبصورة خاصة على الصعيد العربي في مختلف مجالات القياس التي تهتم ببناء وإعداد ادوات جديدة متنوعة لجمع البيانات ، او إعادة تطوير ادوات شائعة الاستخدام ، (Edward,2010) وكذلك إنشاء بنوك الأسئلة (Njriu & Romanoski ,2007) ونظرا لإعتماد الإجراءات والتحليلات الخاصة بإعداد هذه الأدوات على مؤشر دالة المعلومات الذي قد يختلف تقديره باختلاف النماذج الرياضية المختلفة المرتبطة بنظرية (IRT) فقد دعت الحاجة إلى ضرورة إجراء هذه الدراسة التي تستهدف عقد مقارنة بين تقديرات مؤشر دوال المعلومات للمفردات الاختبارية (باستخدام اختبار جون رافن العادي) التي يمكن التوصل إليها باستخدام ثلاثة نماذج شائعة الاستخدام في هذا المجال وهي النماذج (أحادية – ثنائية – ثلاثية)البارمتر، وذلك بهدف التعرف على أي النماذج الثلاثة التي يمكن ان تعطي اعلى و افضل تقدير لهذا المؤشر ، وبالتالي يمكن من خلال النتائج المتوقعة من هذه الدراسة تحديد أي من هذه النماذج يمكن الاعتماد عليها مستقبلا وفقا لكفاءتها في تقديم افضل وادق معلومات عن قدرات الأفراد في حال خضوعها لإجراءات و عمليات 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., 2009) , (Urry ,1977)وتقدر بارمترات التمييز في التحليلات الخاصة بنماذج التحليل المرتبطة بنظرية الاستجابة للمفردة بظل الزاوية التي يصنعها لمماس 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& et al. 1997,213) ، (De Ayala ,1992:516)

، (De Ayala & et al. ,1990) ، (McBride & Martine ,1983)

(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).

(د) بارمترات المعلومات: Information parameters

ترتبط بارمترات المعلومات بدالة المعلومات Information function وهي تتمثل في

مقدار ما تعطيه المفردة من معلومات عند مستويات القدرة المختلفة. (Stocking ,1984 :4)، فمفهوم المعلومات مثل مفهوم دالة الاستجابة على الاختبار، هو لب نظرية الاستجابة للمفردة . والمعلومات في المصطلح السيكومتري أو الإحصائي تعني كيف يقدر شيء ما بصورة دقيقة . فلقد عرف فيشر Sir Ronald Fisher المعلومات الإحصائية بأنها مقلوب الخطأ المعياري (De Ayala,2009,29)، في نظرية الاستجابة للمفردة فإنه من الممكن تحديد مقدار المعلومات التي يمكن أن توفرها المفردة عند كل نقطة خلال متصل (θ)، وكلما كان هناك معلومات أكثر (أعلى) عند مستوى محدد (θ)كلما كان هناك مقياس أدق عند نفس المستوى، وبالتالي فإن أحد مميزات نظرية الاستجابة للمفردة هي القدرة على تحديد أي المفردات تقيس بصورة أفضل عند مستوى (θ)بعينه (Edward H., 2010)

و يعكس الميل slop في دالة الاستجابة على الاختبار عند أي نقطة خلال المنحنى كيف تمايز المفردة تماما بين مستويات (θ) المتجاورة رياضيا فإن هذا الميل يوصف بأنه ميل خط المماس أو ظل الزاوية لدالة الاستجابة على الاختبار عند مستوى بعينه من (θ)، وفي حساب التفاضل و التكامل فإنها معروفة كأول اشتقاق للدالة التي تصف دالة الاستجابة على الاختبار ، وبصورة عامة فإن مقدار المعلومات الذي يمكن أن توفره المفردة عند مستوى معين من (θ) يمكن أن يمثل كالتالي)Fumiko Samajima(2010:,

$$I(\theta) = (\text{Squared IRT Slope})/(\text{Conditional Variance}).$$

وتعرض دالة معلومات الاختبار المقدار الأقصى من المعلومات و الذي يكون متاحا من الاختبار المؤلف من مجموعة بعينها من المفردات. ولتحقيق هذه المعلومات القصوى فإن المفردات يجب أن تصحح بطريقة توزن نفسها مثاليا. فإذا استخدم برنامج للحاسب لتقدير مستويات (θ) مستخدما أسلوب "الارجحية القصوى" أو "بييز" فإن هذه التقديرات تكون موزونة مثاليا و يمكن استخدام دالة معلومات الاختبار بصورة مباشرة كمؤشر لمقدار المعلومات المتاح لكل مستوى (θ) Bryant, (2005) (Lord, 1980 :74) ومن المعروف أن برنامج BILOG-MG3 المستخدم بالدراسة الحالية يتوافر به تحليلات كل من أسلوبي "الارجحية القصوى" و "بييز".

(Michele Zimowski, Eiji Muraki, Robert J., Mislevy and R. Darrell Bock,2003)

نظرية الاستجابة للمفردة وتقدير دالة المعلومات

تعتبر دالة المعلومات المؤشر الأساسي لدقة القياس في نظرية الاستجابة للمفردة (Weiss, 1982). وقد اشتق هذا المؤشر كدالة لمستوى القدرة المقدر للمفحوص وبارامترات المفردات الاختبارية (مثل التمييز والتخمين) ، (De Champlain, 2009) فإذا أعطت المفردات تمييزا عاليا عند مستوى قدرة معين؛ فإن معلومات الاختبار سوف تكون عالية. أما إذا أعطت تمييزا منخفضا فإن المعلومات ستكون عندئذ منخفضة والمفردة التي يكون بها بارامتر (a) عاليا، وبارامتر التخمين (c) منخفضا توفر معلومات أكثر من (a) المنخفضة ، وكذلك (c) المرتفعة (Chang & (Ying, 1999) (Drasgow, F. ,2005). Stark, S. Chernyshenko, O. S., ,1984) (Stocking . (Vispoel, 1993)

ونظرا للاعتبارات السابقة والتي تتعلق بحساسية دالة المعلومات لكل من بارمترات الصعوبة والتمييز، والتخمين فإن الباحث الحالي يرى أن الاعتماد على اختبار المصفوفات العادي لرافن في دراسته الحالية يلبي الهدف الرئيس لهذه الدراسة نظرا لاعتباره 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 (60) مفردة اختبارية هي قوام هذا الاختبار (متيرد، ٢٠٠٠، ١٢١)، وبالتالي تكون هناك إمكانية لإبراز الفروق بين نماذج التحليل المستهدفة من الدراسة فيما يتعلق بالبارمترات الخاصة بقيم دوال المعلومات التي تعتمد إلى حد كبير على هذه النوعية من البارمترات.

ويمكن تقدير دوال معلومات الاختبار، باستخدام النماذج 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 مؤشرات دوال المعلومات لكل من المفردة والاختبار باستخدام نماذج نظرية الاستجابة للمفردة المبينة فيما يلي:

تقديرات نماذج الاستجابة للمفردة لدوال المعلومات:

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) وهي (Hambelton&Swaminathan ,1985:46) Bryant, , (2005) (علام، ٢٠٠٥):

١- النموذج أحادي البارامتر : One-Parameter Logistic Model

يعتمد هذا النموذج على فلسفة الاحتمالات والتي تتناسب مع طبيعة الظاهرة السلوكية كونها ظاهرة احتمالية وليست تحديدية (Embretson SE, Reise SP,2000) . حيث يرى "راش" أن استجابة الفرد (v) على المفردة (i) وهى (X_{vI}) تُحدّد وفقا لمتغيرين أساسيين هما متغير قدرة الفرد (θ_v)، ومتغير صعوبة المفردة (b_i)، و أنه إذا تزايدت قدرة الفرد (θ_v) عن مستوى صعوبة المفردة (b_i) فإن الإحتمال الأكبر حينئذ هو حدوث الاستجابة الصواب، أى أن (X_{vI}=1) أما إذا حدث العكس فمن المتوقع حدوث الاستجابة الخطأ (X_{vI}=0)، وحيث إن الظواهر السلوكية تتسم بعدم الثبات، فإن هناك صعوبة للتنبؤ بشكل قاطع فى شكل الاستجابة الصادرة عن الفرد، وبالتالي فإن العلاقة بين قدرة الفرد وصعوبة المفردة يجب ألا تكون علاقة قاطعة، بل هي علاقة احتمالية(Nakamura,2006) (lai,et al.2005) (Anzaldua,2005) وقد توصل "راش" إلى الصورة الرياضية لهذا النموذج الموضحة فيما يلى (De Ayala,2009:11-19)، (Hambelton ,1985:46) :

$$p(x_{vi} = x \,/\, \theta_v b_i) = \frac{e^{(\theta_v - b_i)}}{1 + e^{(\theta_v - b_i)}}$$

حيث x =(0, 1) ، x هى استجابة الفرد على المفردة، θ_v قدرة الفرد، b_i صعوبة المفردة.

e = Exp هو أساس اللوغاريتم الطبيعى ويساوى (2.273).

$P(x_{vi})$ احتمال حدوث الاستجابة الصواب أوالخطأ.

ومن الخصائص التي يتصف بها نموذج "راش" أنه يتعامل مع عدد أقل من متغيرات المفردة كما يفترض ان جميع المفردات تمتلك قوة تمييز متساوية (Massof, , 2011)، وان تخمين الإجابة عنها يكون في أضيق صورة.

(Hambelton ,1985:46) ، (Anzaldua,2005) ، (Nakamura,2006)

ويستخدم هذا النموذج فى تحليلات برنامج ، BILOG-MG المستخدم فى الدراسة الحالية أيضا.

دالة المعلومات للنموذج أحادي البارمتر

وتقدر دالة المعلومات للمفردة في هذا النموذج باستخدام الصيغة التالية Linden& (Hambleton, 2000)

:( ) Edward ,2010

$$I_i(\theta) = P_i(\theta)\, Q_i(\theta)$$

$P_i(\theta)$ هو احتمال حدوث الإجابة الصواب على المفردة (i) للفرد الذي له مستوى القدرة (θ) حيث:

، $Q_i(\theta)$ هو احتمال الاستجابة الخطأ حيث : $Q_i(\theta) = 1 - P_i(\theta)$ ، (θ) هي مستوى القدرة للفرد ذات الاهتمام

وقد استخدم نموذج راش في أعداد ضخمة من الدراسات والتي اكدت دقة التقديرات الخاصة بالمؤشرات الإحصائية التي يمكن الحصول عليها باستخدام هذا النموذج ومن بين هذه الدراسات دراسة فان (Fan,2009) والتي استهدفت مقارنة النماذج الثلاثة لنظرية الاستجابة للمفردة من حيث خصائص فقرات الاختبار على عينة من الطلاب المتقدمين لاختبار تكساس للمهارات الأساسية، وقد انتهت الدراسة إلى تفوق هذا النموذج على النموذجين الآخرين الثنائي والثلاثي البارمتر. كما توصلت دراسة (Robert,2010) على عينة من بيانات اختبار القبول

للجامعات الأمريكية أن نموذج راش الاكثر دقة في تقدير كل من القدرة والصعوبة مقارنة بالنموذج الثلاثي البارمتر.

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 و هذا النموذج يسمى نموذج برينبوم Birnbauum (Hambelton ,1985:46):

ويلاحظ من الصيغة الرياضية الخاصة بالنموذج أنها تتشابه مع الصيغة الرياضية و الخاصة 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 و قيمته حوالي (1.7) أما (a_i) فهو معامل تمييز المفردة و يتراوح مداه من (−2 إلى +2) وحدة معيارية على افتراض أن توزيع القدرة يتراوح على الفترة المغلقة من (−3 إلى +3)

دالة المعلومات للنموذج ثنائي البارمتر:

وتقدر دالة المعلومات لمفردات الاختبار في هذا النموذج بالصيغة التالية)
Linden& Hambleton, 2000)
(,2010):Fumiko Samajima

$$Ii\ (\theta) = ai^2\ Pi\ (\theta)\ Qi\ (\theta)$$

حيث ai هو متغير تمييز المفردة (i) ،

، $Pi\ (\theta)$ هو احتمال حدوث الإجابة الصواب على المفردة (i) ؛ للفرد الذي له مستوى القدرة (θ) حيث:

، $Qi\ (\theta)$ هو احتمال الاستجابة الخطأ حيث :

$Qi\ (\theta) = 1 - Pi\ (\theta)$,

، (θ) هي مستوى القدرة للفرد ذات الاهتمام

وقد توصلت بعض الدراسات إلى ان 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 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 ، 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 الاختبارية ثنائية التقسيم كما أعطى النموذج أخطاء معيارية للتقدير أقل من نظائرها بالنموذج أحادي 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. كما توصلت دراسة (Rizavi, S ., S, M.; P. M; Lu, Ying; Way, W. D..; 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) و تفترض 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( ) :

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

معادلة رقم (١) لحساب احتمالية الاستجابة في النموذج ثلاثي البارامتر

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)

دالة المعلومات للنموذج الثلاثي البارمتر

وتقدر دالة المعلومات بنود الاختبار في هذا النموذج بالصيغة التالية Edward H. (2010)

:(Linden& Hambleton, 2000)

$$I_i(\theta) = a^2\left[\frac{Q_i(\theta)}{P_i(\theta)}\right]\left[\frac{P_i(\theta)-c^2}{(1-c^2)}\right]$$

حيث $Pi\ (\theta)$ هي احتمال حدوث الاستجابة الصواب للفرد الذي قدرته (θ) على المفردة (i) ،

C تقديرا متغير التخمين على المفردة ، a تقديرا لمتغير تمييز المفردة

، Qi هي احتمال حدوث الاستجابة غير الصحيحة حيث : $Qi = 1.0 - Pi(\theta)$

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 (٢٠٠١) 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 ان النموذج الثلاثي البارمتر هو الأكثر دقة في انتقاء البيانات مقارنة بالنموذجين الآخرين.

ويتضح 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 الاكثر دقة في تقدير كل من القدرة والصعوبة مقارنة بالنموذجين الآخرين الثنائي و الثلاثي 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 ، ٢٠٠١) أن النموذج الثلاثي البارمتر هو النموذج الأكثر دقة في انتقاء البيانات مقارنة بالنموذجين الآخرين ، مما يشير إلى تعارض النتائج فما يتعلق بدقة تقديرات معالم المفردات الاختبارية وكذلك معالم الأفراد (متمثلة في تقديرات القدرة) باختلاف النموذج المستخدم ، وعلى الرغم أن وجهة الدراسة الحالية لاتتجه نحو حسم هذا الخلاف فيما يتعلق بدقة تقدير معالم المفردات الاختبارية (الصعوبة والتمييز)، إلا أن الدراسة الحالية تتجه نحو الكشف عن دقة تقديرات قيم دوال معلومات المفردات الأختبارية التي تعتمد بطبيعة الحال في تقديرها لقيم هذه الدوال على متغيرات (الصعوبة والتمييز) وحيث ان نتائج الدراسات المشار إليها آنفا أشارت لوجود اختلافات في تقديراتها وفقا للنموذج المستخدم ،ولم يجد الباحث (في حدود علمه) بعد مسحه للتراث أية دراسات أجنبية أو عربية تهتم بتقصي دقة تقدير دوال معلومات بنود الاختبار لأي نموذج من النماذج المرتبطة بنظرية (IRT) ، وهذا مما أثار عدد من التساؤلات لدى الباحث الحالي : هل إذا كانت هناك اختلافات بين هذه النماذج في تقديراتها لقيم دوال معلومات مفردات الاختبار فهل تؤثر قيم دوال المعلومات التي يمكن تقديرها من خلال نموذج الاستجابة المستخدم (في ضوء الاختلافات المتوقعة على دقة تقديرات القدرة لدى الأفراد؟، وما النموذج الذي يمكن ان يعطي أكبر قيمة لمعلومات المفردات وما تأثير ذلك على تقديرات القدرة للأفراد؟ ، هذا ماتحاول ان تجيب عنه الدراسة الحالية.

دالة الاختبار للمعلومات Test Information Function:

تقدم دالة الاختبار للمعلومات وسيلة قوية لمصمم الاختبار يمكن من خلالها التعرف على خواص الاختبار وكيف يقدم الاختبار أفضل قياس عبر متصل السمة أو القدرة المستهدف قياسها من الاختبار، فهي تعتبر أداة تساعد في تقدير مدى الدقة المقدمة للاختبار عند أي مستوى للسمة أو القدرة، حيث يدل أقصى ارتفاع لمنحنى دالة المعلومات للاختبار عن مستوى معين من السمة أو القدرة على أن هذا الاختبار يكون أكثر دقة للقياس عند هذا المستوى ، وتقدر دالة معلومات الاختبار من تقديرات دوال معلومات المفردات التي يتألف منها الاختبار (Rose& et al. ,2008) (Hullin &. et al. ,1983:54-55) (Bjorner JB,2004) (Verkuilen,2006) (Hameblton,& Traub,1974: 273-281)

ويمكن الحصول على ثلاثة بارمترات لدوال معلومات مفردات الاختبار باستخدام بالنماذج الثلاثة السابقة من خلال التحليلات الخاصة ببرنامج BILOG-MG بيانها فيما يلي (Verkuilen,2006):

- أقصى معلومات للمفردة الاختبارية Maximum Information
- مستوى القدرة المناظر لأقصى معلومات للمفردة. Point of Maximum Information
- متوسط معلومات المفردة الاختبارية عند المستويات المختلفة من القدرة. Information Average

يتضح من العرض السابق لدوال معلومات مفردات الاختبار الاختلاف الواضح للصيغ الرياضية الخاصة بتقدير دوال المعلومات باختلاف النموذج – أحادي او ثنائي أو ثلاثي البارمتر – حيث يعتمد كل نموذج على متغيرات تختلف عن المتغيرات التي يعتمد عليها في النموذجين الآخرين ، حيث يعتمد النموذج الأحادي البارمتر على متغير صعوبة المفردة فقط، في حين يعتمد النموذج الثنائي البارمتر على متغيري الصعوبة والتمييز ، ويعتمد النموذج الثلاثي البارمتر على متغيرات الصعوبة والتمييز والتخمين ، ونظرا لأن كمية المعلومات التي يمكن أن تقدمها المفردة حول مستوى قدرة الفرد المستجيب للمفردة قد تتأثر كما تبين سابقا بمتغيرات التمييز والتخمين فإن ذلك قد يؤدي إلى اختلاف تقدير قيم دوال المعلومات ، وبالتالي قد يؤدي بدوره إلى اختلافات في تقدير قدرات الأفراد ،التي يمكن الحصول عليها من التحليلات الخاصة بتلك النماذج مما يستوجب الكشف عن تأثير تلك الاختلافات على دقة تقديرات القدرة الناتجة من التحليلات الخاصة بالنماذج الثلاث وهذا ما تستهدفه الدراسة الحالية من خلال بعض المؤشرات التي يمكن من خلالها الكشف عن دقة تلك التقديرات وهي:

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 المؤشرات Samajima (2010,):

- **مؤشرات عامة حيث تعطي قيمة مختصرة عن دقة التقدير للاختبار ككل:**
  - وتتمثل في تقديرات الثبات ومؤشر التقارب والدقة والكفاءة.
- **مؤشرات شرطية حيث يتم حسابها عند كل نقطة من القدرة:**
  - وتتمثل في الجذر التربيعي لمتوسط مربعات الخطأ Root Mean Square Error (RMSE)
  - التحيز Bias
  - الخطأ المعياري لتقدير القدرة Standard Error of Estimation (SEE)

أولا المؤشرات العامة لدقة تقدير القدرة :

(أ) الثبات: Reliability

تؤكد نظرية الاستجابة للمفردة على دقة القياس المشروطة conditional measurement حيث تسمح بتحديد دقة القياس عند كل مستوى من القدرة باستخدام دالة فيشر للمعلومات، $I(\theta)$ وتشير دالة المعلومات إلى دقة القياس الشرطي على مستوى القدرة الكامنة المقاسة Samajima (2010,). وبالتالي فإن دقة القياس يمكن تحديدها للمستويات المنخفضة والمرتفعة على حد سواء، باستخدام ثبات قيم دوال معلومات مفردات الاختبار Meijar & Nering (1999:191,) ، و يرى كل من(Nicewander & Thomasson, 1999) (Divgi ,1989) 146:أيضا أن معاملات الثبات مفيدة في تطبيقات نظرية الاستجابة للمفردة لعدة أسباب من بينها :

- معاملات الثبات يمكن تفسيرها كمربع الارتباط بين الدرجة الملاحظة والسمة الكامنة، أي أنه يمكن استخدامها لتقيس قوة العلاقة بين المتغيرات الكامنة والملاحظة.
- تعتبر تقديرات معاملات الثبات دوال بسيطة لدالة المعلومات في نظرية الاستجابة للمفردة، فثبات تقديرات القدرة في الاختبار الذي أعد في ضوء نظرية الاستجابة للمفردة يمكن أن يقدر عن طريق إعادة الاختبار أو البنوك المتكافئة، غير أن هذه الأساليب قد تكون مكلفة وتستهلك وقتا كبيرا ولهذا فإن الأسلوب الأمثل لتقدير الثبات هو المشابه لمعامل ألفا أو (كيودر ريتشاردسون 20) وهذا الأسلوب له خصائص ثلاثة هي Nicewander & Thomasson

(,1999:239) :

١- أنه يمكن تقدير الثبات بناء على تطبيق اختبار لمرة واحدة فقط.

٢- أنه يستخدم معاملات الثبات المرتبطة بدالة المعلومات في نظرية الاستجابة للمفردة.

٣- أنه قابل للاستخدام لكل من تقديرات القدرة بالأساليب البييزية Bayesian techniques (أسلوب أوين ، أسلوب البعدي المتوقع ، أسلوب البعدي الأقصى) وأساليب الاحتمال الأقصى Maximum Likelihood Estimation techniques.

بناء على ما سبق؛ فقد أجريت بعض الدراسات لمعاملات الثبات في سياق الاختبار الذي اجري بناؤه باستخدام نظرية الاستجابة للمفردة، كما قارنت دراسات أخرى بين دقة تقديرات القدرة إعتمادا على مؤشر الثبات بين الاختبارات الورقية الاختبارات التواؤمية المحوسبة حيث قارن (Moreno & Segall, 1997) بين النسختين الورقية والتواؤمية المحوسبة لبطارية الاستعدادات المهنية فوجدا أن تقديرات الثبات لكل اختبار فرعي من البطارية المحوسبة كانت مساوية للصورة الورقية أو أعلى منها .

و تجدر الإشارة إلى أن معاملات الثبات باستخدام معامل ألفا على سبيل المثال يمكن أن يسمح بمقارنة مباشرة بين دقة تقديرات القدرة، وهناك دراسات عرضت للعلاقة بين الثبات التقليدي ومقاييس المعلومات في نظرية الاستجابة للمفردة مثل دراسة (Thissen, 1990) الذي ناقشت الأساليب العامة لحساب ثبات درجات الاختبارات، وتوصلت إلى عدم وجود فروق بين معاملات الثبات المشتقة من تقدير الاحتمال الأقصى والأخرى من التقدير البييزي.

أما في سياق نظرية الاستجابة للمفردة فقد اقترح (Mokken, 1997)معاملات ثبات لاستخدامها كدوال للاستجابة على المفردة، ويؤكد (Meijar & Nering 1999:192)أن لهذه الأساليب خصائص إحصائية أفضل من مؤشرات الثبات التقليدية مثل معامل ألفا كرونباخ وذلك عند استخدامه في سياق نظرية الاستجابة للمفردة، كما قام (Nicewander & Thomasson 1999)بوضع ثلاث معاملات ثبات جديدة باستخدام تقديرات الاحتمال الأقصى والتقدير البييزي، وتعتبر هذه التقديرات دوال لدالة معلومات الاختبار ويسهل مقارنتها كذلك مع مؤشرات الدقة الاخرى، قد أثبتت الدراسة أن تقديرات الثبات الثلاثة لم تختلف عن معاملات الثبات الحقيقية بأكثر من (0.02)

(ب) مؤشر التقارب Convergence:

يمثل مؤشر التقارب نسبة الحالات التي وفر لها إجراء أسلوب الاحتمال الأقصى تقدير قدرة جيد، وتكون درجة القدرة مقاربة convergent إذا لم تساوي موجب أو سالب ما لا نهاية ($\pm \infty$) infinity. ولها خطأ معياري مساو لواحد أو أقل، ويتم التعرف على التقارب بالرسم من خلال المحور الأفقي الذي يمثل القدرة من (÷ 2.5 إلى +2.5) مثلا والمحور الرأسي يمثل النسبة المئوية للتقارب من (0 إلى 100)، ويكون الاختبار دقيقا كلما ارتفعت النسبة المئوية للتقارب عند كل نقطة من القدرة (Vispoel ,1993: 128-130)

(جـ) الدقة Fidelity :

وتعرف على أنها العلاقة بين القدرة المقدرة والقدرة الحقيقية ($r\theta\hat{\theta}$). ويمكن أن تمثل بالرسم عن طريق المحور الأفقي الذي يمثل طول الاختبار والمحور الرأسي الذي يمثل قيم الدقة كدالة لطول الاختبار . (Vispoel, Wang & Bleiler, 1997: 48 -49) ، فالعلاقة المرتفعة تعني دقة عالية (Wang & Vispoel 1998:120) ومربع الارتباط بين القدرة المقدرة والقدرة الحقيقية يعطي مؤشرا للثبات تماما كما في العلاقة بين الدرجة الملاحظة والدرجة الحقيقية التي تمثل ثبات الاختبار في النظرية التقليدية (Vispoel ,1993: 128) .

(د) كفاءة الاختبار Efficiency:

تعني كفاءة الاختبار انه الاختبار الذي يعطي أفضل معلومات ممكنة عن المفحوص بأقل عدد ممكن من المفردات.(Vispoel et al., 1994) (Wang & Vispoel 1998) فالمفردات الأقل تعني كفاءة أعلى، وكلما زاد طول الاختبار قلت الكفاءة (Rocklin 1994:9)بافتراض تثبيت الخطأ المعياري لتقدير القدرة.

ثانيا: 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)، و يوضح 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 كما يحسب الخطأ المعياري لقياس تقديرات القدرة باستخدام قيم دوال المعلومات أيضا حيث يحسب عن طريق الجذر التربيعي لمقلوب دالة المعلومات. وفي أساليب تقدير القدرة باستخدام الاحتمال الأقصى؛ و الخطأ المعياري يرتبط بصورة مباشرة مع ثبات الاختبار الذي يعد بدوره مربع الارتباط بين الدرجة المقدرة والدرجة الحقيقية ويمكن حساب الخطأ المعياري بالمعادلة الآتية من :(Wang & Vispoel, 1998:119)

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

(ج) التحيز Bias

ويعرف بأنه الفرق الواقع بين متوسط الدرجة المقدرة والدرجة الحقيقية. (Vispoel et al., 2010) Adedoyin,.1999), (Wang & Hanson 1999,263), (، ويمكن أن يعبر عنه كما يلي :

التحيز = مجموع (القدرة المقدرة – القدرة الحقيقية)/عدد المفحوصين عند مستوى القدرة الحقيقي المقدم.

وتكون قيمة التحيز مناسبة عندما تقترب من الصفر بدون زيادة في الخطأ المعياري (Wang & Hanson, 1999:264) ، فإذا انخفض فيكون بخسا للتقدير underestimate وإذا ارتفع فيكون إفراط للتقدير overestimate (Vispoel, 1993:130)، (Weiss, 1982:482). ويعد التحيز في تقديرات القدرة مشكلة في اختبارات الإجازة أو الشهادة أو أي اختبار تصنيفي يستخدم درجة فاصلة cut score ففي هذه الحالة يمكن أن يؤثر التحيز في الدرجة الفاصلة، وبصفة عامة عندما تتخذ قرارات معينة بناء على تقديرات القدرة فإن التحيز سيسبب تأثيرا سلبيا على صدق الاختبار. (Wang & Hanson ,1999:263)

ومن خلال العرض السابق للأساليب التي تستخدم في الكشف عن دقة تقديرات القدرة الناتجة عن التحليلات، فسوف يستخدم الباحث الحالي إثنان من هذه الأساليب وهما : الخطأ المعياري لتقدير القدرة (SEE)، ، وكذلك تقدير ثبات دوال معلومات الاختبار (Reliainf) وذلك للكشف عن دقة تقديرات القدرة التي يمكن الحصول عليها من تحليلات النماذج الثلاثة المستخدمة بالدراسة الحالية ، نظرا لأن هذين المؤشرين متوافرين ببرنامج التحليل المستخدم بالدراسة الحالية وهو برنامج BILOG-MG3 (Zimowski, et.al.,2003)

مشكلة الدراسة:

تتلخص مشكلة الدراسة في الإجابة على السؤال العام التالي:

هل تختلف تقديرات قيم دوال معلومات وكذلك تقديرات القدرة لعينة متلقي الاختبار باختلاف نموذج التحليل المستخدم (النموذج أحادي والثنائي والثلاثي البارمتر) ويتفرع من هذ السؤال الأسئلة الفرعية التالية:

١. هل تختلف قيم كل من دوال معلومات مفردات الاختبار (أقصى قيمة للمعلومات – مستويات القدرة المناظرة لأقصى قيمة – متوسط قيم دوال معلومات المفردة عند مستويات القدرة المختلفة) باختلاف نموذج التحليل ؟
٢. هل تختلف دقة تقديرات القدرة – متمثلة في قيمة الخطأ المعياري للتقدير – للعينة المستخدمة باختلاف نموذج التحليل المستخدم ؟
٣. هل هناك علاقة بين دقة تقديرات القدرة للأفراد وقيمة دالة معلومات الاختبار؟

أهمية الدراسة

تبدو أهمية الدراسة من خلال النقاط التالية:

أولاً : الأهمية النظرية والتطبيقية:

١. الدراسة تهتم بتحليل اختبار المصفوفات المتتابعة (العادي) 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.
٢. تحاول الدراسة التعرف على أي من نماذج التحليل الثلاثة المستهدفة من الدراسة التي يمكن أن يعطي أكبر قيمة لدوال المعلومات والذي يمكن بدورها أن تعطي أفضل وأدق تقدير للقدرة المستهدف قياسها من 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 دوال معلومات بنود الاختبار باستخدام نماذج مختلفة من التحليل.
٢. الكشف عن مدى تأثر تقديرات القدرة بنموذج التحليل المستخدم.

٣. استيضاح العلاقة بين دقة تقديرات القدرة التي تم استخراجها من تحليلات النماذج الثلاثة ومتغير قيم دوال معلومات بنود الاختبار.

محددات الدراسة:

- اقتصرت الدراسة على بيانات اختبار المصفوفات المتتابعة (العادي) لجون رافن.
- اقتصرت الدراسة على تطبيق الأداة الخاصة بها على الطلاب الذكور دون الإناث.
- اقتصرت الدراسة في تطبيق الأداة الخاصة بها على طلاب المرحلة الثانوية من التعليم العام دون الخاص.
- اقتصرت الدراسة على طلاب المستويات من المستوى الثاني حتى المستوى السادس.
- اقتصرت الدراسة على التحليلات بثلاث نماذج مرتبطة بنظرية الاستجابة للمفردة (النموذج الأحادي اللبارمتر – النموذج ثلاثي 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 التاسع عشر.
- اقتصرت الدراسة على تطبيق الأداة الخاصة بها في العام الدراسي ١٤٣٠ – ١٤٣١هـ

فروض الدراسة

١. لا تختلف قيم دوال معلومات مفردات الاختبار (أقصى قيمة– مستويات القدرة المناظرة لأقصى قيمة – متوسط قيم دوال معلومات المفردة عند مستويات القدرة المختلفة) باختلاف نموذج التحليل المستخدم.

٢. لا تختلف دقة تقديرات القدرة – ممثلة بالخطا المعياري للتقدير – وثبات قيم دوال المعلومات باختلاف نموذج التحليل المستخدم .

٣. لاتوجد علاقة بين دقة تقديرات القدرة للأفراد ودوال معلومات مفردات الاختبار.

إجراءات الدراسة

١- استخدمت الدراسة اختبار المصفوفات المتتابعة (العادي) "لجون رافن"، وهو أحد الاختبارات العالمية المشهورة في قياس 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 ، حيث 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 (1) والخاطئة بالدرجة (0)، لكل مفردة من مفردات الاختبار (فؤاد أبو حطب،١٩٩٦)(Raven, J.,1989) (صفوت فرج ،١٩٨٠)

وقد وقع اختيار الباحث 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 ، كما تحققت الدراسة أيضا من ثبات الاختبار باستخدام معادلة كيودار- ريتشاردسون (صيغة 21) وبلغت (0.98) وهي الصيغة التي تناسب المفردات ثنائية الإجابة ، كما استخدم الباحث المجموع الكلي للطلاب في السنة الدراسية السابقة كمحك خارجي لصدق الاختبار على عينة استطلاعية قوامها (157) من طلاب الصفوف الثلاثة من المرحلة الثانوية بواقع عدد [(42) من الصف الثالث الثانوي +(61) من الصف الثاني ، (54) طالب من الصف الأول]، وقد بلغ قيمة معامل ارتباط بيرسون بين الاختبار المحك (0.88) وبالتالي كانت هناك ثقة في دقة البيانات التي تم الحصول عليها باستخدام هذا الاختبار.

٢- أجري تطبيق الاختبار على عينة عشوائية طبقية من المرحلة الثانوية بلغت تعدادها (858) طالبا من جميع الصفوف الدراسية بالمرحلة الثانوية (الأول – حتى الثالث) أجري سحبها من خمسة من

المدارس الثانوية بالمنطقة التعليمية بمحافظة جنوب القاهرة .

٣- تم تحليل استجابات عينة الطلاب لجميع بنود الاختبار باستخدام نماذج التحليل الثلاث (أحادي – ثنائي – ثلاثي البارمتر) وذلك باستخدام برنامج BILOG-MG3.

٤- إيجاد تقديرات القدرة للأفراد لكل نموذج على حدة.

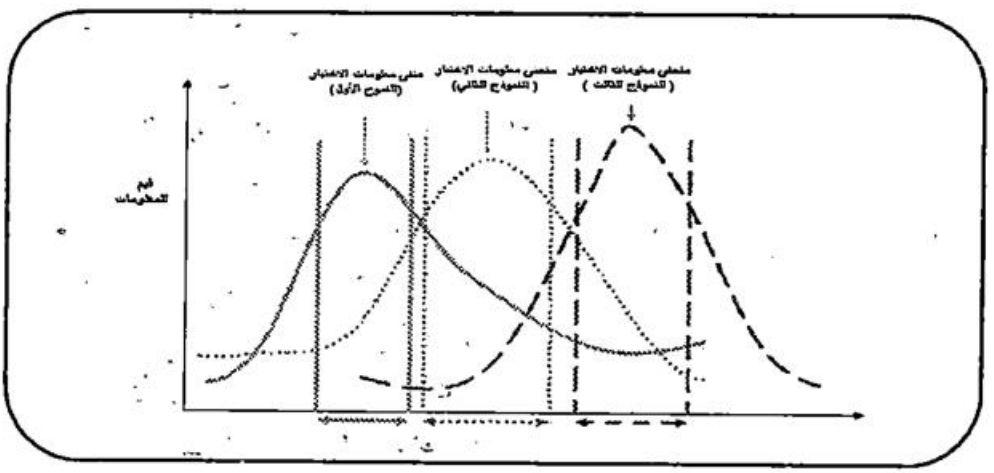
٥- إيجاد مؤشرات دوال المعلومات لكل بند على حدة باستخدام نماذج التحليل الثلاثة وهي:

(أ) أقصى معلومات للمفردة الاختبارية Maximum Information

ويشير هذا المتغير إلى أقصى قيمة لدالة المعلومات لكل بند من بنود الاختبار نتيجة التحليلات الخاصة بكل نموذج من النماذج الثلاثة المستهدفة على حدة.

(ب) مستوى القدرة المناظر لأقصى معلومات للمفردة. Point of Maximum Information

ويتحدد من خلال متوسط نقاط القدرة التي تقابل أعلى قيم لدوال مفردات الاختبار (قمة منحنى دوال معلومات الاختبار) ، ويرى الباحث أن نقاط القدرة التي تناظر أعلى قيم لدوال معلومات الاختبار قد تنتمي إلى مستوى محدد من القدرة وقد تختلف باختلاف نموذج التحليل، والشكل التالي يوضح الأساس المنطقي لهذا الإجراء:



شكل (1)

الاختلافات المتوقعة لفترات القدرة التي تناظر أقصى قيمة لدوال معلومات بنود الاختبار (التي تقابل قمة منحنى دوال معلومات الاختبار الناتجة عن التحليل بالنماذج الثلاثة)

وسوف يعتمد الباحث على تحليل قياس التباين المتكرر في عقد مقارنة بين مستويات القدرات الثلاثة والتي تقابل أقصى قيم لتقديرات المعلومات.

(ج) متوسط معلومات المفردة الاختبارية عند المستويات المختلفة من القدرة. Information Average

ويشير هذا المتغير إلى متوسط قيم دوال المعلومات لكل مفردة اختبارية على متصل القدرة المقاسة.

٦- استخدمت الدراسة أسلوب تحليل القياس المتعدد للقياسات الثلاثة للقدرة باستخدام نماذج التحليل الثلاثة.

ومن المعروف أن المقارنات الثنائية التي يعتمد عليها هذا النوع من التحليلات بين النماذج الثلاثة ، تعتمد على المتوسطات (متوسطات تقديرات قيم لمعلومات المفردات الاختبارية) وحيث ان هذه المتوسطات تتأثر بالقيم المتطرفة في حالة وجود توزيع ملتوي لقيم أعلى معلومات فسوف يقوم الباحث بالتحقق من اعتدالية توزيع قيم معلومات للمفردات قبل استخدام أسلوب تحليل القياس المتكرر، وإن لم يتحقق توافر شرط الاعتدالية فسوف تستخدم أساليب أخرى بديلة.

منهج الدراسة و الأساليب 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.
- استخدام مؤشر الملاءمة الإحصائي مربع كاي المتوافر بتحليلات برنامج BILOG-MG3 للتحقق من توافر افتراض استقلالية المحل لمفردات اختبار المصفوفات المتتابعة.
- أسلوب تحليل تباين القياس المتكرر Repeated Measures المتوافر بتحليلات برنامج SPSS للتحقق من وجود فروق في تباين قياسات دوال المعلومات للفروض الثلاثة الأولى (أسلوب فريدمان كبديل في حالة عدم توافر شروط استخدام تباين القياس المتكرر)
- استخدام اختبار (LSD) لعقد المقارنات الثنائية Pairwise Comparisons (وليست

المتعددة نظرا لأن التحليلات تجرى على مجموعة واحدة فقط من الأفراد ، ولا تجرى على مجموعات متعددة.

- اختبار (**كلمجروف – سميرنوف**) المتوافر بتحليلات برنامج SPSS للتحقق من مدى اعتدالية توزيع قيم القياسات الخاصة بجميع المتغيرات موضع المقارنة.

عرض نتائج الدراسة ومناقشتها

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

(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:

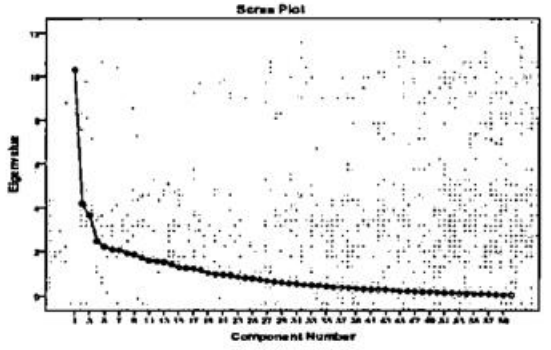
جدول (1)

الجذور الكامنة للعوامل الثلاثة الأولى الناتجة عن التحليل العاملي لمفردات 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 للتتباين 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 ، وهذا مايبدو واضحاً من مقياس Scree plot وهو مايعرف باختبار الهضبة أو الفرز ،الذي يظهر انحدار شديد بين قيم الجذر الكامن بين العامل الأول والثاني بينما يقل هذا الانحدار بشكل واضح بين الجذور الكامنة للعوامل الأخرى مما يشير إلى مؤشر آخر لتمتع الاختبار بأحادية البعد. (Gessaroli & De Champlain , 2005)



شكل (1)

تمثيل بياني لقيم الجذور الكامنة الناتجة عن تحليل مفردات اختبار الدراسة

باستخدام أسلوب التحليل العاملي

(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).

- 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 راعى الباحث في تطبيق اختبار الدراسة إعطاء وقتا مفتوحا للإجابة عن جميع فقرات الاختبار لحين انتهاء كل المفحوصين من محاولة الإجابة عن جميع فقرات الاختبار ، ولم يتبين للباحث وجود أي ردود أفعال من الطلاب عن ضيق الوقت المخصص للإجابة عن لفقرات الاختبار ،وبالتالي فلم يكن لعامل السرعة أي تأثير في تحديد نوع الاستجابة على فقرات الاختبار مما يعني تحقق افتراض السرعة.

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

لقد أسفرت نتائج المطابقة للأفراد والمفردات عن النتائج التالية:

- مطابقة المفردات Item Fit لتوقعات النماذج اللوغاريتمية (الأحادية والثنائية 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 ، وبالتالي كانت نتائج التحليلات التي تم التوصل إليها من هذه الخطوة هي النتائج النهائية التي تم الاعتماد عليها في المعالجات الإحصائية لمتغيرات الدراسة.

- مطابقة الأفراد Person Fit لتوقعات النماذج اللوغاريتمية (الأحادية والثنائية والثلاثية البارمتر)، وقد تبين أن إجمالي عدد الطلاب وعددهم (24) طالباً جاء مؤشر المطابقة لقدراتهم خارج حدود المطابقة المسموح بها (في احد التحليلات الخاصة بأحد النماذج الثلاثة على الأقل) وتم إستبعادهم من التحليلات النهائية.

ثانيا : نتائج اختبار اعتدالية توزيعات تقديرات جميع متغيرات الدراسة:

لقد أسفرت التحليلات الخاصة باختبار **كلموجروف – سميرنوف** للتحقق من مدى اعتدالية توزيع التقديرات الخاصة بكل من : أقصى قيمة للمعلومات وكذلك نقاط القدرة التي تحقق أقصى معلومات للمفردات الاختبارية ومتوسط تقديرات المعلومات لمفردات الاختبار وكذلك أخطاء قياس القدرات ،وكذلك ثبات تقدير دوال معلومات بنود الاختبار ، والتي أجري التوصل إليها من التحليلات الخاصة بنماذج التحليل الثلاث (الأحادي والثنائي والثلاثي) البارمتر – عن النتائج التالية:

١- نتائج التحقق من مدى اعتدالية توزيع تقديرات أقصى قيم لدوال معلومات مفردات :

جدول (2)

قيمة ودلالة اختبار كلمجروف- سميرنوف لتوزيع تقديرات أقصى قيم لدوال معلومات مفردات اختبار المصفوفات المتتابعة المقدرة من نماذج التحليل

	أقصى	أقصى معلومات	أقصى معلومات
قيمة (z) لاختبار كلمجروف-	1.322	1.202	1.029
مستوى الدلالة	0.242	0.111	0.23

يتضح من الجدول السابق أن قيم اختبار (كلموجروف – سميرنوف) لأقصى قيم المعلومات لمفردات الاختبار عبر الفترات (المستويات المختلفة من القدرة) والتي تم الحصول عليها من خلال التحليلات الخاصة بالنماذج الثلاثة : الأحادي والثنائي والثلاثي البارمتر على الترتيب، قد بلغت (1.322 ، 1.202 ، 1.029) وهي قيم غير دالة عند مستوى (0.05) مما يعني أن جميع توزيعات تقديرات قيم قياسات متغير أقصى معلومات لدوال معلومات الاختبار جاءت جميعها بشكل اعتدالي.

٢- نتائج التحقق من مدى اعتدالية توزيع تقديرات نقاط القدرة التي تقابل أقصى قيم لدوال معلومات مفردات :

جدول (3)

قيمة ودلالة اختبار كلمجروف- سميرنوف لتوزيع تقديرات نقاط القدرة التي تقابل أقصى قيم لدوال معلومات مفردات اختبار المصفوفات المتتابعة المقدرة من نماذج التحليل

	النقطة المقابلة لأقصى معلومات (النموذج الأحادي)	النقطة المقابلة لأقصى معلومات (النموذج الثنائي)	النقطة المقابلة لأقصى معلومات (النموذج الثلاثي)
قيمة (z) لاختبار	0.531	0.837	0.375
مستوى الدلالة	0.941	0.485	0.714

يتضح من الجدول السابق أن قيم اختبار (كلموجروف – سميرنوف) لتقديرات نقاط القدرة التي تقابل أقصى قيم لدوال معلومات مفردات والتي تم الحصول عليها من خلال التحليلات الخاصة بالنماذج الثلاثة : الأحادي والثنائي والثلاثي البارمتر، قد جاءت على النحو التالي (531،.837، 375.) على الترتيب وهي جميعها قيم غير دالة عند مستوى (0.05) مما يعني أن جميع توزيعات قيم قياسات نقاط القدرة التي تقابل أقصى قيم لمعلومات لدوال معلومات الاختبار جاءت جميعها بشكل اعتدالي.

٣- نتائج التحقق من مدى اعتدالية توزيع تقديرات متوسط تقديرات المعلومات لمفردات الاختبار:

جدول (4)

قيمة ودلالة اختبار كلمجروف- سميرنوف لتوزيع تقديرات متوسط تقديرات المعلومات لمفردات اختبار المصفوفات المتتابعة

	متوسط تقديرات المعلومات (النموذج الأحادي)	متوسط تقديرات المعلومات (النموذج الثنائي)	متوسط تقديرات المعلومات (النموذج الثلاثي)
قيمة (z) لاختبار كلمجروف-سميرنوف	1.113	1.051	1.118
مستوى الدلالة	0.094	0.219	0.164

يتضح من الجدول السابق أن قيم اختبار (كلموجروف – سميرنوف) متوسط تقديرات المعلومات لمفردات الاختبار والتي تم الحصول عليها من خلال التحليلات الخاصة بالنماذج الثلاثة : الأحادي والثنائي والثلاثي البارمتر. بلغت (1.113 ،1.051، 1.118) على الترتيب، وجميعها قيم غير دالة عند مستوى (0.05) مما يعني أن جميع توزيعات هذه التقديرات جاءت جميعها بشكل اعتدالي.

٤- نتائج التحقق من مدى اعتدالية توزيع تقديرات قيم ثبات دوال معلومات بنود الاختبار:

جدول (5)

قيمة ودلالة اختبار كلمجروف- سميرنوف لتوزيع قيم ثبات تقديرات دوال المعلومات لمفردات اختبار المصفوفات المتتابعة

	قيم ثبات تقديرات المعلومات (النموذج الأحادي)	قيم ثبات تقديرات المعلومات (النموذج الثنائي)	قيم ثبات تقديرات المعلومات (النموذج الثلاثي)
قيمة (z) لاختبار كلمجروف-سميرنوف	1.365	1.008	0.854
مستوى الدلالة	0.088	0.262	0.459

يتضح من الجدول السابق أن قيم اختبار (كلموجروف – سميرنوف) لقيم ثبات تقديرات دوال المعلومات بنود الاختبار والتي تم الحصول عليها من خلال التحليلات الخاصة بالنماذج الثلاثة : الأحادي والثنائي والثلاثي البارمتر. بلغت (1.365، 1.008 ، 854.) على الترتيب وجميعها قيم غير دالة عند مستوى (0.05) مما يعني أن جميع توزيعات هذه الأخطاء جاءت جميعها بشكل اعتدالي.

٥- نتائج التحقق من مدى اعتدالية توزيع قيم أخطاء تقديرات القدرة:

جدول (6)

قيمة ودلالة اختبار كلمجروف- سميرنوف لتوزيع قيم أخطاء تقديرات القدرة

	قيم أخطاء تقديرات القدرة (النموذج الأحادي)	قيم أخطاء تقديرات القدرة (النموذج الثنائي)	قيم أخطاء تقديرات القدرة (النموذج الثلاثي)
قيمة (z) لاختبار كلمجروف- سميرنوف	1.128	1.049	1.086
مستوى الدلالة	0.08	0.12	0.09

كما يتضح من الجدول السابق أن قيم اختبار (كلموجروف – سميرنوف) لقيم تقديرات أخطاء تقديرات القدرة والتي تم الحصول عليها من خلال التحليلات الخاصة بالنماذج الثلاثة : الأحادي والثنائي والثلاثي البارمتر (1.128، 1.049، 1.086) على الترتيب، بلغت وهي جميعها قيم غير دالة عند مستوى (0.05) مما يعني أن جميع توزيعات هذه التقديرات جاءت جميعها بشكل اعتدالي.

ويتضح للباحث من النتائج السابقة أن جميع التوزيعات الخاصة بتقديرات دوال المعلومات الناتجة من تحليلات النماذج الثلاثة جاءت على نحو اعتدالي ، كما اتضح للباحث أيضا من النتائج السابقة أن جميع التوزيعات الخاصة بتقديرات قيم ثبات دوال المعلومات وكذلك أخطاء تقديرات القدرة الناتجة من تحليلات النماذج الثلاثة جاءت على نحو اعتدالي مما يتيح عقد المقارنات في ضوء جميع المتغيرات موضع الدراسة باستخدام أسلوب القياس المتكرر.

ثالثاً :عرض ومناقشة نتائج التحليلات الخاصة باختبار فروض الدراسة:

١- نتائج التحليلات الخاصة بدوال المعلومات.

(أ) نتائج التحليلات الخاصة بأقصى قيمة لدوال المعلومات

جدول (7)

المتوسط والإنحراف للقيم العظمى لدوال المعلومات لمفردات اختبار المصفوفات المتتابعة

نموذج التحليل	متوسط القيم العظمى لدوال المعلومات	الإنحراف المعياري
أحادي البارمتر	0.242	0.000000
ثنائي البارمتر	0.355315	0.2888369
ثلاثي البارمتر	0.37039	0.4072798

يتضح من الجدول السابق ان هناك فروق ظاهرية بين متوسطات القيم العظمى لدوال معلومات مفردات الاختبار حيث بلغت أكبر قيمة لمتوسط القيم العظمى لدوال معلومات الاختبار القيمة (0.37) تقريبا، وكانت ناتجة من تحليلات النموذج ثلاثي البارمتر ، بينما بلغت أقل قيمة لمتوسط القيم العظمى لدوال معلومات بنود الاختبار القيمة (0.24) وكانت للنموذج أحادي البارمتر، بينما بلغ متوسط القيم العظمى لدوال معلومات المفردات (.355315) وكانت للنموذج الثنائي البارمتر.

جدول (8)

اختبارات دلالة تباين القيم العظمى لدوال معلومات مفردات اختبار المصفوفات المتتابعة الناتجة عن تحليلات النماذج موضع المقارنة

الاختبار	قيمة الاختبار	قيمة التباين	درجات الحرية	مستوى الدلالة
Pillai's Trace	0.136	4.582	2	0.014
Wilks' Lambda	0.864	4.582	2	0.014
Hotelling's Trace	0.158	4.582	2	0.014
Roy's Largest Root	0.158	4.582	2	0.014

يتضح من الجدول السابق الذي يتضمن نتائج الاختبارات المختلفة التي تبرز دلالة التباين بين القيم العظمى لدوال المعلومات الناتجة عن التحليل بالنماذج الثلاثة (الأحادي – الثنائي – الثلاثي) البارمتر ، ان جميع الاختبارات وأشهرها اختبار ولكس لامدا Wilks' Lambda قد بلغت قيمته 864. بتباين قدره (4.582) وهي قيمة دالة عند مستوى دلالة قدره (0.14) . مما يشير إلى وجود دلالة للتباين بين القيم العظمى لدوال معلومات الاختبار الناتجة عن التحليل باستخدام نماذج التحليل الثلاث ترجع إلى نموذج التحليل المستخدم.

جدول (9)

اختبار "ماكولي " للدورية بين قياسات القيم العظمى لمعلومات مفردات اختبار المصفوفات المتتابعة الناتجة عن نموذج التحليل

قيمة اختبار ماكولي	مربع كاي التقريبية	درجات الحرية	مستوى الدلالة
0.716	19.368	2	0.00

يتضح من الجدول السابق ان قيمة اختبار ماكولي للدورية بلغت (716.) وهي قيمة دالة حيث بلغت قيمة قيمة الدلالة المقترنة بها القيمة (0.00) مما يشير إلى وجود اختلافا جوهريا بين تجانس

القياسات الثلاثة الناتجة عن التحليل بالنماذج الثلاثة المستخدمة مما يشير إلى عدم تحقق شرط الدورية ومما يستوجب في مثل هذه الحالة استخدام اختبار " جرين هاوس – جيسر" Greenhouse-Geisser عند اختبار دلالة التباين بين القياسات الثلاثة الناتجة عن التحليل باستخدام النموذج المستخدم ، وهو مايتضح من خلال الجدول التالي:

جدول (10)

تحليل تباين القياسات الثلاثة الناتجة عن تحليل مفردات اختبار المصفوفات المتتابعة بالنموذج المستخدم

مصدر التباين	الاختبار	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة (ف)	مستوى الدلالة
بين القياسات	Sphericity Assumed	0.591	2	0.296	5.456	0.005
	Greenhouse-Geisser	0.591	1.558	0.379	5.456	0.01
	Huynh-Feldt	0.591	1.592	0.371	5.456	0.01
	Lower-bound	0.591	1	0.591	5.456	0.023
داخل القياسات (الخطأ)	Sphericity Assumed	6.391	118	0.054		
	Greenhouse-Geisser	6.391	91.908	0.07		
	Huynh-Feldt	6.391	93.946	0.068		
	Lower-bound	6.391	59	0.108		

يتضح من الجدول السابق أن قيمة اختبار التباين باستخدام طريقة "جرين هاوس – جيسر " بلغت (5.456) وهي قيمة دالة حيث بلغت قيمة الدلالة الخاصة بها (0.01) مما يشير إلى وجود فروق جوهرية بين متوسطات القياسات الخاصة بالقيم العظمى لدوال معلومات بنود الاختبار وللتعرف على اتجاه هذه الفروق يجب استخدام جدول المقارنات الثنائية Pairwise Comparisons بين متوسطات القياسات موضع المقارنة والذي يعتمد على طريقة أقل مربعات صغري بين الفروق LSD وهي موضحة بالجدول التالي:

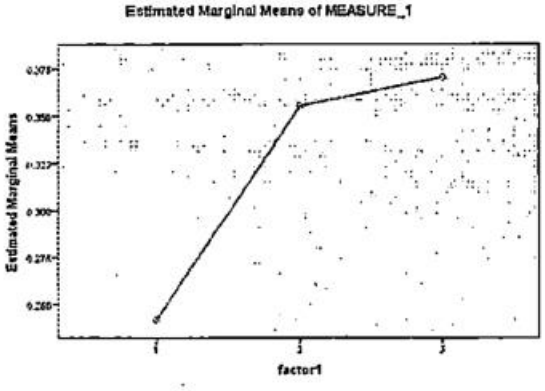
جدول (11)

المقارنات الثنائية بطريقة (LSD) بين متوسطات القيم العظمى لدوال معلومات مفردات اختبار المصفوفات المتتابعة باستخدام نماذج التحليل

طرف المقارنة الأول	طرف المقارنة الثاني	الفرق بين متوسطي طرفي المقارنة (الأول – الثاني)
النموذج الأحادي	النموذج الثنائي	-.113*
	النموذج الثلاثي	-.128*
النموذج الثنائي	النموذج الأحادي	.113*
	النموذج الثلاثي	-.015

* دال عند مستوى (0.05)

يتضح من الجدول السابق أن هناك فروقاً جوهرية بين متوسطي قيم دوال معلومات بنود الاختبار باستخدام نموذجي التحليل الأحادي والثنائي البارمتر في صالح النموذج ثنائي البارمتر. كما يتضح أيضا أن هناك فروقاً جوهرية بين متوسطي قيم دوال معلومات بنود الاختبار باستخدام نموذجي التحليل الأحادي والثلاثي البارمتر في صالح نموذج التحليل ثلاثي البارمتر ،كما يتضح أيضا عدم وجود فروق جوهرية بين نموذجي التحليل الثنائي والثلاثي البارمتر.



شكل (2)

تمثيل بياني لمتوسط القيم العظمى لدوال مفردات اختبار المصفوفات المتتابعة للنماذج الثلاث موضع المقارنة [الأحادي (1) و الثنائي (2) والثلاثي (3)]

يتضح من التمثيل البياني السابق أن أعلى متوسط لقيم العظمى لدوال معلومات بنود الاختبار كانت ناتجة عن التحليل باستخدام النموذج الثلاثي البارمتر ، يلي ذلك متوسط القيم العظمى لدوال معلومات بنود الاختبار باستخدام نموذج الثنائي البارمتر ثم جاء متوسط القيم العظمى دوال معلومات النموذج أحادي البارمتر في كأقل قيمة لمتوسطات القيم العظمى دوال معلومات بنود الاختبار. مما يشير إلى ان نتائج التحليل باستخدام نموذج التحليل ثلاثي البارمتر تحقق أفضل نتائج إذا ماقورنت بنتائج التحليلات الخاصة بنموذجي التحليل الآخرين، ويرى الباحث الحالي ان الفروق التي جاءت في صالح النموذجين الثنائي والثلاثي مقابل النموذج الأحادي البارمتر قد تكون نتيجة منطقية نظرا لاعتماد كل من هذين النموذجين على بارمترات إضافية في تقدير أقصى قيمة لدوال معلومات الاختبار حيث يعتمد النموذج الثنائي على بارمتر التمييز كما يعتمد النموذج الثلاثي على بارمتري التمييز والتخمين مما قد يؤدي إلى إعطاء تقديرات أعلى لأقصي قيمة لمعلومات المفردة الاختبارية وهذه النتيجة تؤكد النتائج التي توصلت إليها بعض الدراسات التي اجريت بهدف التعرف على دقة تقديرات النموذج الثنائي والثلاثي البارمتر حيث توصلت العديد من الدراسات دقة تقدير قدرات الأفراد باستخدام النموذج الثنائي البارمتر وكذلك الثلاثي البارمتر في مقابل النموذج الأحادي البارمتر . ومن بين هذه الدراسات دراسة Van Der Maas, Han L. J.(2011,) التي أثبتت تفوق النموذج الثنائي البارمتر على النماذج الأخرى في اتخاذ القرارات فيما يتعلق بالفروق الفردية للأفراد وكذلك دراستي)Daniel,(Courville,2004) (2010, التي توصلتا إلى تفوق النموذج الثنائي البارمتر على النموذج الأحادي .وهذه النتيجة تتسق كذلك مع ما توصلت إليه دراسة (Jaio,2003) حيث 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 والثلاثي البارمتر في انتقاء مفردات اختبار القبول لكليات الطب بالجامعات الأمريكية ،كما اتسقت كذلك نتائج دراسة العكايلة (٢٠٠٧) إلى الدقة العالية لتقدير النموذج الثلاثي البارمتر بالمقارنة بدقة تقديرات الطرق الكلاسيكية، كما تتسق هذه النتائج مع النتائج التي انتهت إليها 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 اتسقت كذلك مع نتائج دراسة الوليلي (٢٠٠١) التي انتهت إلى ان النموذج الثلاثي البارمتر هو الأكثر دقة في انتقاء البيانات مقارنة بالنموذجين الآخرين، وقد يعزو الباحث دقة هذا النموذج بالمقارنة بالنموذجين الآخرين نظرا لاعتماده في تقدير قيم دوال المعلومات على متغير التخمين وهو المتغير الذي لايعتمد عليه النموذجين الآخرين في تقديرهما لنفس الدوال.

(ب) نتائج التحليلات الخاصة بنقاط القيم العظمى لدوال المعلومات الناتجة عن استخدام النماذج الثلاث:

جدول (12)

المتوسط والإنحراف المعياري لنقاط القدرة التي تقابل أقصى قيم لدوال معلومات مفردات اختبار المصفوفات المتتابعة

نموذج التحليل	متوسط نقاط القدرة المقابلة للقيم العظمى لدوال المعلومات	الإنحراف
أحادي البارمتر	-1.416240	1.5878068
ثنائي البارمتر	2.000372	2.2794945
ثلاثي البارمتر	1.804467	3.7153143

يتضح من الجدول السابق ان هناك فروقاً ظاهرية بين متوسطات قيم نقاط القدرة التي تحقق عندها أقصى قيمة لدوال معلومات بنود الاختبار حيث بلغت أكبر قيمة لمتوسط نقاط القدرة القيمة (2) تقريبا وكانت ناتجة من تحليلات النموذج ثنائي البارمتر ، بينما بلغت أقل قيمة لمتوسط نقاط القدرة التي تحقق عندها أقصى قيمة لدوال بنود الاختبار القيمة (-1.42) تقريبا وكانت للنموذج أحادي البارمتر . مما يعنى أن النموذج الثنائي البارمتر يعطي أقصى معلومات للأفراد ذوي القدرة المرتفعة يلية النموذج ثلاثي البارمتر ، أما النموذج الأحادي البارمتر يعطي أقصى قيمة لدوال المعلومات لدى الفراد منخفضي القدرة.

جدول (13)

اختبارات دلالة تباين نقاط القدرة التي تناظر أقصى معلومات ناتجة عن تحليل مفردات اختبار المصفوفات المتتابعة بالنماذج الثلاثة

الاختبار	قيمة الاختبار	قيمة التباين	درجات الحرية	مستوى الدلالة
Pillai's Trace	.513	30.558	2	.000
Wilks' Lambda	.487	30.558	2	.000
Hotelling's	1.054	30.558	2	.000
Roy's Largest	1.054	30.558	2	.000

يتضح من الجدول السابق الذي يتضمن نتائج الاختبارات المختلفة التي تبرز دلالة التباين بين نقاط القدرة التي تحقق عندها أقصى قيمة لدوال لمعلومات بنود الاختبار الناتجة عن التحليل بالنماذج الثلاثة (الأحادي – الثنائي – الثلاثي) البارمتر ، ان جميع الاختبارات وأشهرها اختبار ولكس لامدا Wilks' Lambda قد بلغت قيمته (.487) بتباين قدره (30.558) وهي قيمة دالة عند مستوى دلالة قدره (0.00) . مما يشير إلى وجود دلالة للتباين بين نقاط القدرة التي يتحقق عندها أقصي قيم لدوال معلومات بنود الاختبار الناتجة عن التحليل باستخدام نماذج التحليل الثلاث تعود إلى نموذج التحليل المستخدم.

جدول (14)

اختبار "ماكولي " للدورية بين تقديرات متوسطات نقاط القدرة التي تناظر أقصى معلومات لمفردات اختبار المصفوفات المتتابعة الناتجة عن التحليل بالنموذج المستخدم

قيمة اختبار ماكولي	مربع كاي التقريبية	درجات الحرية	مستوى الدلالة
.548	34.883	2	0.00

يتضح من الجدول السابق ان قيمة اختبار ماكولي للدورية بلغت (548.) وهي قيمة دالة حيث بلغت قيمة قيمة الدلالة المقترنة بها القيمة (0.00) مما يشير إلى وجود اختلافا جوهريا بين تجانس القياسات الثلاثة الناتجة عن التحليل بالنماذج الثلاثة المستخدمة مما يشير إلى عدم تحقق شرط الدورية ومما يستوجب في مثل هذه الحالة استخدام اختبار " جرين هاوس - جيسر" Greenhouse-Geisser عند اختبار دلالة التباين بين القياسات الثلاثة الناتجة عن التحليل باستخدام النموذج المستخدم ، وهو مايتضح من خلال الجدول التالي:

جدول (15)

تحليل تباين القياسات الثلاثة لنقاط القدرة التي تناظر أقصى معلومات و الناتجة عن التحليل باستخدم نموذج التحليل المستخدم

مصدر التباين	الاختبار	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة (ف)	مستوى الدلالة
بين القياسات	Sphericity Assumed	441.691	2 .	220.846	35.576	.000
	Greenhouse-Geisser	441.691	1.377	320.662	35.576	.000
	Huynh-Feldt	441.691	1.400	315.594	35.576	.000
	Lower-bound	441.691	1.000	441.691	35.576	.000
داخل القياسات (الخطأ)	Sphericity Assumed	732.520	118	6.208		
	Greenhouse-Geisser	732.520	81.269	9.014		
	Huynh-Feldt	732.520	82.574	8.871		
	Lower-bound	732.520	59.000	12.416		

يتضح من الجدول السابق أن قيمة اختبار التباين (f) باستخدام طريقة " جرين هاوس - جيسر " بلغت (35. 576) وهي قيمة دالة حيث بلغت قيمة الدلالة الخاصة بها (0.00) مما يشير إلى وجود فروق جوهرية بين متوسطات القياسات الخاصة بنقاط القدرة التي تحققت عندها أقصى قيم لدوال معلومات بنود الاختبار وللتعرف على اتجاه هذه الفروق يجب استخدام جدول المقارنات الثنائية Pairwise Comparisons بين متوسطات القياسات موضع المقارنة والذي يعتمد على طريقة أقل مربعات صغري بين الفروق LSD وهي موضحة بالجدول التالي:

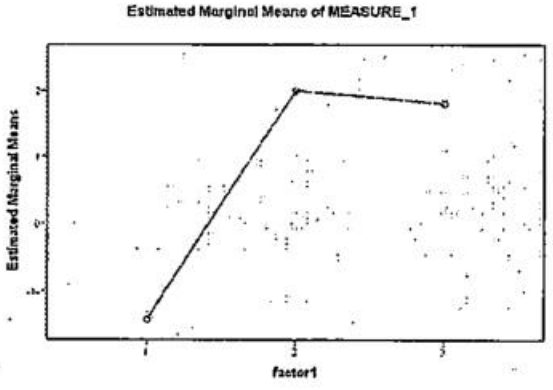
جدول (16)

المقارنات الثنائية بطريقة (LSD) بين متوسطات تقديرات نقاط القدرة التي تقابل أقصى معلومات لمفردات اختبار المصفوفات المتتابعة باستخدام نماذج التحليل

طرف المقارنة الأول	طرف المقارنة الثاني	الفرق بين متوسطي طرفي المقارنة (الأول – الثاني)
النموذج الأحادي	النموذج الثنائي	-3.417*
	النموذج الثلاثي	-3.221*
النموذج الثنائي	النموذج الأحادي	3.417*
	النموذج الثلاثي	.196
النموذج الثلاثي	النموذج الأحادي	3.221*
	النموذج الثنائي	-.196

* دال عند مستوى (0.05)

يتضح من الجدول السابق أن هناك فروقاً جوهرية بين متوسطي قيم نقاط القدرة التي تحقق عندها أقى قيم لدوال معلومات بنود الاختبار باستخدام نموذجي التحليل الأحادي والثنائي البارمتر في صالح نموذج التحليل ثنائي البارمتر. كما يتضح أيضا أن هناك فروق جوهرية بين متوسطي قيم دوال معلومات بنود الاختبار باستخدام نموذجي التحليل الأحادي والثلاثي البارمتر في صالح نموذج التحليل الثلاثي البارمتر ، كما يتضح أيضا عدم وجود فروق جوهرية بين نموذجي التحليل الثنائي والثلاثي البارمتر



شكل (3)

تمثيل بياني لمتوسط نقاط القدرة التي تناظر أقصى معلومات لدوال مفردات اختبار المصفوفات المتتابعة للنماذج الثلاث موضع المقارنة [الأحادي (1) و الثنائي (2) والثلاثي (3)]

يتضح من التمثيل البياني السابق أن متوسط نقاط القدرة التي تحقق عندها أقصى قيم لدوال معلومات بنود الاختبار كانت ناتجة عن التحليل باستخدام النموذج الثنائي البارمتر ، يلي ذلك متوسط النقاط التي تناظر القيم العظمى لدوال معلومات بنود الاختبار باستخدام نموذج الثلاثي البارمتر، ثم جاء متوسط نقاط القدرة التي تحقق عندها أقصى قيم لدوال معلومات بنود الاختبار النموذج أحادي البارمتر كأقل قيمة لمتوسطات تلك النقاط من القدرة. مما يشير إلى ان نتائج التحليل باستخدام نموذج التحليل ثنائي البارمتر تحقق أفضل نتائج إذا ماقورنت بنتائج التحليلات الخاصة بنموذجي التحليل الآخرين. وهذه النتيجة تتسق مع النتائج التي توصلت إليها بعض الدراسات التي اجريت بهدف التعرف على دقة تقديرات التنموذج الثنائي ، مثل دراسة .Van Der& Han(2011,) التي أثبتت تفوق النموذج الثنائي البارمتر على النماذج الأخرى في اتخاذ القرارات فيما يتعلق بالفروق الفردية للأفراد ومن بين هذه الدراسات أيضا دراسة)Daniel,2010) وكذلك (Courville,2004) التي توصلت الدراسة إلى تفوق النموذج الثنائي البارمتر وكذلك دراسة (Jaio,2003) التي أشارت إلى أن النموذج الثنائي كان افضل النماذج الثلاث، كما اتفقت مع نتائج دراسة (Aimee,et,al,2008) التي أفادت بدقة هذا النموذج الثنائي البارمتر إذا ماقورن بكل من النموذجين الأحادي والثلاثي البارمتر وأيضا مع نتائج دراسة العكايلة (٢٠٠٧) التي أكدت الدقة العالية لتقدير النموذج الثلاثي البارمتر ، كما تتسق هذه النتائج مع النتائج التي انتهت إليها دراسة (DeMars,2001) أن هناك أفضلية لاستخدام النموذج الثلاثي البارمتر إذا ماقورن بالنموذج الأحادي البارمتر ، كما اتسقت كذلك مع نتائج دراسة **الوليلي (٢٠٠١)** التي انتهت إلى ان النموذج الثلاثي البارمتر هو الأكثر دقة في انتقاء البيانات مقارنة بالنموذج الأحادي.

ويرى الباحث الحالي ان الفروق التي جاءت في صالح النموذجين الثنائي والثلاثي مقابل النموذج الأحادي البارمتر وذلك عند مستوى أعلى من القدرة قد تكون نتيجة منطقية أيضا نظرا لاعتماد كل من هذين النموذجين على بارمترات إضافية في تقدير أقصى قيمة لدوال معلومات الاختبار حيث يعتمد النموذج الثنائي على بارمتر التمييز كما يعتمد النموذج الثلاثي على بارمتري التمييز والتخمين وذلك نظرا لأن تقديرات القدرات ذات المستوى الأعلى يعتمد على مفردات اختبارية أكثر صعوبة وأكثر تمييزا عن تقديرات القدرة ذات المستوى الأقل والتي ليست بحاجة على مفردات تمتلك مستويات أعلى من الصعوبة والتمييز ومما يتطلب في بعض الأحيان لجوء الفرد على أسلوب التخمين كأسلوب بديل للوصول على الاستجابة الصواب ، وبالتالي فإن إعطاء قيم أعلى للمعلومات عند نقاط قدرة (مستويات قدرة أعلى) يكون مرتبطا على حد كبير بتقديرات التمييز والتخمين والذي يعتمد عليهما كل من هذين النموذجين ، ولايعتمد عليهما النموذج الأحادي البارمتر ،مما قد يؤدي إلى

إعطاء تقديرات أعلى لأقصي قيمة لمعلومات المفردة الاختبارية باستخدام تجليلات هذين النموذجين.

(ج) نتائج التحليلات الخاصة بالمقارنة بين متوسطات قيم دوال معلومات بنود الاختبار الناتجة من التحليلات الخاصة بالنماذج الثلاثة

جدول (17)

المتوسطا والانحراف المعياري لقيم متوسطات دوال معلومات مفردات اختبار المصفوفات المتتابعة الناتجة عن التحليل باستخدام النماذج موضع المقارنة

نموذج التحليل	متوسط قيم متوسط دوال المعلومات	الإنحراف المعياري
أحادي البارمتر	.132303	.0598805
ثنائي البارمتر	.192660	.1706060
ثلاثي البارمتر	.169083	.1562833

يتضح الجدول السابق ان هناك فروق ظاهرية بين متوسطات قيم دوال معلومات بنود الاختبار حيث بلغت أكبر قيمة لمتوسط قيم دوال معلومات بنود الاختبار القيمة (0.19) تقريبا وكانت ناتجة من تحليلات النموذج ثنائي البارمتر ، بينما بلغت أقل قيمة لمتوسط قيم دوال معلومات بنود الاختبار القيمة (0.13) وكانت للنموذج أحادي البارمتر. مما يعنى أن النموذج ثنائي البارمتر يعطي أكبر متوسط لقيم دوال معلومات بنود الاختبار يلية النموذج ثلاثي البارمتر ، أما النموذج الأحادي البارمتر يعطي أقل قيمة لدوال المعلومات بنود الاختبار.

جدول (18)

اختبارات دلالة تباين متوسطات تقديرات متوسطات قيم المعلومات الناتجة عن تحليل مفردات اختبار المصفوفات المتتابعة بالنماذج الثلاثة

الاختبار الإحصائي	قيمة الاختبار	قيمة التباين	درجات الحرية	مستوى الدلالة
Pillai's Trace	.233	8.794	2	.000
Wilks' Lambda	.767	8.794	2	.000
Hotelling's Trace	.303	8.794	2	.000
Roy's Largest Root	.303	8.794	2	.000

يتضح من الجدول السابق الذي يتضمن نتائج الاختبارات المختلفة التي تبرز دلالة التباين بين متوسطات قيم دوال معلومات بنود الاختبار الناتجة عن التحليل بالنماذج الثلاثة (الأحادي – الثنائي – الثلاثي) البارمتر ، ان جميع الاختبارات وأشهرها اختبار ولكس لامدا 'Wilks Lambda قد بلغت قيمته (.767) وهي قيمة دالة عند مستوى دلالة قدره (0.00) . مما يشير إلى وجود دلالة للتباين بين متوسطات قيم دوال معلومات بنود الاختبار الناتجة عن التحليل بين النماذج الثلاث تعود إلى نموذج التحليل المستخدم.

جدول (19)

اختبار "ماكولي " للدورية بين تقديرات متوسطات قيم معلومات

لمفردات اختبار المصفوفات المتتابعة الناتجة عن التحليل بالنموذج المستخدم

قيمة اختبار ماكولي	مربع كاي التقريبية	درجات الحرية	مستوى الدلالة
.346	61.536	2	0.00

يتضح من الجدول السابق ان قيمة اختبار ماكولي للدورية بلغت (.346) وهي قيمة دالة حيث بلغت قيمة قيمة الدلالة المقترنة بها القيمة (0.00) مما يشير إلى وجود اختلاف جوهريا بين تجانس القياسات الثلاثة الناتجة عن التحليل بالنماذج الثلاثة المستخدمة مما يشير إلى عدم تحقق شرط الدورية ومما يستوجب في مثل هذه الحالة استخدام اختبار " جرين هاوس – جيسر" Greenhouse-Geisser عند اختبار دلالة التباين بين القياسات الثلاثة الناتجة عن التحليل باستخدام النموذج المستخدم ، وهو مايتضح من خلال الجدول التالي:

جدول (20)

تحليل تباين القياسات الثلاثة لمتوسطات قيم معلومات مفردات اختبار الدراسة

و الناتجة عن التحليل باستخدم نموذج التحليل المستخدم

مصدر التباين	الاختبار	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة (ف)	مستوى الدلالة
بين القياسات	Sphericity Assumed	.111	2	.056	8.367	.000
	Greenhouse-Geisser	.111	1.209	.092	8.367	.003

مصدر التباين	الاختبار	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة (ف)	مستوى الدلالة
	Huynh-Feldt	.111	1.221	.091	8.367	.003
	Lower-bound	.111	1.000	.111	8.367	.005
داخل القياسات (الخطأ)	Sphericity Assumed	.783	118	.007		
	Greenhouse-Geisser	.783	71.348	.011		
	Huynh-Feldt	.783	72.033	.011		
	Lower-bound	.783	59.000	.013		

يتضح من الجدول السابق أن قيمة اختبار التباين (f) باستخدام اختبار "جرين هاوس – جيسر" بلغت (8. 367) وهي قيمة دالة حيث بلغت قيمة الدلالة الخاصة بها (0.003) مما يشير إلى وجود فروق جوهرية بين متوسطات القياسات الخاصة بقيم دوال معلومات بنود الاختبار وللتعرف على اتجاه هذه الفروق يجب استخدام جدول المقارنات الثنائية **Pairwise Comparisons** بطريقة **Lsd** وهي موضحة بالجدول التالي:

جدول (21)

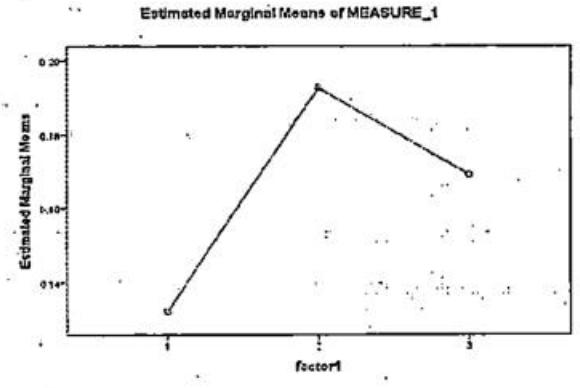
المقارنات الثنائية بطريقة (LSD) بين متوسطات تقديرات قيم دوال معلومات مفردات اختبار الدراسة باستخدام نماذج التحليل موضع المقارنة

طرف المقارنة الأول	طرف المقارنة الثاني	الفرق بين متوسطي طرفي المقارنة (الأول – الثاني)
النموذج الأحادي	النموذج الثنائي	-.060*
	النموذج الثلاثي	-.037*
النموذج الثنائي	النموذج الأحادي	.060*
	النموذج الثلاثي	.024*

*دال عند مستوى (0.05)

يتضح من الجدول السابق أن هناك فروقاً جوهرية بين متوسطي قيم دوال معلومات بنود الاختبار باستخدام نموذجي التحليل الأحادي والثنائي البارمتر في صالح نموذج التحليل ثنائي البارمتر. كما يتضح وجود فروق جوهرية بين متوسطي قيم دوال معلومات بنود الاختبار باستخدام

نموذجي التحليل الأحادي والثلاثي البارمتر في صالح نموذج التحليل الثلاثي البارمتر ، كما يتضح أيضاً وجود فروق جوهرية بين نموذجي التحليل التنائي والثلاثي البارمتر في صالح النموذج الثنائي البارمتر.



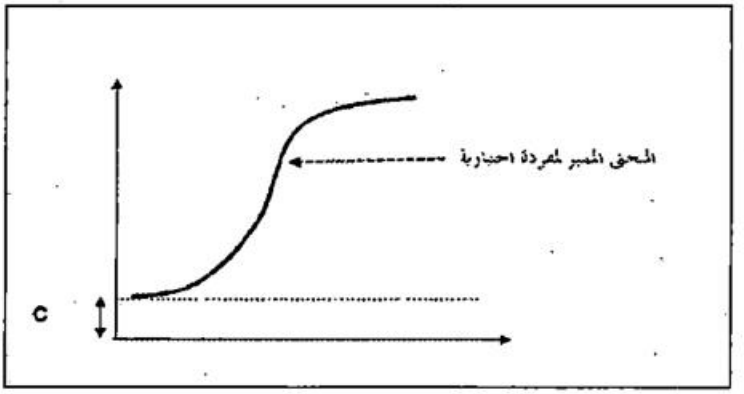
شكل (4)

تمثيل بياني لمتوسطات تقديرات متوسط قيم دوال مفردات اختبار الدراسة للنماذج الثلاث موضع المقارنة [الأحادي (1) و الثنائي (2) والثلاثي (3)]

يتضح النتائج السابقة وكذلك التمثيل البياني السابق أن أعلى متوسط لقيم دوال معلومات بنود الاختبار كانت ناتجة عن التحليل باستخدام النموذج الثنائي البارمتر ، يلي ذلك متوسط قيم دوال معلومات بنود الاختبار الناتجة عن التحليل باستخدام نموذج الثلاثي البارمتر، بينما جاءت أدني قيمة لمتوسط دوال معلومات بنود الاختبار تلك القيمة التي نتجت من التحليلات الخاصة بالنموذج أحادي البارمتر. مما يشير إلى ان نتائج التحليل باستخدام نموذج التحليل ثنائي البارمتر تحقق أفضل نتائج إذا ماقورنت بنتائج التحليلات الخاصة بنموذجي التحليل الآخرين. وتتسق هذه النتيجة مع عدد من النتائج التي انتهت إليها بعض الدراسات التي السابقة ، مثل دراسة Van Der& Han.(2011,) التي أثبتت تفوق النموذج الثنائي البارمتر على النماذج الأخرى في اتخاذ القرارات فيما يتعلق بالفروق الفردية للأفراد ومن بين هذه الدراسات أيضا دراسة (Raykov,Tet.al.,2010) إلى تفوق النموذج الثنائي البارمتر في الوصول إلى اعلى تقدير

للثبات إذا ماقون بالنموذج الأحادي البارمتر للبيانات الاختبارية ثنائية التقسيم كما أعطى النموذج أخطاء معيارية للتقدير أقل من نظائرها بالنموذج الأحادي البارمتر وكذلك دراسة Daniel,) (2010, ودراسة (Courville,2004) فقد توصلتا إلى تفوق النموذج الثنائي البارمتر على النموذج الأحادي .وهذه النتيجة تتسق كذلك مع ما توصلت إليه دراسة (Jaio,2003) حيث 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 والثلاثي البارمتر .

ويرى الباحث أن النتيجة التي تتعلق بتفوق النموذج الثنائي البارمبر على النموذج الثلاثي البارمتر قد تعود إلى متغير التخمين 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 (5) – (Lord ,1980: 12-13) وقد يرى الباحث 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 المعلومات التي يمكن الحصول عليها باستخدام النموذج الثالث.



شكل (5)

قيمة التخمين (c) لمفردة اختبارية

ومن خلال النتائج السابقة الخاصة بالتقديرات الثلاث لقيم دوال معلومات مفردات اختبار الدراسة (أقصى معلومات – نقاط القدرة المقابلة لأقصى معلومات – متوسط المعلومات) يمكن رفض الفرض الأول الذي ينص على: **لا تختلف تقديرات القدرة للأفراد باختلاف نموذج التحليل المستخدم.**

رابعاً :مناقشة وعرض النتائج الخاصة بدقة تقديرات القدرة

اعتمدت الدراسة في الكشف عن دقة تقديرات قدرات عينة الدراسة الناتجة عن تحليلات نماذج التحليل موضع المقارنة على استخدام مؤشرين :

(أ) مؤشر ثبات دوال معلومات الاختبار.

(ب) مؤشر الأخطاء المعيارية لقياس القدرة:

(أ) : نتائج التحليلات الخاصة بمؤشر ثبات دوال معلومات الاختبار.

أسفرت التحليلات الخاصة بهذا المؤشر عن البيانات الموضحة بالجدول التالي:

جدول (22)

المتوسط والإنحراف المعياري لقيم ثبات دوال معلومات الاختبار

نموذج التحليل	متوسطات قيم ثبات دوال	الإنحرافات المعيارية
أحادي البارمتر	.114333	.0484756
ثنائي البارمتر	.196622	.1701749
ثلاثي البارمتر	.140343	.1064882

يتضح من نفس الجدول السابق ان هناك فروق ظاهرية بين متوسطات قيم الثبات لتقدير دوال معلومات بنود الاختبار حيث بلغت أكبر قيمة لمتوسط ثبات تقدير دوال معلومات بنود الاختبار القيمة (0.19) تقريبا وكانت ناتجة من تحليلات النموذج الثنائي البارمتر ، بينما بلغت أقل قيمة لمتوسط قيم ثبات تقديرات دوال معلومات بنود الاختبار القيمة (0.11) وكانت للنموذج أحادي البارمتر مما يعنى أن النموذج الثنائي البارمتر يعطي أكبر متوسط لقيم ثبات تقديرات دوال معلومات

بنود الاختبار يلية النموذج ثلاثي البارمتر ، أما النموذج الأحادي البارمتر فكان له أقل قيمة.

جدول (23)

اختبارات دلالة تباين قيم ثبات دوال معلومات مفردات اختبار الدراسة
الناتجة عن تحليل مفردات اختبار المصفوفات المتتابعة بالنماذج الثلاثة

الاختبار	قيمة الاختبار	قيمة التباين	درجات الحرية	مستوى الدلالة
Pillai's Trace	.252	9.747	2	.000
Wilks' Lambda	.748	9.747	2	.000
Hotelling's Trace	.336	9.747	2	.000
Roy's Largest Root	.336	9.747	2	.000

يتضح من الجدول السابق الذي يتضمن نتائج الاختبارات المختلفة التي توضح دلالة التباين بين متوسطات قيم الثبات لتقدير دوال معلومات بنود الاختبار الناتجة عن التحليل بالنماذج الثلاثة (الأحادي – الثنائي – الثلاثي) البارمتر ، ان جميع الاختبارات وأشهرها اختبار ولكس لامدا Wilks' Lambda قد بلغت قيمته (.748) وهي قيمة دالة عند مستوى دلالة قدره (0.00) . مما يشير إلى وجود دلالة للتباين بين متوسطات قيم الثبات لتقدير دوال معلومات بنود الاختبار الناتجة عن التحليل بين النماذج الثلاث تعود إلى نموذج التحليل المستخدم.

جدول (24)

دلالة اختبار ماكولي للتحقق من شرط الدورية لتقديرات الثبات
لدوال المعلومات لمفردات اختبار الدراسة

قيمة اختبار ماكولي	مربع كاي التقريبية	درجات الحرية	مستوى الدلالة
.555	34.169	2	0.00

يتضح من الجدول السابق ان قيمة اختبار ماكولي للدورية بلغت (.555) وهي قيمة دالة حيث بلغت قيمة قيمة الدلالة المقترنة بها القيمة (0.00) مما يشير إلى وجود اختلاف جوهري بين تجانس

القياسات الثلاثة الناتجة عن التحليل بالنماذج الثلاثة المستخدمة مما يشير إلى عدم تحقق شرط الدورية ومما يستوجب في مثل هذه الحالة استخدام اختبار " جرين هاوس – جيسر" Greenhouse-Geisser عند اختبار دلالة التباين بين القياسات الثلاثة الناتجة عن التحليل باستخدام النموذج المستخدم ، وهو مايتضح من خلال الجدول التالي:

جدول (25)

تحليل تباين القياسات الثلاثة لثبات قيم دوال معلومات

مفردات اختبار الدراسة الناتجة عن نموذج التحليل المستخدم

مصدر التباين	الاختبار	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة (ف)	مستوى الدلالة
بين القياسات	Sphericity Assumed	.212	2	.106	14.482	.000
	Greenhouse-Geisser	.212	1.384	.153	14.482	.000
	Huynh-Feldt	.212	1.406	.151	14.482	.000
	Lower-bound	.212	1.000	.212	14.482	.000
داخل القياسات (الخطأ)	Sphericity Assumed	.865	118	.007		
	Greenhouse-Geisser	.865	81.650	.011		
	Huynh-Feldt	.865	82.980	.010		
	Lower-bound	.865	59.000	.015		

يتضح من الجدول السابق أن قيمة اختبار التباين (f) باستخدام اختبار " جرين هاوس – جيسر" بلغت (14. 482) وهي قيمة دالة حيث بلغت قيمة الدلالة الخاصة بها (0.00) مما يشير إلى وجود فروق جوهرية بين متوسطات القياسات الخاصة بقيم ثبات تقديرات دوال معلومات بنود الاختبار وللتعرف على اتجاه هذه الفروق يمكن استخدام جدول المقارنات الثنائية Pairwise Comparisons بطريقة Lsd وهي موضحة بالجدول التالي:

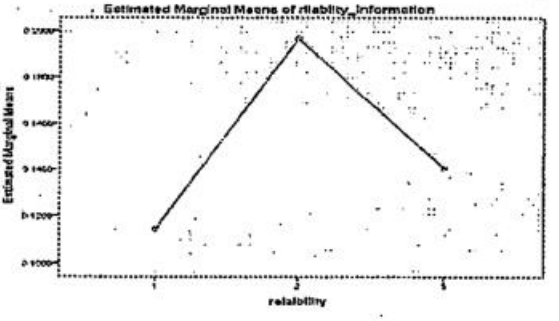
جدول (26)

المقارنات الثنائية بطريقة (LSD) بين متوسطات تقديرات ثبات قيم دوال معلومات مفردات اختبار الدراسة باستخدام نماذج التحليل موضع المقارنة

طرف المقارنة الأول	طرف المقارنة الثاني	الفرق بين متوسطي طرفي المقارنة (الأول – الثاني)
النموذج الأحادي	النموذج الثنائي	-.082*
	النموذج الثلاثي	-.026*
النموذج الثنائي	النموذج الأحادي	.082*
	النموذج الثلاثي	.056*

*دال عند مستوى (0.05)

يتضح من الجدول السابق أن هناك فروقاً جوهرية بين متوسطي قيم الثبات لتقدير دوال معلومات بنود الاختبار باستخدام نموذجي التحليل الأحادي والثنائي البارمتر في صالح نموذج التحليل ثنائي البارمتر. كما يتضح أيضا أن هناك فروقاً جوهرية بين متوسطي قيم الثبات لتقدير دوال معلومات بنود الاختبار باستخدام نموذجي التحليل الأحادي والثلاثي البارمتر في صالح نموذج التحليل الثلاثي البارمتر ، كما يتضح أيضا وجود فروق جوهرية بين نموذجي التحليل الثنائي والثلاثي البارمتر في صالح النموذج الثنائي البارمتر.



شكل (6)

تمثيل بياني لمتوسطات تقديرات ثبات قيم دوال مفردات اختبار الدراسة للنماذج الثلاث موضع المقارنة [الأحادي (1) و الثنائي (2) والثلاثي (3)]

يتضح من التمثيل البياني السابق السابق أن أعلى متوسط قيم الثبات لتقدير دوال معلومات بنود الاختبار كانت ناتجة عن التحليل باستخدام النموذج الثنائي البارمتر ، يلي ذلك متوسط قيم الثبات لتقدير دوال معلومات بنود الاختبار الناتجة عن التحليل باستخدام نموذج الثلاثي البارمتر، بينما جاءت أدني قيمة لمتوسط قيم الثبات لتقدير دوال معلومات كانت ناتجة عن التحليلات الخاصة بالنموذج أحادي البارمتر. مما يشير إلى ان نتائج التحليل باستخدام نموذج التحليل ثنائي البارمتر تحقق أفضل نتائج إذا ماقورنت بالنتائج الخاصة بنموذجي التحليل الآخرين.

وتتسق هذه النتيجة النتائج التي انتهت إليها بعض الدراسات ، مثل دراسة (Raykov,T,et.al.,2010) التي انتهت إلى إلى تفوق النموذج الثنائي البارمتر في الوصول إلى اعلى تقدير للثبات إذا ماقون بالنموذج الأحادي البارمتر للبيانات الاختبارية ثنائية التقسيم كما أعطى النموذج أخطاء معيارية للتقدير أقل من نظائرها بالنموذج الأحادي البارمتر ، و كذلك دراسة (Raykov,T,et.al.,2010)والتي أشارت إلى تفوق النموذج الثنائي البارمتر في الوصول إلى اعلى تقدير للثبات إذا ماقون بالنموذج الأحادي البارمتر للبيانات الاختبارية ثنائية التقسيم كما أعطى النموذج أخطاء معيارية للتقدير أقل من نظائرها بالنموذج الأحادي البارمتر وكذلك دراسة Van Der& Han.(2011,) التي أثبتت تفوق النموذج الثنائي البارمتر على النماذج الأخرى في اتخاذ القرارات فيما يتعلق بالفروق الفردية للأفراد وكذلك دراسة كل من Daniel,(2010, Courville,2004) التي توصلتا الدراسة إلى تفوق النموذج الثنائي البارمتر على النموذج الأحادي .وهذه النتيجة تتسق كذلك مع ما توصلت إليه دراسة (Jaio,2003) حيث 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.

ويرى الباحث مثلما أشار سابقا إلى ذلك من قبل، ان النتيجة التي تتعلق بتفوق النموذج الثنائي

البارمبر على النموذج الثلاثي البارمتر قد تعزى إلى متغير التخمين حيث أن قيمة التخمين (c) قد تعزي إلى سهولة المفردة ولاتعزي لمستوى التخمين لدى الممتحن الغير قادر على الإجابة الصواب ، مما قد يؤثر في النهاية على دقة المعلومات التي يمكن الحصول عليها باستخدام النموذج الثالث.

(ب) : نتائج التحليلات الخاصة بدقة تقدير قدرات عينة الدراسة باستخدام الأخطاء المعيارية للقياس

وقد اعتمدت الدراسة على مؤشر الخطا المعياري لقياس القدرة كمؤشر لتحديد دقة قياس وتقدير قدرات الأفراد ، فالتقدير الأدق هو ذلك التقدير الذي يمتلك أقل خطا معياري للقياس ، وقد اعتمدت الدراسة في إجراءاتها على دلالة الفروق بين متوسطات الأخطاء المعيارية لتقدير قدرات الأفراد الناتجة عن التحليل الخاص بالنماذج المختلفة وذلك لتحديد دقة تقديرات القدرة الناتجة الخاصة بكل نموذج على حدة والجداول التالية توضح النتائج المتعلقة بهذا الجانب:

جدول (27)

المتوسط والإنحراف المعياري لقيم الأخطاء المعيارية لتقدير القدرة المشتقة من نماذج التحليل

نموذج التحليل	متوسطات قيم الأخطاء المعيارية لتقديرات	الإنحرافات المعيارية
أحادي البارمتر	.367598	.0515899
ثنائي البارمتر	.329070	.1312613
ثلاثي البارمتر	.360789	.1902273

يتضح من الجدول السابق أن هناك فروق ظاهرية بين متوسطات الأخطاء المعيارية لتقديرات القدرة الناتجة عن التحليل الخاص بكل نموذج من النماذج الثلاثة المستخدمة بالدراسة، حيث بلغت أكبر قيمة لمتوسط الأخطاء المعيارية القيمة (0.368) تقريبا وكانت ناتجة من تحليلات النموذج أحادي البارمتر ، بينما بلغت أقل قيمة لمتوسط الأخطاء المعيارية القيمة (0.329) تقريبا

وكانت للنموذج الثنائي البارمتر ، مما يعنى أن تقديرات القدرة باستخدام التحليلات الخاصة بالنموذج الثنائي البارمتر كانت بأقل الأخطاء المعيارية.

جدول (28)

اختبارات دلالة تباين قيم الأخطاء المعيارية لتقديرات

الناتجة عن تحليل مفردات اختبار المصفوفات المتتابعة بالنماذج الثلاثة

الاختبار	قيمة الاختبار	قيمة التباين	درجات الحرية	مستوى الدلالة
Pillai's Trace	.183	51.414	2	.000
Wilks' Lambda	.817	51.414	2	.000
Hotelling's Trace	.224	51.414	2	.000
Roy's Largest Root	.224	51.414	2	.000

يتضح من الجدول السابق الذي يتضمن نتائج الاختبارات المختلفة التي تبرز دلالة التباين بين متوسطات قيم دوال معلومات بنود الاختبار الناتجة عن التحليل بالنماذج الثلاثة (الأحادي – الثنائي – الثلاثي) البارمتر ، ان جميع الاختبارات وأشهرها اختبار ولكس لامدا **Wilks' Lambda** قد بلغت قيمته (817 .) وهي قيمة دالة عند مستوى دلالة قدره (0.00) . مما يشير إلى وجود دلالة للتباين بين متوسطات الخطاء المعيارية لتقديرات القدرة الناتجة عن التحليل بين النماذج الثلاث تعود إلى نموذج التحليل المستخدم.

جدول (29)

دلالة اختبار ماكولي للتحقق من شرط الدورية لقيم أخطاء تقديرات القدرة الناتجة عن تحليل استجات عينة الدراسة على اختبار المصفوفات المتتابعة باستخدام نماذج التحليل موضع المقارنة

قيمة اختبار ماكولي	مربع كاي التقريبية	درجات الحرية	مستوى الدلالة
.564	263.620	2	0.00

يتضح من الجدول السابق ان قيمة اختبار ماكولي للدورية بلغت (.564) وهي قيمة دالة حيث بلغت قيمة قيمة الدلالة المقترنة بها القيمة (0.00) مما يشير إلى وجود اختلافا جوهريا بين تجانس القياسات الثلاثة الناتجة عن التحليل بالنماذج الثلاثة المستخدمة مما يشير إلى عدم تحقق شرط الدورية ومما يستوجب في مثل هذه الحالة استخدام اختبار " جرين هاوس – جيسر" Greenhouse-Geisser عند اختبار دلالة التباين بين القياسات الثلاثة الناتجة عن التحليل باستخدام النموذج المستخدم ، وهو مايتضح من خلال الجدول التالي:

جدول (30)

تحليل تباين القياسات الثلاثة لقيم الأخطاء المعيارية لتقديرات القدرة الناتجة عن تحليل استجات عينة الدراسة على اختبار المصفوفات المتتابعة باستخدام نموذج التحليل المستخدم

مصدر التباين	الاختبار	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة (ف)	مستوى الدلالة
بين القياسات	Sphericity Assumed	.391	2	.195	17.667	.000
	Greenhouse-Geisser	.391	1.393	.281	17.667	.000
	Huynh-Feldt	.391	1.395	.280	17.667	.000
	Lower-bound	.391	1.000	.391	17.667	.000
داخل القياسات (الخطأ)	Sphericity Assumed	10.194	922	.011		
	Greenhouse-Geisser	10.194	641.964	.016		
	Huynh-Feldt	10.194	643.300	.016		
	Lower-bound	10.194	461.000	.022		

يتضح من الجدول السابق أن قيمة اختبار التباين (f) باستخدام اختبار " جرين هاوس – جيسر " بلغت (17.667) وهي قيمة دالة حيث بلغت قيمة الدلالة الخاصة بها (0.00) مما يشير إلى وجود فروق جوهرية بين متوسطات الأخطاء المعيارية لتقدير القدرة الناتجة عن التحليل باستخدام النماذج الثلاث موضع المقارنة وللتعرف على اتجاه هذه الفروق يجب استخدام جدول المقارنات الثنائية Pairwise Comparisons بطريقة LSD وهي موضحة بالجدول التالي:

جدول (31)

المقارنات الثنائية بطريقة (LSD) بين متوسطات تقديرات الأخطاء المعيارية لتقديرات القدرة الناتجة عن تحليل استجات

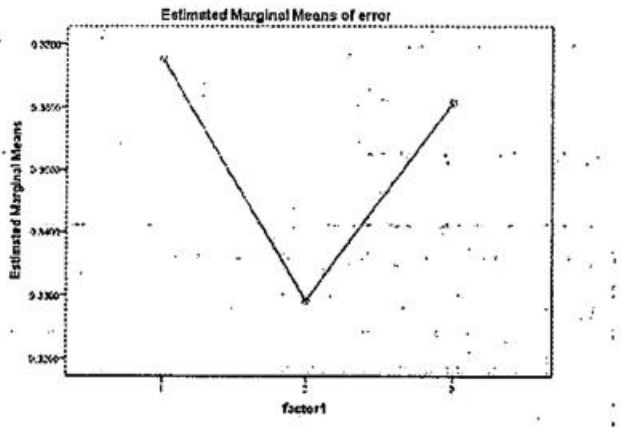
عينة الدراسة على اختبار المصفوفات المتتابعة باستخدام نماذج التحليل موضع المقارنة

طرف المقارنة الأول	طرف المقارنة الثاني	الفرق بين متوسطي طرفي المقارنة (الأول – الثاني)
النموذج الأحادي	النموذج الثنائي	.039*
	النموذج الثلاثي	.007
النموذج الثنائي	النموذج الأحادي	-.039*
	النموذج الثلاثي	-.032*

*دال عند مستوى (0.05)

يتضح من الجدول السابق وجود فروق جوهرية بين متوسطي الأخطاء المعيارية لتقديرات القدرة الناتجة عن التحليل باستخدام النموذجين الأحادي والثنائي البارمتر، في صالح النموذج الأحادي البارمتر، كما يتضح أيضا أن هناك فروق جوهرية بين متوسطي الأخطاء المعيارية لتقديرات القدرة الناتجة عن التحليل باستخدام النموذجين الثنائي والثلاثي البارمتر في صالح نموذج التحليل الثلاثي البارمتر ، كما يتضح عدم وجود فروق جوهرية بين نموذجي التحليل الأحادي والثلاثي البارمتر .

ويتضح من النتيجة السابقة أن متوسط الأخطاء المعيارية لتقديرات القدرة الناتجة عن التحليل باستخدام النموذج الثنائي البارمتر جاء اقل بصورة جوهرية من متوسطي تقديرات القدرة التي اسفرت عنها التحليلات الخاصة بالنموذجين الآخرين (الأحادي والثلاثي) البارمتر مما يشير إلى ان تقديرات القدرة الخاصة بهذا النموذج كانت أكثر دقة من التقديرات التي أسفرت عنها التحليلات الخاصة بالنموذجين الآخرين .



شكل (7)

تمثيل بياني لمتوسطات تقديرات الأخطاء المعيارية لتقديرات القدرة الناتجة عن تحليل استجابات عينة الدراسة باستخدام النماذج الثلاث موضع المقارنة [الأحادي (1) و الثنائي (2) والثلاثي (3)]

كما يبدو واضحا من التمثيل البياني الذي يشير على ان متوسط الأخطاء المعيارية لتقديرات القدرة الناتجة عن التحليلات الخاصة بالنموذج الثنائي البارمتر جاء أقل بصورة واضحة عن متوسطي الأخطاء المعيارية لتقديرات القدرة الناتجة عن التحليل باستخدام النموذجين الثنائي والثلاثي البارمتر.

وتتسق هذه النتيجة أيضا مع النتائج التي انتهت إليها بعض الدراسات ، مثل دراسة (Raykov,T., et,al.,2010) التي انتهت إلى تفوق النموذج الثنائي البارمتر في الوصول إلى اعلى تقدير للثبات إذا ماقون بالنموذج الأحادي البارمتر للبيانات الاختبارية ثنائية التقسيم كما أعطى النموذج أخطاء معيارية للتقدير أقل من نظائرها بالنموذج الأحادي البارمتر ، وكذلك دراسة Van Der& Han.(2011.) التي أثبتت تفوق النموذج الثنائي البارمتر على النماذج الأخرى وكذلك دراسة Daniel,(2010)التي أثبتت تفوق النموذج الثنائي البارمتر على النماذج الأخرى في دقة تقديرات القدرة ، ودراسة (Courville,2004) التي توصلت الدراسة إلى دقة النموذج الثنائي البارمتر إذا ماقورن بالنموذج الأحادي، وهذه النتيجة تتسق كذلك مع ما توصلت إليه دراسة (Jaio,2003) حيث 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.

ومن خلال النتائج السابقة الخاصة بدقة تقديرات القدرة الناتجة عن التحليل نموذج التحليل المستخدم يمكن رفض الفرض الثاني الذي ينص على " لا تختلف دقة تقديرات القدرة – ممثلة بالخطا المعياري للتقدير – وثبات قيم دوال المعلومات باختلاف نموذج التحليل المستخدم."

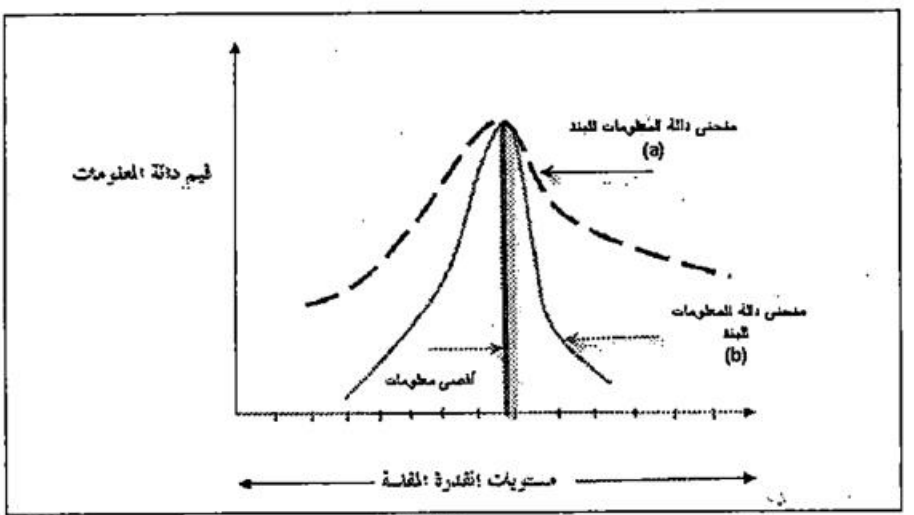
خامسا : النتائج الخاصة بالعلاقة بين تقديرات قيم دوال المعلومات ودقة تقديرات القدرة

وقد اقتصر الباحث في الكشف عن العلاقة بين تقديرات دوال المعلومات الناتجة عن نموذج التحليل المستخدم (الأحادي والثنائي والثلاثي البارمتر) ودقة تقديرات قدرات أفراد عينة الدراسة (كما قدرت بمؤشري ثبات دوال المعلومات والخطأ المعياري للتقدير) على استخدام التحليلات الكيفية والتي تقتصر على مقابلة نتائج التحليلات الخاصة بالفروض السابقة الخاصة بقيم دوال المعلومات ونتائج التحليلات الخاصة بمؤشري ثبات قيم دوال المعلومات والأخطاء المعيارية لتقديرات القدرة .

لقد اتضح من الجداول السابقة أرقام (16 , 21) التي انتهت إليها الدراسة ، أن النموذج الثنائي البارمتر قد تفوق على النموذجين الآخرين (الأحادي والثلاثي البارمتر) في نوعين من تقديرات دوال معلومات بنود الاختبار وهما (متوسط دوال معلومات بنود الاختبار ، و نقاط القدرة التي تقابل أقصى معلومات لكل بند من بنود الاختبار) حيث أسفرت التحليلات الخاصة بهذا النموذج عن أعلى تقدير دال لمتوسط متوسطات قيم دوال المعلومات بنود الاختبار وكذلك أعلى تقديرات دالة للمعلومات عند مستويات اعلى من مستويات القدرة للأفراد إذا قورن بالنموذجين الآخرين ، كما تبين أيضا أن نفس النموذج يعطي أعلى قيم لثبات تقدير دوال معلومات بنود الاختبار كما يعطي أقل تقديرات للأخطاء المعيارية لتقديرات القدرة ، بينما نجد ان النموذج الثالث وكذلك الثاني أسفرت تحليلاتهما عن أعلى تقديرات لأقصى قيم لدوال معلومات بنود الاختبار بصورة عامة إذا قورن كلاهما بالنموذج الأحادي البارمتر، في حين لاتوجد فروق جوهرية بين هذين النموذجين في هذا النوع من المعلومات ، كما تضح من الجدول رقم (11) بينما افتقرت نتائج تقدير المعلومات التي أسفرت عنها تحليلات النموذج الثالث إلى دقة تقديرات القدرة كما بدا ذلك من خلال مؤشرات دقة تقديرات القدرة لهذا النموذج ، كما هو مبين بالجدول رقم (27 ، 32) وبالتالي افتقر هذا النموذج وكذلك النموذج الأحادي البارمتر إلى نفس الدقة التي تمتع بها النموذج الثنائي البارمتر، وإن دل ذلك على شئ فإنما يدل على أن هناك علاقة بين تقديرات قيم دوال معلومات بنود الاختبار فيما يتعلق بمتوسط قيم دوال معلومات بنود الاختبار من ناحية ودقة تقديرات القدرة من ناحية اخرى كما بدت من تحليلات النموذج الثاني فقط.

وقد تفسر هذه النتيجة والتي تتمثل في تفوق النموذج الثنائي البارمتر على النموذجين الاخرين

في دقة تقديرات القدرة ، بأن النموذج الثنائي البارمتر تفوق على النموذجين الآخرين فيما يتعلق بمتغير متوسط دوال معلومات بنود الاختبار - على الرغم من عدم وجود فروق جوهرية بين النموذجين (الثنائي والثلاثي) البارمتر فيما يتعلق بتقديرات قيم أقصى معلومات لبنود اختبار الدراسة - حيث يرى الباحث أن متغير متوسط قيم معلومات بنود المعلومات عبر متصل القدرة المقاسة أفضل واكثر صدقا وواقعية عن متغير أقصى معلومات لبنود الاختبار والذي يتعلق فقط بنقاط محددة على متصل القدرة المقاسة وليس لجميع النقاط على متصل القدرة ، فمن المعروف أن أنه كلما كانت قيم المعلومات التي تعطيها المفردة أعلى عند مستوى معين من القدرة كلما كان أفضل من بند آخر يعطي معلومات أقل عند نفس المستوى من القدرة وبالتالي فإنه من المنطقي والطبيعي أنه كلما كان إزداد المتوسط العام لقيم دوال المعلومات لمفردة ما من مفردات الاختبار لجميع نقاط (مستويات) القدرة على متصل القدرة ، كلما كانت هذه المفردة أفضل من مفردة أخرى تمتلك متوسط عام أقل لقيم دوال المعلومات على نفس متصل القدرة، ويمكن للباحث توضيح ذلك بالشكل التالي:



شكل (7)

منحنى دالة المعلومات لمفردتين اختباريتين افتراضيتين

ويوضح الشكل السابق أن منحنى دالة المعلومات للمفردة (a) – الممثل بالمنحنى المتقطع – له نفس قيمة أقصى معلومات لمنحنى دالة المعلومات للمفردة (b) – الممثل بالمنحنى المتصل – ولكن

المفردة (a) تعطي معلومات أعلى على متصل القدرة المقاسة عن المعلومات التي تعطيها المفردة (b) وبالتالي فإن متوسط المعلومات التي تعطيها المفردة (a) على جميع نقاط القدرة على متصل القدرة المقاسة أعلى من متوسط المعلومات التي تعطيها المفردة (b) وبالتالي قد تكون المفردة (a) أدق من المفردة (b) في تقديم معلومات أفضل عند جميع مستويات المقاسة.

جدول (33)

قيم متغيري (أقصى معلومات – متوسط المعلومات)لبعض بنود اختبار الدراسة الناتجة عن التحليل باستخدام النموذج الثلاثي البارمتر ، وقيم متوسط المعلومات لنفس المفردات الناتجة عن التحليل باستخدام النموذج الثنائي البارمتر

رقم المفردة	متوسط تقدير المعلومات	
	النموذج ثلاثي البارمتر (PL3)	النموذج ثنائي البارمتر (PL2)
7	0.2071	0.2851
11	0.479	0.6173
12	0.3614	0.4167
18	0.206	0.2782
23	0.2582	0.3596
28	0.1587	0.1848
29	0.1889	0.248
30	0.2146	0.2759
32	0.1266	0.1753
33	0.0811	0.1067
34	0.1907	0.2499
38	0.2108	0.2343
39	0.1864	0.1934
40	0.1128	0.1489
42	0.2159	0.256
44	0.2291	0.2611

رقم المفردة	متوسط تقدير المعلومات	
	النموذج ثلاثي البارمتر (PL3)	النموذج ثنائي البارمتر (PL2)
45	0.2235	0.2932
46	0.1717	0.1801
47	0.0965	0.1797
49	0.6105	0.629
51	0.3475	0.4411
55	0.3200	0.4787
56	0.1626	0.3994
57	0.1732	0.2580
58	0.0855	0.2467

فمن خلال النتائج الواردة بالجدول السابق يتضح أن قيمة متوسط دوال المعلومات الناتجة من تحليلات النموذج الثنائي البارمتر أعلى من نظائرها الناتجة من تحليلات النموذج الثلاثي البارمتر وذلك لجميع المفردات المعروضة بالجدول السابق فعلى سبيل المثال ، بلغ متوسط دوال المعلومات للمفردة رقم (12) الناتجة من تحليلات النموذج الثلاثي البارمتر بلغت (0.3614) فيي حين أن قيمة متوسط دوال المعلومات الناتجة من تحليلات النموذج الثاني للمفردة نفسه بلغت (0.4167)، مما يعني أن نفس المفردة قد أعطت معلومات أعلى بشكل عام مع تحليلات النموذج الثنائي البارمتر عن المعلومات التي اعطيت بنفس المفردة مع تحليلات النموذج الثلاثي البارمتر وكذلك الحال مع جميع المفردات الواردة بالجدول السابق، وبالتالي فإن دقة تقديرات القدرة التي تميز بها النموذج الثاني إذا ماقورن بالنموذجين الآخرين قد ترتبط بمتغير متوسط قيم معلومات بنود الاختبار أكثر من تعلقها بقيم أقصى معلومات لنفس المفردات. وتتسق هذه النتيجة مع النتائج التي توصلت إليها بعض الدراسات ، مثل دراسة .Van Der& Han)(2011, التي أثبتت تفوق النموذج الثنائي البارمتر على النماذج الأخرى وتوصلت كذلك دراسة (Raykov,T,et.al.,2010) إلى تفوق النموذج الثنائي البارمتر في الوصول إلى اعلى تقدير للثبات إذا ماقون بالنموذج الأحادي البارمتر للبيانات الاختبارية ثنائية التقسيم كما أعطى النموذج أخطاء معيارية للتقدير أقل من نظائرها بالنموذج الأحادي البارمتر ،وكذلك دراستي (Courville,2004)Daniel,((2010, التي أشارتا إلى دقة النموذج الثنائي البارمتر إذا ماقورن بالنموذج الأحادي، وكذلك دراسة (Jaio,2003) التي أظهرت 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، وبالتالي يمكن استنتاج أن دقة المعلومات التي اتسم بها النموذج الثنائي البارمتر إقترنت بدقة تقديرات القدرة كما تمثلت في متوسط قيم دوال معلومات بنود الاختبار مما يفيد بوجود ثمة علاقة بين المتوسط العام لقيم دوالمفردات الاختبار ودقة تقديرات القدرة التي اتضحت من خلال مؤشري الخطأ المعياري للقياس وثبات قيم دوال المعلومات.

ومن خلال النتائج السابقة الخاصة بعلاقة دوال معلومات مفردات الاختبار بدقة تقديرات القدرة الناتجة عن التحليل بالنموذج موضع المقارنة يمكن رفض الفرض الأول الذي ينص على: **لاتوجد علاقة بين بين دقة تقديرات القدرة للأفراد ودوال معلومات مفردات الاختبار.**

الخلاصة

لقد تبين من خلال نتائج الدراسة الحالية أن النموذج الثنائي البارامتر أفضل النماذج الثلاثة في تقديمه لأعلى تقديرات لقيم دوال معلومات فقرات الاختبار يليه النموذج الثلاثي البارمتر ثم الأحادي البارمتر، وهي نفس النتائج التي تم التوصل إليها فيما يتعلق بمتغيري الخطأ المعياري لتقدير القدرة وثبات تقدير دوال المعلومات وذلك كمؤشرين لدقة تقدير قدرات الممتحنين.

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 في تقديرات قدرات الأفراد.
٢. دراسة للكشف عن استخدام قيم دوال معلومات الاختبار للتعرف على الحجم المناسب من من مفردات والتي يمكن أن تعطي نفس تقديرات القدرة.
٣. دراسة تأثير حجم عينة الأفراد على دقة تقدير قيم دوال معلومات الاختبار.

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 وثبات التقديرات الخاصة بقيم دوال معلومات تلك المفردات المقارنة بالنموذجين الآخرين.

٢- استخدام دوال معلومات المفردات الاختبارية في تخفيض حجم عينة فقرات الاختبار- كإجراء أساسي يجب إدراجه ضمن إجراءات إعداد الاختبار – إلى الحجم المناسب الذي يعطي نفس تقديرات القدرة.

المراجع

١- أبو هاشم ، السيد محمد ، (٢٠٠٦). دراسة مقارنة بين النظرية التقليدية ونموذج راش في اختيار فقرات مقياس مداخل الدراسة لدى طلاب الجامعة ، مجلة كلية التربية ، جامعة الزقازيق ، العدد (٥٢).

٢- 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.

٣- الحكماني، رحاب سعيد ، (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.

٤- زكري ، علي عبد الله (٢٠٠٩). الخصائص السيكومترية لتطوير اختبار أوتيس لينون للقدرة العقلية مقدرة وفق القياس الكلاسيكي ونموذج ر اش لدى طلبة المرحلة المتوسطة بحافظة صبيا التعليمية ، رسالة دكتوراة غير منشورة ، بكلية التربية .

٥- الزيد ، جواهر محمد (٢٠٠٧). فاعلية نظرية الاستجابة للمفردة في بناء الاختبارات (إعداد بنك أسئلة في مقرر علم النفس التربوي). رسالة دكتوراه غير منشوره.جامعة الإمام محمد بن سعود الإسلامية ، كلية العلوم الاجتماعية.

٦- 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.

٧- علام، صلاح الدين محمود (٢٠٠٥). نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي ، القاهرة ،دار الفكر العربي.

٨- صفوت فرج (١٩٨٠) .القياس النفسي ، القاهرة، دار الفكر العربي.

٩- فؤاد أبو حطب.(١٩٩٦). القدرات العقلية ، القاهرة، مكتبة الأنجلو ، ط٥.

١٠- 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.

١١- 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.

١٢- معوض،خليل ميخائيل.(٢٠٠٧). القدرات العقلية.مصر.مركز الإسكندرية للكتاب. ط٤.

١٣- 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.

١٤- 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.

١٥- 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.

16- 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.

17- Aime,M,(2008). **A comparison of (CAT) scores based on based on different Measurement Models for testlets**; Educational and psychological measurement,31(3):371-398.

18- 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.

19- Azevedo,C.L.N. (2009): **Some Observations on the Identification and Interpretation of the 3PL IRT Model**; Applied Psychological Measurement , 33(7): 89–114.

20- 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.

21- Bryant, D. U. (2005). **A Note on Item Information in Any Direction for the Multidimensional Three-Parameter Logistic Model**. Psychometrika. 70(1):123-144

22- 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.

23- Daniel, Robert C.; Embretson, Susan E.(2010). **Designing Cognitive Complexity in Mathematical Problem-Solving Items.** Applied Psychological Measurement, 34(5) :348-364.

24- De Ayala,R.J.(2009).**The Theory and Practice of Item Response Theory**.NY.the Guilford press.

25- 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.

26- DeMars,C.(2001). **Group differences based on IRTscores:Does the model mater?** Educational and psychological measurement,61(1):60-70.

27- Edward, H. (2010) : **Interpretation of the Three-Parameter Testlet Response Model and Information Function**. Applied Psychological Measurement;,. 34 (7) :467-482.

28- Embretson SE, Reise SP (2000). **Item Response Theory for Psychologists**.London: Lawrence Erlbaum Associates.

29- Fan,a .(2009).**Item Response Theory Models :an Empirical Comparison of Their Item Statistics**. Educational and psychological measurement, 44 (2):.256-270.

30- Fumiko Samajima (2010). **Estimation of Reliability Using the item Information Function,** Applied Psychological Measurement feb., Vol. 18 No. 3 PP.229-244 .

31- 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

32- Hattie,J.:Methodology Review(1985).**Assessing Unidimensionality of Tests and Items** Applied Psychological Measurement;, 9 (2):139-164.

33- Hulin,C.Drasgow,F.& Parson,K.(1983):**Item Response Theory** : Pplications to Psychological Measurement .Hoomewood,Illinois:Dow Jones-Irwin.

34- 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,

35- Linden WJ, Hambleton RK.(2000).**Handbook Of modern item response theory**.Berlin: Springer.

36- 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.

37- Lord,f.(1980).**Applicatios Of Item Response Theory To Practical Testing Proplems**.hillsdale,n.j:lawrenve Erlbaum.

38- 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.

39- 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.

40- 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 Questionnaires.** Ophthalmic Epidemiology; Feb, 18 (1) :1-19.

41- Noel Y, Dauvier B. (2007). **A Beta Logistic Item Response Model For Continuous Bounded Responses**. Applied Psychological Measurement,29 (3): 184-203.

42- Reckase, M.D.(1999).**The Past And Future Of Multidimensional Item Response Theory.** Applied Psychological Measurement; 21(3):25-36.

43- 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.

44- 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 Simulations 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.

45- Robert, J and Bock, d.(2003): **Item Analysis And Test Scoring With Binary Logistic Models**, Scientific Software Inc USA.

46- Robert,B.C.(2010). **Comparisons Between Item Response Theory Models** , counseling psychologist,college park,vol.21,no.3,pp234-260.

47- 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.

48- 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.

49- 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.

50- 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)

51- 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..

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.

53- 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.

54- 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.

55- 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.

56- 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.

57- 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.

58- Zimowski, M., Muraki, E., Mislevy, R. J., & Bock, R. D. (2003). BILOG-MG 3: **Item Analysis And Test Scoring With Binary Logistic Models**. Chicago: Scientific Software.

Analysis Influence Of Logarithmic Models (One And Two And Three) Parameter On Estimates Of Item Test Information Functions And Their Relationship To Accurately Estimate The Ability

(Comparative study)

D. Mohamed Mansour Mohamed El Shafei

Assistant Research Professor, Department of Research at the National Center For Examinations and Educational Evaluation

Abstract

Study aimed to compare estimates of, item information functions of progressive matrices test (black and white picture) "by John Raven," It (the maximum value of information , points of ability corresponding to the maximum values for the items information - the average values of information) resulting from the analysis of three models linked to the item response model (pL1 – pL2 – pL3), and the relationship between estimates of, item information functions and the accurately estimate the ability resulting from the analysis using the three models are the subject of comparison, The study used in the analysis of their sample of students from five secondary schools at south Cairo learning region. (858) students from all grades (one to three grade), The study used the method analyzing the measurement repeated in statistical process for testing hypotheses of the study, The study found that the form of two – parameter logistic model is the best than other models in introducing the highest estimates of the values of functions, information of test items followed by form three – parameter logistic model then one – parameter logistic model, the same results reached with respect to two variables: standard error to estimate the ability and reliability of estimating functions of information and therefore as indicators of the accuracy of estimating thc ability of examiners, it was found that the form of bilateral Parameter led

his analysis to the lowest estimate of the errors, the standard estimates of ability and the estimates for the stability of the values of information items were higher than those estimated including its Form of others, The study also found a link between the accuracy of estimates of ability and information functions, items test, where it became clear that with increasing values of (average) information items test there was an increase in the accuracy of estimates of ability resulting from the analysis using the item response model.