مقاومة العلامة الحقيقية لاختلاف عدد أبعاد الاختبار وقوة الارتباط بينها واختلاف توزيع قدرة المفحوصين: دراسة محاكاة

يوسف عبدالقادر ابوشندي *

ملخص

هدفت هذه الدراسة إلى تفحص أثر عدد أبعاد الاختبار على تقدير العلامة الحقيقية Domain Score، كما هدفت إلى تفحص أثر قوة الارتباط بين أبعاد الاختبار (0.2, 0.9)، وتوزيع قدرة المفحوصين(طبيعي Normal، متسق Uniform) على تقدير العلامة الحقيقية. ولتحقيق غرض الدراسة تم توليد استجابات لأربع مجموعات من المفحوصين (1000 مفحوص لكل مجموعة) على سبعة اختبارات (ثلاثة منها أحادية البعد، وثلاثة منها ثنائية البعد، وواحد ثلاثي البعد).

كشفت نتائج الدراسة عن اختلاف تقديرات العلامة الحقيقية باختلاف كل من عدد أبعاد الاختبار(بعد واحد، بعدان، ثلاثة أبعاد)، واختلاف توزيع معلمة قدرة المفحوصين (طبيعي، متّسق). وعدم تأثرها باختلاف قوة الارتباط بين الأبعاد(0.2, 0.9).

الكلمات الدالة: العلامة الحقيقية، نظرية استجابة الفقرة،،أحادية البعد، تعددية البعد، توزيع معلمة القدرة، الارتباط بين الأبعاد.

المقدمة

تعد الاختبارات من أكثر أدوات القياس شيوعاً، وتعتمد كثير من القرارات على نتائج هذه الاختبارات مثل: القرارات المتعلقة بالتعيين، أو الترفيع، أو التصنيف، أو تعديل الخطط الدراسية. ويلجأ بعض الباحثين للاختبارات كأدوات جمع بيانات لدراساتهم. ويستخدمها الكثير من المدرسين كوسيلة إثارة دافعية لطلبتهم؛ فدافعية الطلبة ترتفع عند حصولهم على علامة مرتفعة في الاختبارات التحصيلية.

وتقيس بعض الاختبارات قدرة واحدة، وأخرى قدرات متعددة، وقد اهتمت نظريتا القياس (التقليدية والحديثة) بعدد أبعاد الاختبار، ففي بعض نماذج نظرية استجابة الفقرة Item Response Theory (IRT) يفترض وجود قدرة واحدة تفسر أداء المفحوص على الاختبار، وتسمى هذه بالنماذج أحادية البعد Unidimensional Models، وبعض النماذج تفترض وجود أكثر من قدرة تفسر أداء المفحوص وتسمى بالنماذج متعددة الأبعاد Multidimensional Models. وأكد البعض من خبراء القياس المعاصر (Mckinley & Way, 1992; Reckase, 1997) على ان عمليات التفكير تنطوي على العديد من المكونات، لذا ينبغي التعامل معها من منظور متعدد الأبعاد، واتجه هؤلاء الخبراء إلى توسيع نماذج نظرية استجابة الفقرة أحادية البعد إلى نماذج متعددة الأبعاد التي من خلالها يمكن وصف التفاعلات بين معالم القدرات للمفحوصين ومعالم الفقرات.

تختلف نماذج نظرية استجابة الفقرة متعددة الأبعاد عن نماذج نظرية استجابة الفقرة أحادية البعد، وخاصة في المتغيرات التي تتضمنها الصيغ الرياضية لهذه النماذج؛ ففي النماذج أحادية البعد يمثل منحنى خصائص الفقرة Item Characteristic Curve (ICC) العلاقة بين احتمال الاستجابة الصحيحة على الفقرة، وبين القدرة التي يقيسها الاختبار في فضاء ثنائي البعد Two Dimensional Space(حيث يمثل محور السينات مستوى قدرة المفحوصين والصادات احتمال الإجابة الصحيحة على الفقرة$(P_i(\theta)$)، أما في النماذج متعددة الأبعاد فتمثل العلاقة بواسطة سطح خصائص الفقرة Item Response Surface (IRS) في فضاء متعدد الأبعاد (حيث يمثل احد المحاور احتمال الإجابة الصحيحة على الفقرة والمحاور الأخرى كل منها يمثل قدرة من قدرات الاختبار). ويشير تمييز الفقرة متعددة الأبعاد إلى نقطة تقع على سطح خصائص الفقرة، يكون عندها الميل أكبر ما يمكن، ولان التعامل مع سطح وليس منحنى اقترح ريكيس وماكنلي

* كلية العلوم التربوية، جامعة الزرقاء. تاريخ استلام البحث 2010/11/30، وتاريخ قبوله 2011/10/5.

(Reckase& Mckinley, 1991) معادلة لتقدير تمييز الفقرة متعددة الأبعاد:

$$MDISC_i = (\Sigma a^2_{ik})^{1/2} \quad (1)$$

حيث تمثل MDISCi (Multidimensional Discrimination) التمييز متعدد الأبعاد للفقرة i، a^2_{ik} مربع معامل تمييز الفقرة i على البعدk . وتشتمل صعوبة الفقرة متعددة الأبعاد على قيمتين كما بيَن ريكيس (Reckase, 1985) إحداهما تمثل قيمة المسافة من نقطة الأصل الى النقطة التي تحدد تمييز الفقرة، والأخرى تمثل جتا الزاوية المحصورة بين خط المسافة ومحور القدرة.

ومَيز آدمز وويلسون وولنج (Adams, Wilson & Wang, 1997) بين نوعين من الاختبارات متعددة الأبعاد هما:

- **الاختبار متعدد الأبعاد بين الفقرات** (Between-Item Multidimensional Test): حيث يكون الاختبار متعدد الأبعاد بين الفقرات إذا احتوى على اختبارات جزئية أحادية البعد، وفي هذا الاختبار تنتمي كل فقرة من فقراته لاختبار جزئي واحد، ولا تتقاطع مع أي اختبار جزئي آخر، وتمتاز الاختبارات الجزئية بأنها مستقلة و تقيس خصائص مترابطة ومن الأمثلة على هذا النوع اختبار في مادة الرياضيات يحتوي على بعد الهندسة وبعد الجبر وبعد القياس.
- **الاختبار متعدد الأبعاد داخل الفقرات** (Within- Item Multidimensional Test)**:** حيث يحتوي هذا النوع من الاختبارات على فقرات تقيس أكثر من بعد، يتطلب هذا النوع من الفقرات أكثر من قدرة من المفحوص حتى يستجيب عليها بشكل صحيح.

وقد أشارت كثير من الأدبيات (Ackeman, 1994; Embretson & Reise, 2000; Knol & Berger, 1991; Levy, 2006) إلى أن النماذج متعددة الأبعاد يمكن أيضا التمييز بينها من حيث كونها نماذج تعويضية Compensatory Models ونماذج غير تعويضية Non Compensatory Models. فقد بين اكرمان (Ackerman, 1994) أنه في النماذج التعويضية يستطيع الأفراد الضعاف في قدرة معينة أو أكثر تعويض ضعفهم؛ وبعبارة أكثر تحديداً تسمح هذه النماذج بتفاعل بين القدرات متعددة الأبعاد حيث أن النقص في إحداها لدى المفحوص يمكن أن يعوض بقدراته الأخرى المرتفعة. ومن الأمثلة على النماذج التعويضية: النموذج اللوجستي ثنائي المعلمة للفقرات الثنائية الذي طوره ريكيس(Reckase, 1985) وصيغته الرياضية:

$$P(X_{ij}=1|\theta_j, d_{i)} = (e^{aik\,\theta jk + di}) / (1+ e^{aik\,\theta jk + di}) \quad (2)$$

حيث تمثل $P(X_{ij}=1|\theta_j, d_i$ احتمال استجابة المفحوص j الذي له القدرة θ_j استجابة صحيحة على الفقرة i التي لها معامل صعوبة d_i. $a_{ik} = (a_{i1},a_{i2},\ldots,a_{im})$: متجه من معالم تمييز الفقرة i على كل بعد من الأبعاد. d_i: قيمة غير متجهه لمعلمة صعوبة الفقرة، $\theta_k = (\theta_1, \theta_2,\ldots, \theta_m)$ يمثل متجه من معالم القدرات.

من المعالم الرئيسية التي يتم تقديرها في نماذج نظرية استجابة الفقرة بأنواعها معلمة العلامة الحقيقية، حيث بيَن لورد (Lord, 1980) أنَ تقدير العلامة الحقيقية عنصراً جوهرياً لفهم عملية القياس، ولفهم أثر أخطاء القياس ولخصائص العلامة المشاهدة كقياسات قابلة للخطأ. وقد حاول هامبلتون وسواميناثان (Hambleton & Swaminathan, 1985) أن يوضحا مفهوم العلامة الحقيقية من خلال العلاقة بينها وبين علامة القدرة θ، من خلال العلاقة:

$$(3) = 1/n \sum P_i(\theta)|\zeta\theta$$

حيث تشير $|\zeta\theta$: إلى العلامة الحقيقية، و $P_i(\theta)$: تمثل احتمال إجابة المفحوص على الفقرة i إجابة صحيحة، n: عدد فقرات الاختبار.

وقد شغل تقدير العلامة الحقيقية الكثير من المتخصصين في السنوات الاخيرة، ومنهم بومريتش ونيسوندر (Pommerich& Nicewander, 1998) حيث بيَنا أن نماذج نظرية استجابة الفقرة تزودنا بطريقة سهلة لتقدير العلامة الحقيقية. وبيَن بوك وثيسن وزيموسكي (Bock, Thissen& Zimowski, 1997) عند مقارنتهما لطريقة نظرية استجابة الفقرة مع طرق أخرى تبيَن أن العلامات المقدرة حسب نظرية استجابة الفقرة كانت أسهل وأدق من التقديرات بالطريقة الأخرى التي تقوم على حساب نسبة العلامات الصحيحة حسب النظرية التقليدية Classical Percent-Correct Score.

أهمية الدراسة

حاولت الكثير من الدراسات تفحص أثر عدد أبعاد الاختبار وقوة الارتباط بين هذه الأبعاد على تقديرات معالم الفقرات وقدرة المفحوصين، والبعض منها حاول بالتحديد دراسة أثر عدد أبعاد الاختبار على العلامة الحقيقية (Camilli, Wang& Fesq, 1992; Cook, Dorans, Eignor& Peterson, 1983; De Champlain, 1995; Pommerich& Nicewander, 1998; Wang, Harris & Rosses, 2002) ولم تتناول أي من هذه الدراسات أثر قوة العلاقة الارتباطية بين أبعاد الاختبار على تقدير العلامة الحقيقية، على الرغم من أن العديد من الدراسات (Ackerman,1989; Batley&Boss,1993; Bolt& Lall, 2003; De Ayala, 1993; Torre& Patz, 2002; Wang& Wang, 2003; Way, Ansley& Forsyth,1988) كشفت عن وجود اثر

لقوة العلاقة الارتباطية بين الأبعاد على تقديرات معالم الفقرات (الصعوبة والتمييز) وقدرة المفحوصين. وأغفلت أيضاً الدراسات أثر اختلاف خصائص توزيع معلمة قدرة المفحوصين على تقديرات العلامة الحقيقية، سيما أن بعض الدراسات (Batley&Boss,1993; Kirisci& Hsu, 2001) أثبتت تأثر تقديرات معالم الفقرات وقدرة المفحوصين بخصائص توزيع قدرات المفحوصين. لذا تأتي أهمية هذه الدراسة من أهمية اختيار نموذج نظرية استجابة الفقرة المناسب لتقدير احتمال إجابة المفحوص على فقرات الاختبارات إجابة صحيحة Pi(θ) والذي يدخل في تقدير العلامة الحقيقة، حيث تختلف النماذج باختلاف عدد أبعاد الاختبار. كما تأتي أهمية هذه الدراسة من كونها الدراسة الأولى (حسب علم الباحث) التي تتفحص أثر قوة الارتباط بين الأبعاد على تقدير العلامة الحقيقة. كما وتأتي أهميتها من كونها الدراسة الأولى (حسب علم الباحث) التي تتفحص أثر خصائص توزيع معلمة قدرة المفحوصين θ على تقدير العلامة الحقيقية. ومن المؤمل أن يستفيد من نتائجها الباحثون في نظرية استجابة الفقرة بشكل عام وفي تقدير العلامة الحقيقية بشكل خاص، كما يمكن أن يستفيد منها طلبة الدراسات العليا في تخصص القياس والتقويم المهتمين بتقدير معالم نماذج نظرية استجابة الفقرة.

مشكلة الدراسة وأسئلتها

اهتمت العديد من الدراسات بتفحص أثر عدد أبعاد الاختبار وقوة الارتباط بين هذه الأبعاد على تقديرات نماذج نظرية استجابة الفقرة، ومن هذه الدراسات (McKinley; McKinley & Reckase ; Muraki & Englehard. as cited in Batley &Boss,1993) التي أشارت إلى أن النماذج متعددة الأبعاد أكثر دقة في تقدير وتفسير معالم فقرات الاختبار متعدد الأبعاد من النماذج أحادية البعد. وبعض الدراسات اهتمت بتفحص أثر توزيع معلمة القدرة للمفحوصن على تقديرات نماذج نظرية استجابة الفقرة منها (Batley & Boss, 1993; Kirisci & Hsu, 2001). وقد حاول السيكومتريون كما بين ميرا وروبين وسيرسي (Meara, Robin & Searsi, 2000) منذ سنوات عديدة مضت أن يظهروا أهمية تقييم عدد أبعاد البيانات الاختبارية، وأثر التقييم الخاطئ لعدد الأبعاد على مصداقية نتائج الاختبارات وعلى حساب وتفسير علامات الاختبارات، وتقدير معالم الفقرات والمفحوصين. وبشكلٍ عام توصلت الدراسات أعلاه إلى أنَ عملية إجراء تقديرات أحادية البعد لبيانات متعددة الأبعاد قد عانت من صعوبة في التفسير للنتائج ولم تعكس خصائص حقيقية للبيانات.

حاولت هذه الدراسة تفحص أثر عدد أبعاد الاختبار (ثلاثة أبعاد، بعدان، بعد واحد) وقوة الارتباط بين هذه الأبعاد (0.2, 0.9)، واختلاف توزيع معلمة القدرة للمفحوصين (طبيعي، متسق) على تقديرات العلامة الحقيقية باستخدام بيانات محاكاة Simulated Data. وقد اختار الباحث قيمتي قوة الارتباط بين الأبعاد (0.2, 0.9) لان بعض الدراسات (Torre& Patz, 2002; Ackerman,1989; Way, Ansley& Forsyth,1988; Ansley& Forsyth,1985) بيَنت أنَ تقديرات معالم الفقرات (الصعوبة والتمييز) ومعلمة قدرة المفحوصين (θ) تكون غير دقيقة إذا كانت قوة الارتباط بين الأبعاد0.9، واتفقت أيضاً هذه النتائج مع دراسة ابوشندي (2008). وبيَن الكرمان (Ackerman, 1989) باستخدام التحليل العاملي انه عندما كانت قوة الارتباط بين أبعاد الاختبار (ثلاثة أبعاد) 0.9 تبيَن أن الاختبار أحادي البعد. وبينَت دراسة (Van Abswoude, Van der Ark & Sijtsma, 2004) أن طرائق تحديد عدد أبعاد الاختبار (DETECT, DIMTEST) لا تعطي نتائج دقيقة عندما تكون قوة الارتباط بين الأبعاد 0.2. وبيَنت دراسة ابوشندي (2010) أنَ تقديرات معالم الفقرات(الصعوبة والتمييز) تكون غير دقيقة عند استخدام تقديرات أحادية البعد لاختبارات متعددة الأبعاد عندما تكون قوة الارتباط بين الأبعاد 0.2.

وبشكلٍ أكثر تحديداً حاولت هذه الدراسة الإجابة عن الأسئلة التالية:

1. هل تختلف تقديرات العلامة الحقيقية في نظرية استجابة الفقرة باختلاف عدد أبعاد الاختبار (ثلاثة أبعاد، بعدان، بعد واحد)؟
2. هل تختلف تقديرات العلامة الحقيقية في نظرية استجابة الفقرة باختلاف قوة الارتباط بين الأبعاد (0.2, 0.9)؟
3. هل تختلف تقديرات العلامة الحقيقية في نظرية استجابة الفقرة باختلاف توزيع معلمة القدرة للمفحوصين (طبيعي، متسق)؟

هدف الدراسة

هدفت هذه الدراسة إلى مقارنة تقديرات العلامة الحقيقية للمفحوصين عندما تم توليد استجاباتهم على فقرات شكلت اختبار أحادي البعد، أو ثنائي البعد، أو ثلاثي البعد، وعندما كانت قوة العلاقة الارتباطية بين الأبعاد 0.2 أو 0.9، وعندما توزعت قدرات المفحوصين توزيعاً طبيعياً (بوسط حسابي 0.0 وانحراف معياري 1.0) و توزعت بشكل متَسق (4+ ,4-)، باستخدام بيانات محاكاة.

مصطلحات الدراسة

الاختبار أحادي البعد: اختبار يقيس قدرة واحدة من خلال مجموعة الفقرات المشكّلة لهذا الاختبار.

الاختبار متعدد الابعاد بين الفقرات التعويضي: اختبار يقيس سمتين أو أكثر، ويحتوي على اختبارات جزئية بحيث تنتمي كل فقرة من فقرات هذا الاختبار إلى اختبار جزئي واحد، ولا تتقاطع مع أي اختبار جزئي آخر. وتبيّن الصيغ الرياضية للنماذج التعويضية، انه لكل فقرة معلمة تميز على كل بعد من الأبعاد المكونة لفضاء السمة، ومعلمة صعوبة واحدة.

العلامة الحقيقيةDomain Score :هي المتوسط الحسابي لاحتمالات إجابات المفحوص Pi(θ) عند مستوى القدرة θ على جميع فقرات بعد الاختبار سواء كان البعد يشكل أختبار احادي البعد، أم ثنائي البعد أم ثلاثي البعد.

محددات الدراسة

1. اقتصرت الدراسة على استخدام النماذج التعويضية Compensatory Models، ولم تستخدم النماذج غير التعويضية Non compensatory Models.
2. اقتصرت الدراسة على الاختبار من نوع متعدد الابعاد بين الفقرات، ولم تتناول حالة تعددية البعد داخل الفقرات.
3. اقتصرت الدراسة على اختبار مختلف في عدد الفقرات في كل بعد من أبعاد الاختبار، ولم تتناول تساوي عدد الفقرات في كل بعد، والهدف من ذلك تفحص أثر وجود بعد سائد Dominant Factor بين هذه الأبعاد على تقديرات العلامة الحقيقية.
4. تعتمد النتائج التي تم التوصل إليها على البرامج الحاسوبية المتوافرة في توليد وتحليل البيانات.

الدراسات السابقة

هدفت دراسة كوك ودورانس وايجنور وبترسون(Cook, Dorans, Eignor& Peterson, 1983) تفحص العلاقة بين انتهاك افتراض احادية البعد الذي تم تفحصه باستخدام التحليل العاملي ومعايرة العلامة الحقيقية باستخدام نظرية استجابة الفقرة (المعايرة بالسلسلة). بيانات الدراسة تم الحصول عليها من تطبيق اختبار الاستعداد المدرسي Scholastic Aptitude Test (SAT) ست مرات بجزئيه اللفظي والرياضي. نتائج التحليل العاملي اتفقت مع التحليلات المسبقة لتحديد عدد أبعاد الاختبار (البعد اللفظي والبعد الرياضي). لاحظ كوك ورفاقه ان انتهاك افتراض احادية البعد في الاختبار لوجود بعدين لم يؤثر بشكل واضح على تقديرات العلامة الحقيقية، وعزو ذلك لوجود عامل سائد Dominant Factor بين أبعاد الاختبار.

وفي دراسة كاميلي ووانج وفيسك (Camilli, Wang& Fesq, 1992) التي قاموا من خلالها بتفحص عدد الأبعاد لستة مرات من تطبيق اختبار القبول لكلية الحقوق(LSAT) باستخدام التحليل العاملي، حيث تبين وجود قدرتين هما التفكير الاستقرائي والتفكير الاستنتاجي. أجريت معايرات للعلامة الحقيقية مرة عندما شكلت جميع الفقرات اختبار واحد ومرة عندما شكلت اختبار ثنائي البعد. بيَنت نتائج الدراسة عدم اختلاف في تقدير العلامة الحقيقية في الحالتين.

وفي دراسة دي شامبلين (De champlain, 1995) التي قيَم من خلالها اثر عدد أبعاد الاختبار على تقدير العلامة الحقيقية في نظرية استجابة الفقرة لبيانات تم الحصول عليها من تطبيق اختبار القبول في كلية الحقوق Law School Admission Test (LSAT) على مجموعات مختلفة من المفحوصين (الأمريكان من أصل أفريقي، والأمريكان من أصل اسباني، والبيض). بيَنت دراسات تقييم عدد ابعاد الاختبار انه ثنائي البعد (التفكير التحليلي والمنطقي، والاستيعاب القرائي). لتحقيق غرض الدراسة قورنت التقديرات من نتائج المجموعات كل على حدة مع تقديرات النتائج عند التطبيق الكلي، توصلت النتائج أن الفروق في شكل المنحنيات التي تمثل العلاقة بين قدرة المفحوص والعلامة الحقيقية لم تكن ظاهرة بوضوح، والفروق الأكبر ظهرت عند أطراف المنحنيات؛ أي عند القدرات المنخفضة والقدرات المرتفعة.

وفي دراسة بومريتش ونايسواندر (Pommerich& Nicewander, 1998) التي هدفت الى تفحص أثر كل من: عدد فقرات الاختبار(5, 10, 20)، وحجم مجال الفقرات الكلي (30, 60, 120)، وحجم عينة المفحوصين (25, 50,100) على دقة تقدير العلامة الحقيقية باستخدام ستة طرق مختلفة حسب النظرية الحديثة في القياس (EM, EAP1, EAP2, MLE, MU, OBS) والطريقة التقليدية (Percent-Correct Score). قام الباحثان بتوليد تسعة اختبارات(تختلف فيما بينها في عدد فقرات الاختبار وحجم مجال الفقرات الكلي) وتطبيقها على عينات المفحوصين. بيَنت الدراسة ان طريقتا (Percent-Correct Score, MLE) قد تأثرتا بعدد فقرات الاختبار(5, 10, 20) و حجم عينة المفحوصين (25, 50,100). وبقية الطرق (EM, EAP1, EAP2, MU, OBS) لم تتأثر.

وفي دراسة وانج وهاريس وروسس (Wang, Harris & Rosses, 2002) كان غرضها تفحص أثر عدد الأبعاد للاختبار(1, 2) على منحنى خصائص الفقرة ICC وعلى تقدير العلامة الحقيقية. ولهذا الغرض استخدم الباحثون بيانات تم الحصول عليها من عدة تطبيقات لاختبار القبول لكلية الحقوق Low School Admission Test (LSAT)، الذي يتكون من أربعة أجزاء(التفكير التحليلي Analytical Reasoning (AR)، والتفكير المنطقي Logical Reasoning A(LR:A)، والتفكير المنطقي

Logical Reasoning B (LR:B)، والاستيعاب القرائي Reading Comprehension)، وبعد إجراء التحليل العاملي تم تصنيف هذه الأجزاء في بعدين(التفكير الإستنتاجي Deductive Reasoning وقد ضم الجزء الأول (التفكير التحليلي) والتفكير العام Informal Reasoning وقد ضم الأجزاء الثلاث المتبقية (التفكير المنطقي A,b، الاستيعاب القرائي). وأشارت النتائج أن تعددية البعد تعطي تقديرات متدنية لميل منحنى خصائص الفقرة (تمييز الفقرة)، وتقلل من قيمة العلامة الحقيقية عند ذوي القدرات المرتفعة (حيث قسمت قدرات المفحوصين على نتائج اختبار LSAT إلى ثلاثة مستويات: مرتفعة، متوسطة، منخفضة)، وأظهرت النتائج أن تعددية البعد كان لها أثر واضح على دقة تقديرات المعالم عندما كان مدى القدرة واسع.

وهدفت دراسة ذيابات (2007) إلى فحص أثر التغير في طول الاختبار (عدد الفقرات) على خصائص توزيع العلامة الحقيقية، من خلال تطبيق اختبار اوتيس- لينون للقدرة العقلية المكيفة للبيئة الأردنية على عينة حجمها 2100 طالب وطالبة من الصفوف الأساسية (الرابع، الخامس، السادس) في مديرية تربية الرمثا. كشفت نتائج الدراسة عن تأثر خصائص توزيع العلامة الحقيقية بتغير طول الاختبار، حيث يزداد تباين توزيع العلامة الحقيقية بازدياد طول الاختبار. وبينت الدراسة أنَ خصائص توزيع العلامة الحقيقية للاختبارات الفرعية تقترب من خصائص توزيعها للاختبار الكلي كلما زاد طول الاختبار.

الطريقة والإجراءات

البيانات المحاكاة

استخدم الباحث في هذه الدراسة البيانات المحاكاة التي حصل عليها من دراسة سابقة له ابوشندي(2010) حيث استخدم برمجية RESGEN4 حسب الإجراءات التالية:

المعالم الحقيقية

تتطلب برامج توليد البيانات تزويدها بالقيم الحقيقية لمعالم الفقرات(معلمة الصعوبة bi، معلمة التمييز a_i، معلمة التخمين c_i) والمفحوصين(معلمة القدرة θ). أما توزيع معلمة القدرة فقد درس الباحث توزيعين مختلفين لمعلمة القدرة: التوزيع الطبيعي المعياري Standard Normal Distribution بوسط حسابي صفر وانحراف معياري1، والتوزيع المتسق Uniform بقيمة ابتدائية 4- وقيمة نهائية 4+.

توليد البيانات

بعدما اعتمدت القيم الحقيقية لمعالم الفقرات، وتوزيع معلمة قدرة المفحوصين تطلب توليد البيانات الخطوات التالية:

1. إدخال المعالم الحقيقية للفقرات الثلاثين الأولى المشكلة للبعد الأول من الاختبار في المجموعة الأولى Block1، حيث لكل فقرة تم إدخال قيمة واحدة لمعلمة صعوبة (لان النموذج المستخدم في توليد البيانات هو النموذج التعويضي Compensatory Models)، وقيمة واحدة لمعلمة التمييز على البعد الأول وصفر لمعلمة التمييز على البعدين الثاني والثالث(لان تعددية البعد بين الفقرات وليس داخل الفقرات). أما البعد الثاني من الاختبار والمتشكل من 20 فقرة فقد تم إدخاله في المجموعة الثانية Block2، وقد تم إدخال قيمة واحدة لمعلمة صعوبة الفقرة، وقيمة واحدة لمعلمة التمييز على البعد الثاني وصفر على البعدين الأول والثالث. والبعد الثالث المتشكل من 10 فقرات تم إدخاله في المجموعة الثالثة Block 3، وتم إدخال قيمة واحدة لمعلمة الصعوبة وقيمة واحدة لمعلمة التمييز على البعد الثالث وصفر على البعدين الأول والثاني.

2. افترض أبوشندي (2010) وجود أربع مجموعات من المفحوصين (تتكون كل منها من 1000 مفحوص):

أ- المجموعة الأولى Group1 (1000 مفحوص): تم الافتراض بان قدرات هذه المجموعة من المفحوصين تتوزع بشكل طبيعي بوسط حسابي صفر وانحراف معياري واحد، وان قوة العلاقة الارتباطية بين الأبعاد الثلاثة هي 0.2 حسب المصفوفة (1, 0.2, 1,0.2,0.2,1).

ب- المجموعة الثانية Group2(1000 مفحوص): تم الافتراض بان قدرات هذه المجموعة من المفحوصين تتوزع بشكل طبيعي بوسط حسابي صفر وانحراف معياري واحد، وان قوة العلاقة الارتباطية بين الأبعاد الثلاثة هي 0.9 حسب المصفوفة (1, 0.9, 1, 0.9, 0.9, 1).

ج- المجموعة الثالثة Group3(1000 مفحوص): تم الافتراض بان قدرات هذه المجموعة من المفحوصين تتوزع بشكل متسق [4+ ,4 -]، وان قوة العلاقة الارتباطية بين الأبعاد الثلاثة هي 0.2 حسب المصفوفة (1, 0.2, 1 ,0.2 ,0.2, 1).

د- المجموعة الرابعة Group 4(1000 مفحوص): تم الافتراض بان قدرات هذه المجموعة من المفحوصين تتوزع بشكل متسق [4+ ,4 -]، وان قوة العلاقة الارتباطية بين الأبعاد الثلاثة هي 0.9 حسب المصفوفة (1, 0.9, 1, 0.9, 0.9, 1).

3. افترض الباحث أن مجموعات المفحوصين الأربع تقدموا لسبعة اختبارات تختلف فيما بينها في البعدية (عدد الأبعاد لكل منها: بعد واحد، وبعدان، وثلاثة أبعاد) وعدد

الفقرات لكل بعد (30,20,10) كالتالي:

اولا: الاختبارات احادية البعد: حيث تشكل الاختبار الأول من 30 فقرة، والثاني من 20 فقرة، والثالث من 10 فقرات.

ثانياً: الاختبارات ثنائية البعد:

- الاختبار الأول: يحتوي البعد الأول على 30 فقرة والثاني 20 فقرة.
- الاختبار الثاني: يحتوي البعد الأول على 30 فقرة والثاني 10 فقرات.
- الاختبار الثالث: يحتوي البعد الأول على 20 فقرة والثاني 10 فقرات.

ثالثاً: الاختبار ثلاثي البعد: يحتوي البعد الأول عل 30 فقرة والثاني على 20 فقرة والثالث على 10 فقرات.

وبعد الحصول على بيانات محاكاة تتمثل في استجابات أربع مجموعات من المفحوصين على سبعة اختبارات، أدخلت هذه الاستجابات الى برمجية NOHARM بعد تكييفها لتوائم البرمجية وبغرض تقدير معالم الفقرات وذلك للاختبارات احادية البعد وثنائية البعد وثلاثية الأبعاد.

إجراءات حساب العلامة الحقيقية

- لان مخرجات برمجية NOHARM لا تحتوي على تقديرات لمعلمة قدرة المفحوصين، أدخلت معالم الفقرات المقدرة من برمجية NOHARM إلى برمجية RESGEN4 لتقدير معلمة قدرة للمفحوصين.
- من تقديرات معالم الفقرات والمفحوصين تم تقدير احتمالات الإجابة الصحيحة Pi(θ) لجميع فقرات الاختبار مقابل كل مستوى من مستويات القدرة، وذلك باستخدام Microsoft Excel حسب المعادلة (2)
- بغرض تقدير العلامة الحقيقية تم حساب الوسط الحسابي لتقديرات الاحتمالات مقابل كل مستوى من مستويات القدرة حسب المعادلة (3)

نتائج الدراسة ومناقشتها

النتائج المتعلقة بالسؤال الأول (هل تختلف تقديرات العلامة الحقيقية باختلاف عدد أبعاد الاختبار (ثلاثة أبعاد، بعدان، بعد واحد)؟)

للاجابة عن هذا السؤال تم تقدير العلامة الحقيقية عندما شكلت الفقرات اختبار ثلاثي البعد (30, 20, 10) او اختبارات ثنائية البعد ((30, 20)، (30, 10)، (20, 10)) أو اختبارات احادية البعد ((30)، (20)، (10)). ومن ثم حسبت الإحصائيات الوصفية (الوسط الحسابي، الانحراف المعياري). ويظهر جدول1هذه الإحصائيات

يلاحظ من الجدول1 عدم وجود اختلافات واضحة في خصائص توزيع تقديرات العلامة الحقيقية باختلاف عدد أبعاد الاختبار(بعد واحد، بعدان، ثلاثة أبعاد)، حيث كانت الفروق بين المتوسطات الحسابية غير ظاهرة. وهذا قد يعود إلى العلاقة الايجابية التي وضحها هامبلتون وسواميناثان (Hambleton & Swaminathan, 1985) بين توزيع العلامة الحقيقية والقدرة.

لذلك لجأ الباحث إلى معاملات ارتباط سبيرمان الرتبية لمحاولة دراسة التغير في العلامة الحقيقية عند اختلاف عدد ابعاد الاختبار، ويظهر جدول 2 معاملات ارتباط سبيرمان بين تقديرات العلامة الحقيقية على افتراض أن الاختبار احادي البعد والتقديرات على افتراض أن الاختبار ثنائي اوثلاثي البعد.

الجدول 1

الإحصائيات الوصفية لتقديرات العلامة الحقيقية عند اختلاف عدد ابعاد الاختبار

عدد أبعاد الاختبار	عدد الفقرات المشكلة للبعد	الوسط الحسابي	الانحراف المعياري
بعد واحد	30	0.493	0.244
	20	0.515	0.252
	10	0.510	0.251
بعدان	30	0.483	0.305
	20	0.525	0.318
	10	0.518	0.308
ثلاثة أبعاد	30	0.492	0.305
	20	0.512	0.323
	10	0.520	0.312

الجدول2

معاملات ارتباط سبيرمان الرتبية بين تقديرات العلامة الحقيقية بين التقديرات أحادية البعد وثنائية وثلاثية البعد

ابعاد الاختبار(عدد الفقرات)	توزيع معلمة القدرة	قوة الارتباط بين الابعاد	احادي وثنائي	احادي وثلاثي
الاول (30)	طبيعي	0.2	-0.015	0.068
		0.9	-0.015	0.068
	متَسق	0.2	0.012	0.060
		0.9	0.012	0.060
الثاني (20)	طبيعي	0.2	-0.015	0.010
		0.9	-0.015	0.057
	متَسق	0.2	0.002	0.012
		0.9	0.002	0.012
الثالث (10)	طبيعي	0.2	-0.090	-0.062
		0.9	-0.039	0.016
	متَسق	0.2	-0.018	-0.013
		0.9	-0.018	-0.013

يظهر من جدول 2 بشكل عام انخفاض معامل الارتباط بين التقديرات، وتراوح بين 0.09- و0.068. ومعامل الارتباط بين التقديرات أحادية البعد وثلاثية البعد بشكل عام أعلى من احادية البعد وثنائية البعد.

وبغرض الوصول إلى صورة واضحة حول تفحص أثر عدد أبعاد الاختبار على تقديرات العلامة الحقيقية، لجأ الباحث إلى الرسم البياني الموضح في ملحق 1 الذي يبين منحنى للعلاقة بين تقديرات علامات قدرة المفحوصين على محور السينات وتقديرات العلامة الحقيقية على محور الصادات. تبين بشكل عام تطابق منحنيات العلامة الحقيقية للتقديرات ثنائية البعد وثلاثية البعد وللأبعاد الثلاثة (30،20،10) والذي يظهر في الرسم البياني في الملحق 1 وعند الحالات المختلفة من توزيع قدرة المفحوصين وقوة الارتباط بين الأبعاد باستثناء حالة التوزيع الطبيعي وقوة الارتباط بين الأبعاد 0.2.

ولوحظ بشكل عام أن قيمة التقديرات عندما كان تشكّل الاختبار من بعدين أو ثلاثة أبعاد للعلامات الحقيقية كانت اقل من التقديرات عندما كان الاختبار أحادي البعد عند القدرة المتدنية (θ >0.0) وأعلى منها عند القدرات المرتفعة (θ <0.0). وذلك عند اختلاف قوة الارتباط بين الأبعاد وتوزيع قدرة المفحوصين.

وقد تفسَر هذه النتائج من خلال اعتماد تقدير العلامة الحقيقية على احتمال إجابة المفحوصين على فقرات الاختبارات إجابة صحيحة $Pi(\theta)$ والذي يعتمد على كل من خصائص الفقرات والمفحوصين، وقد دلت بعض الدراسات كدراسة ابوشندي (2008) وحجازين (2007) على تأثر معالم الفقرات وخاصة معلمة التمييز باختلاف عدد الأبعاد للاختبار، والذي بدوره سيؤثر على احتمال الإجابة الصحيحة عن الفقرات، وبالتالي على العلامة الحقيقية.

وقد اتفقت نتائج هذه الدراسة مع نتائج دراسة وانج وهاريس وروسس (Wang, Harris & Rosses, 2002) التي أظهرت تأثر العلامة الحقيقية بعدد أبعاد الاختبار وخاصة عند ذوي القدرات العالية، واختلفت مع دراسة دي شامبلين (De champlain, 1995) والتي بينت أن الفروق في المنحنيات التي تمثل العلاقة بين قدرة المفحوصين والعلامة الحقيقية لم تكن ظاهرة، واختلفت مع الدراسة نفسها في أن الفروق بين المنحنيات كانت اكبر عند أطراف المنحنيات، أما في هذه الدراسة فكانت المنحنيات اقرب عند الأطراف. واختلفت أيضا هذه الدراسة مع دراسات كاميلي ووانج وفيسك (Camilli, Wang& Fesq, 1992) و كوك ودورانس وايجنور وبترسون(Cook, Dorans, Eignor& Peterson, 1983).

النتائج المتعلقة بالسؤال الثاني (هل تختلف تقديرات العلامة الحقيقية باختلاف قوة الارتباط بين أبعاد الاختبار (0.2, 0.9)؟

للإجابة عن هذا السؤال حسبت الإحصائيات الوصفية (الوسط الحسابي، الانحراف المعياري) لتقديرات العلامة الحقيقية عند اختلاف قوة الارتباط بين الأبعاد، ويبيَن جدول 3 هذه الإحصائيات

يظهر من النتائج في جدول3 عدم الاختلاف الواضح في تقديرات الوسط الحسابي للعلامات الحقيقية (مساوية تقريباً 0.5) عند اختلاف قوة الارتباط بين الأبعاد وكذلك الانحراف المعياري. ومن ثم حسبت معاملات ارتباط سبيرمان الرتبية بين التقديرات المختلفة للعلامات الحقيقية باختلاف قوة الارتباط بين الابعاد(0.2, 0.9) ويظهر جدول4 هذه القيم.

الجدول3

الإحصائيات الوصفية لتقديرات العلامة الحقيقية باختلاف قوة الارتباط بين الأبعاد

عدد الفقرات في كل بعد	قوة الارتباط بين الابعاد	الوسط الحسابي	الانحراف المعياري
30	0.2	0.489	0.286
	0.9	0.490	0.286
20	0.2	0.517	0.301
	0.9	0.518	0.299
10	0.2	0.516	0.292
	0.9	0.517	0.291

الجدول4

معاملات ارتباط سبيرمان الرتبية بين تقديرات العلامة الحقيقية باختلاف معامل الارتباط بين الابعاد

		عدد الفقرات في كل بعد		
توزيع القدرة	عدد أبعاد الاختبار	30	20	10
طبيعي	بعد واحد	1.00	1.00	1.00
	بعدان	1.00	1.00	0.85
	ثلاثة أبعاد	1.00	1.00	0.90
متَسق	بعد واحد	1.00	1.00	1.00
	بعدان	1.00	1.00	1.00
	ثلاثة أبعاد	1.00	1.00	1.00

تبيَن من جدول 4 ارتفاع قيمة معامل الارتباط بين تقديرات العلامة الحقيقية عند اختلاف قوة الارتباط بين الابعاد وعند الحالات المختلفة من عدد ابعاد الاختبار (بعد واحد، بعدان، ثلاثة ابعاد)، وعند اختلاف توزيع معلمة قدرة المفحوصين (طبيعي، متَسق).

ويبَين التمثيل البياني أيضا عدم اختلاف المنحنيات الممثلة لتقديرات العلامة الحقيقية (الملحق 2)، وانطبقت المنحنيات الممثلة لتقديرات العلامة الحقيقية عندما كانت قوة الارتباط بين الابعاد 0.2 مع المنحنيات عندما كانت قوة الارتباط بين الابعاد 0.9 وللحالات المختلفة من عدد ابعاد الاختبار وتوزيع معلمة قدرة المفحوصين.

النتائج المتعلقة بالسؤال الثالث (هل تختلف تقديرات العلامة الحقيقية باختلاف توزيع معلمة القدرة للمفحوصين (طبيعي، متَسق)؟

وللاجابة عن هذا السؤال حسبت الإحصائيات الوصفية (الوسط الحسابي، الانحراف المعياري) لتقديرات العلامة الحقيقية عند اختلاف توزيع قدرة المفحوصين (طبيعي، متَسق)، ويبيَن جدول 5 هذه الإحصائيات.

ظهر من جدول 5 تقارب قيم الوسط الحسابي عند التقديرات المختلفة للعلامات الحقيقية (تقريباً يساوي 0.5) واختلاف ملحوظ في الانحرافات المعيارية عند اختلاف توزيع قدرة المفحوصين، حيث كان قريباً 0.14 عند التوزيع

الطبيعي، و0.39 عند التوزيع المتَسق. ومن ثم حسبت معاملات ارتباط سبيرمان الرتبية بين التقديرات المختلفة للعلامات الحقيقية باختلاف توزيع معلمة المفحوصين (طبيعي، متَسق) ويظهر جدول 6 هذه القيم.

الجدول 5

الإحصائيات الوصفية لتقديرات العلامة الحقيقية باختلاف توزيع معلمة قدرة المفحوصين

عدد الفقرات في كل بعد	توزيع قدرة المفحوصين	الوسط الحسابي	الانحراف المعياري
30	طبيعي	0.489	0.136
	متَسق	0.489	0.381
20	طبيعي	0.523	0.153
	متَسق	0.512	0.395
10	طبيعي	0.521	0.140
	متَسق	0.512	0.388

الجدول 6

معاملات ارتباط سبيرمان الرتبية بين تقديرات العلامة الحقيقية باختلاف توزيع قدرة المفحوصين

قوة الارتباط بين الابعاد	عدد ابعاد الاختبار	30 فقرة	20 فقرة	10 فقرات
0.2	بعد واحد	-0.017	-0.017	-0.017
	بعدان	-0.018	-0.018	-0.011
	ثلاثة ابعاد	-0.050	-0.026	-0.013
0.9	بعد واحد	-0.017	-0.017	-0.017
	بعدان	-0.018	-0.018	-0.001
	ثلاثة ابعاد	-0.050	-0.023	-0.025

يبيَن جدول6 انخفاض قيم معاملات الارتباط بين تقديرات العلامة الحقيقية عند اختلاف توزيع معلمة قدرة المفحوصين، حيث انحصرت القيم بين 0.050- و 0.001-. وهذا يشير الى اختلاف تقديرات العلامة الحقيقية عند اختلاف توزيع معلمة قدرة المفحوصين.

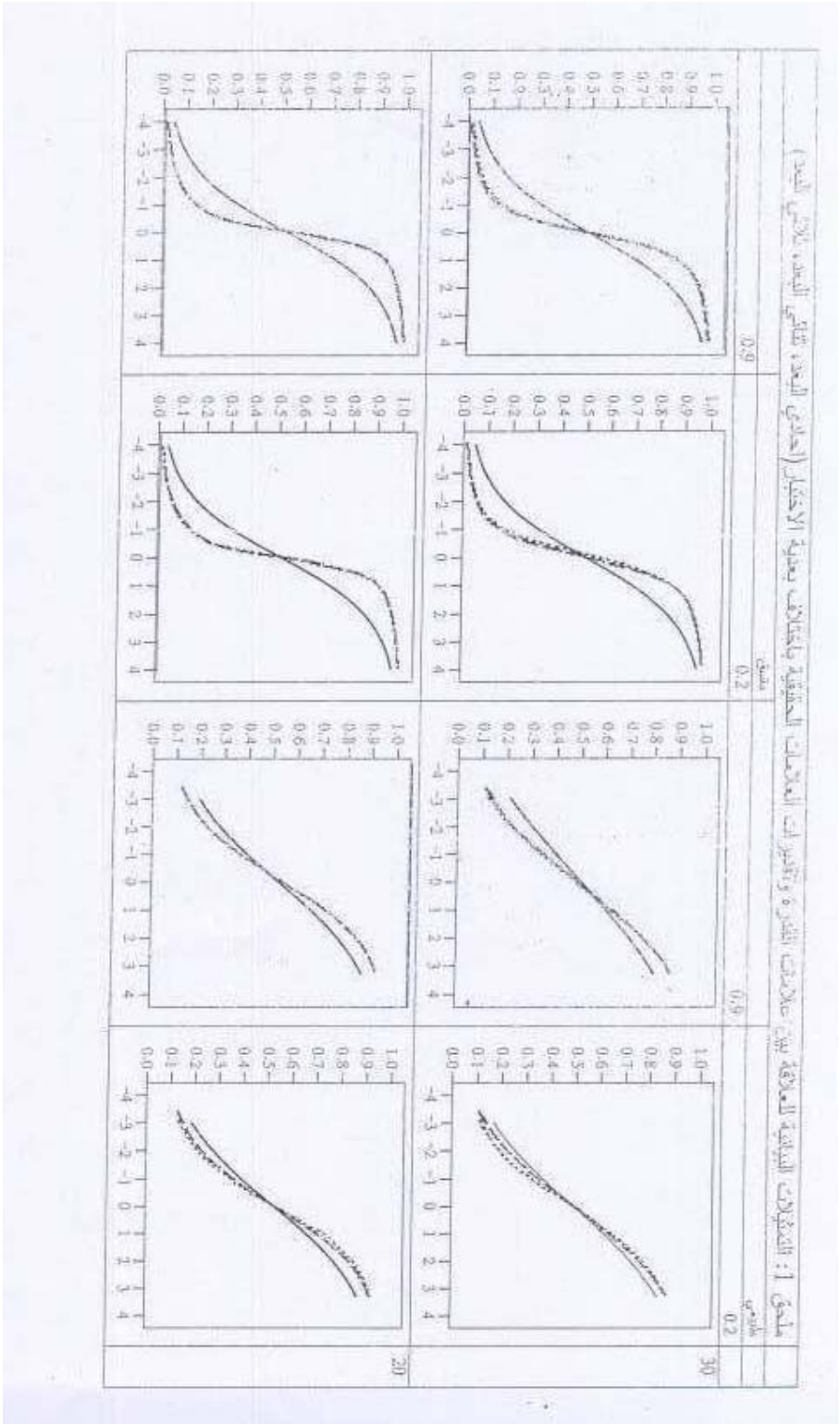
ومن خلال التمثيلات البيانية (ملحق3) تبيَن اختلاف المنحنى الممثل للعلامات الحقيقية عندما توزعت معلمة قدرة المفحوصين بشكل طبيعي عن المنحنى عندما توزعت معلمة قدرة المفحوصين بشكل متَسق، وبشكل عام كانت المنحنيات على اعتبار ان توزيع قدرة المفحوصين متَسق اقل منها على اعتبار ان توزيع معلمة قدرة المفحوصين طبيعي عند القدرات المتدنية (θ< 0.0)، وأعلى عند القدرات المرتفعة (θ> 0.0). وأظهرت التمثيلات البيانية بشكل عام أن مدى توزيع العلامة الحقيقية عندما كان توزيع علامة القدرة متسق اكبر من مدى التوزيع عندما كان توزيع علامات القدرة طبيعي.

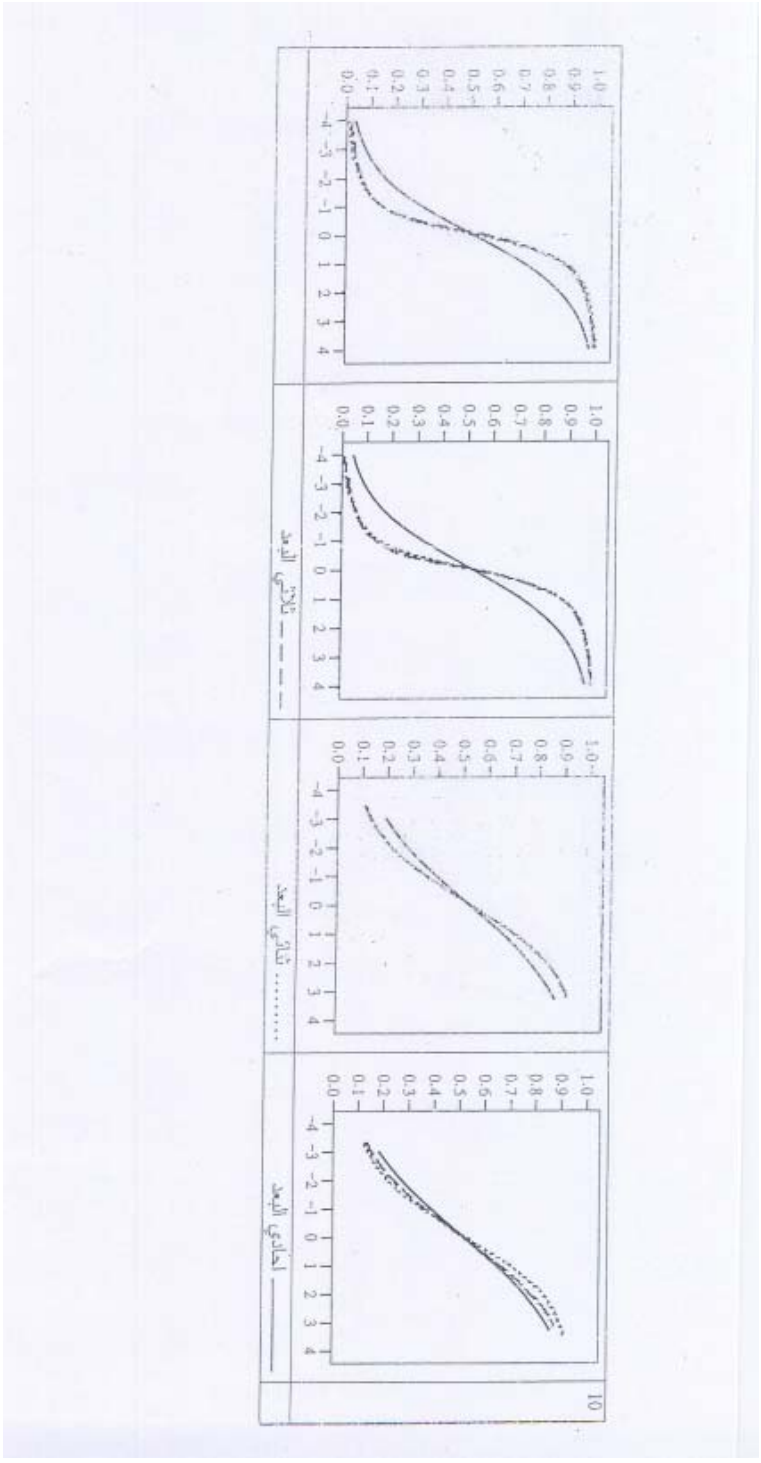
وهذه النتائج يمكن تفسيرها من خلال ما أشار إليه كل من هاملتون وسواميناثان (Hambleton & Swaminathan, 1985) بان توزيع العلامة الحقيقية يكون في مدى واسع عندما تكون الفقرات صعبة (نقطة الانعكاس على منحنى خصائص الفقرة) وعالية التمييز (النقطة التي يكون عندها المنحنى أكثر انحداراً او الميل اكبر ما يمكن) وعند وسط تدريج علامات القدرة. وتكون العلامة الحقيقية ضمن مدى ضيَق اذا كانت علامات القدرة عالية أو متدنية والفقرات صعبة وعالية التمييز.

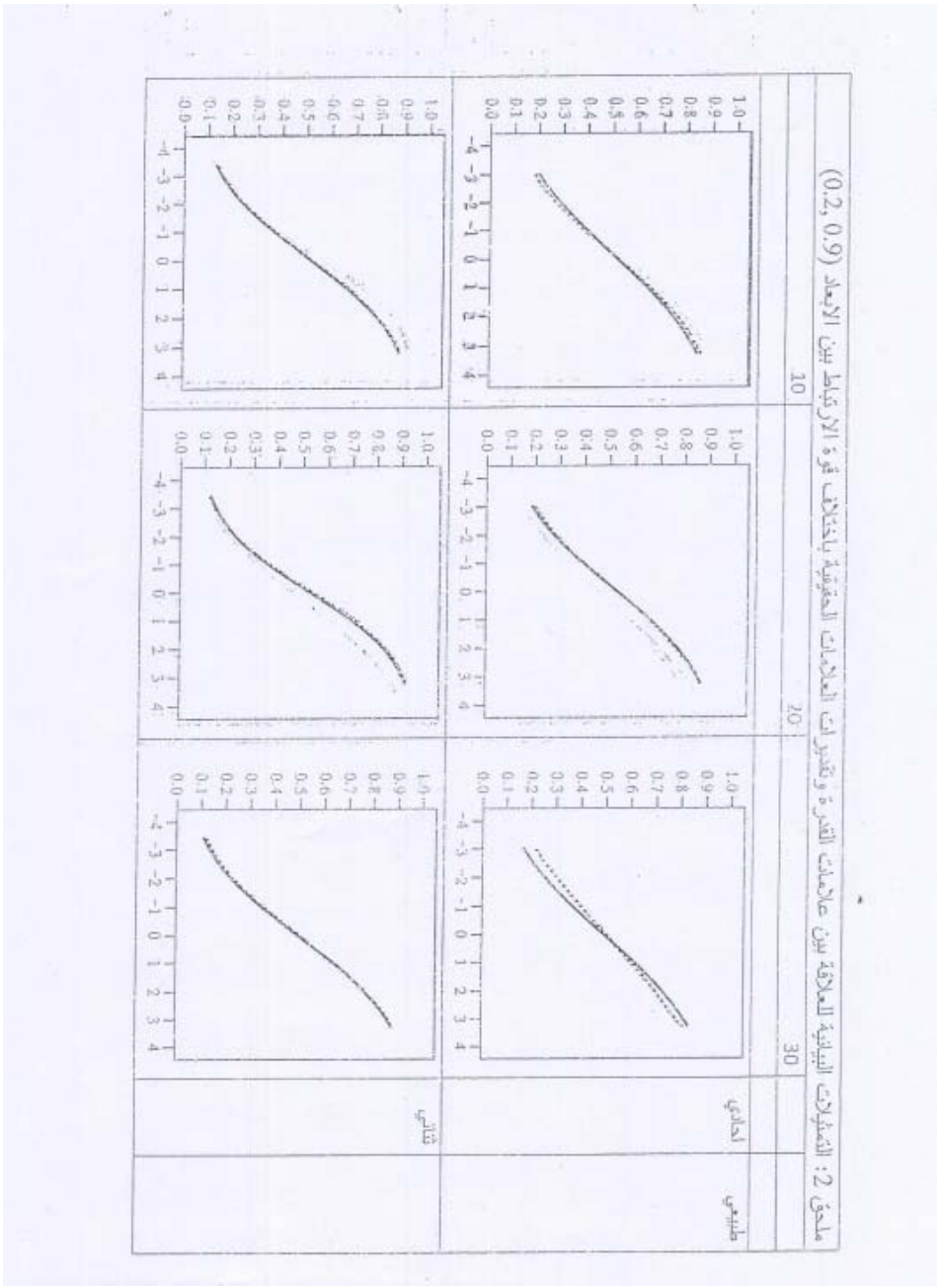
التوصيات

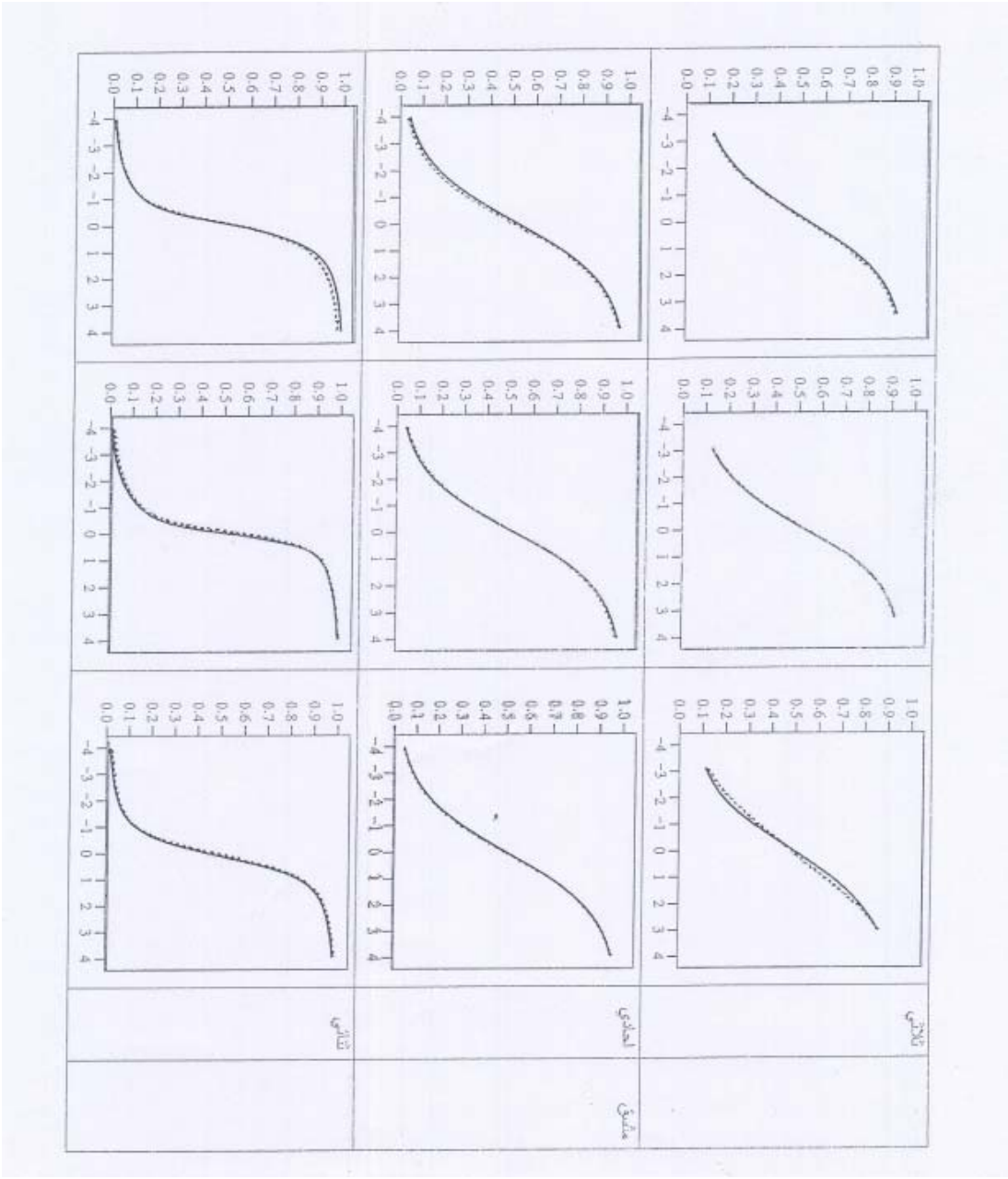
- لأن تقديرات العلامة الحقيقية اختلفت باختلاف عدد ابعاد الاختبار عندما كانت قوة الارتباط بين الابعاد 0.2، وتوزعت قدرات المفحوصين بشكل طبيعي يوصي الباحث بضرورة تحديد البعدية الصحيحة لبيانات الاختبار قبل تقدير العلامة الحقيقية عندما تتوافر هذه الشروط في الاختبار.

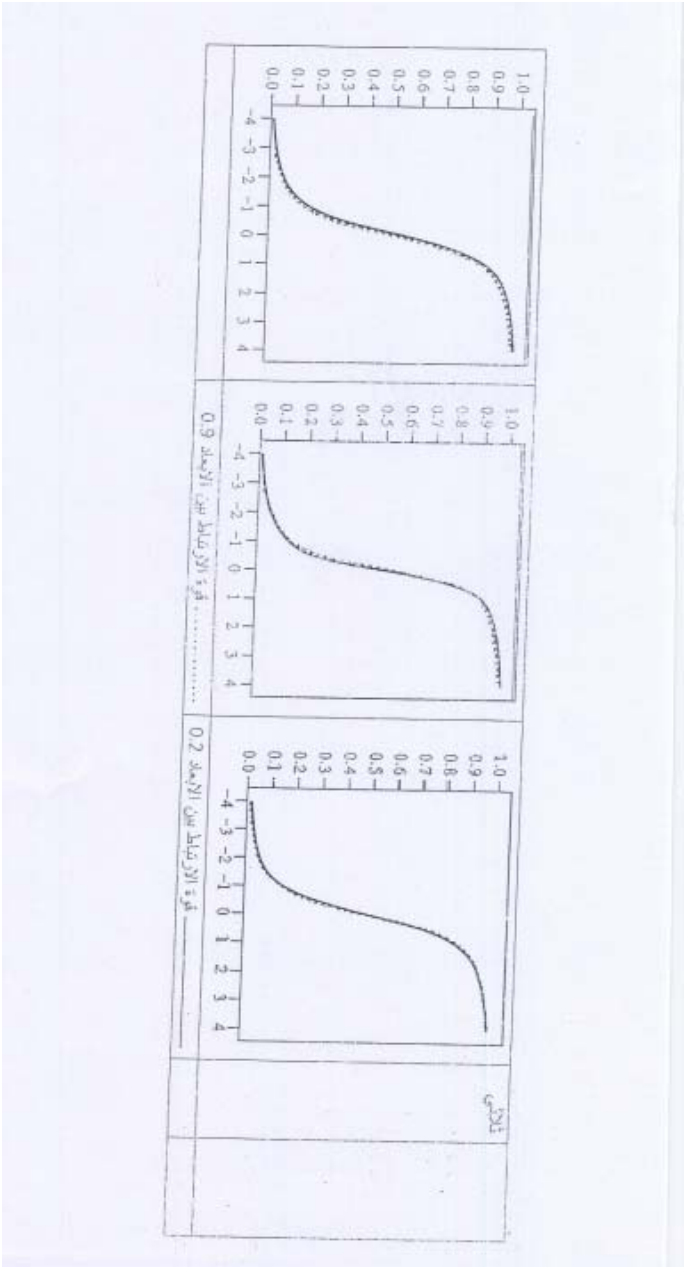
- لأن الدراسة تناولت حالة عدم تساوي عدد الفقرات التي تشكل كل بعد من أبعاد الاختبار يوصي الباحث بإجراء دراسات تتناول الحالة التي تتساوي فيها الفقرات المشكلة لأبعاد الاختبار.
- لأن الدراسة تناولت حالة تعددية البعد بين الفقرات يوصي باجراء دراسات تتناول حالة تعددية البعد داخل الفقرات.
- لان الدراسة تناولت حالة نموذج متعدد الأبعاد تعويضي Compensatory Model يوصي الباحث بإجراء دراسات تتناول النماذج متعددة الأبعاد غير التعويضية Non Compensatory Models.

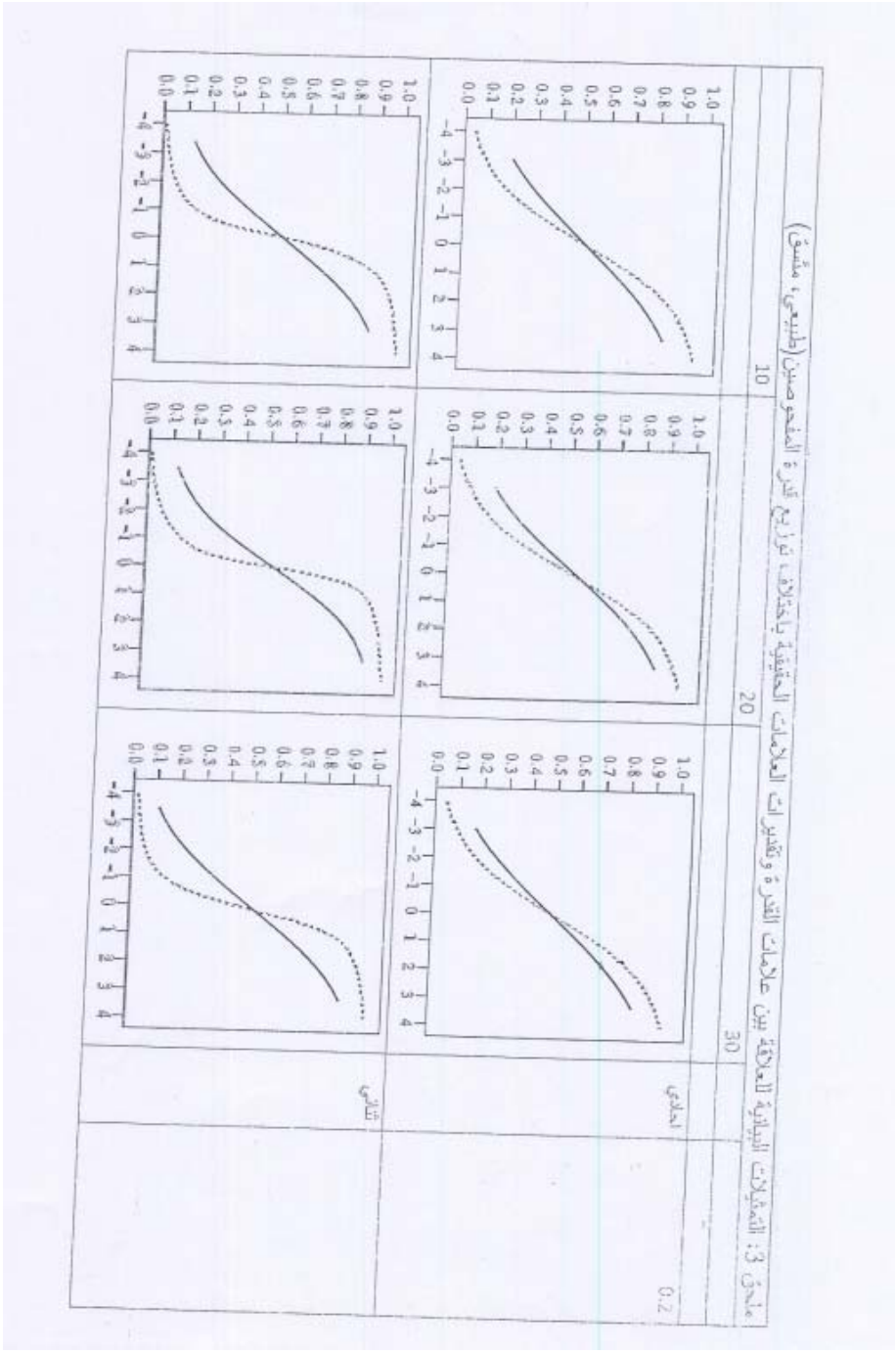


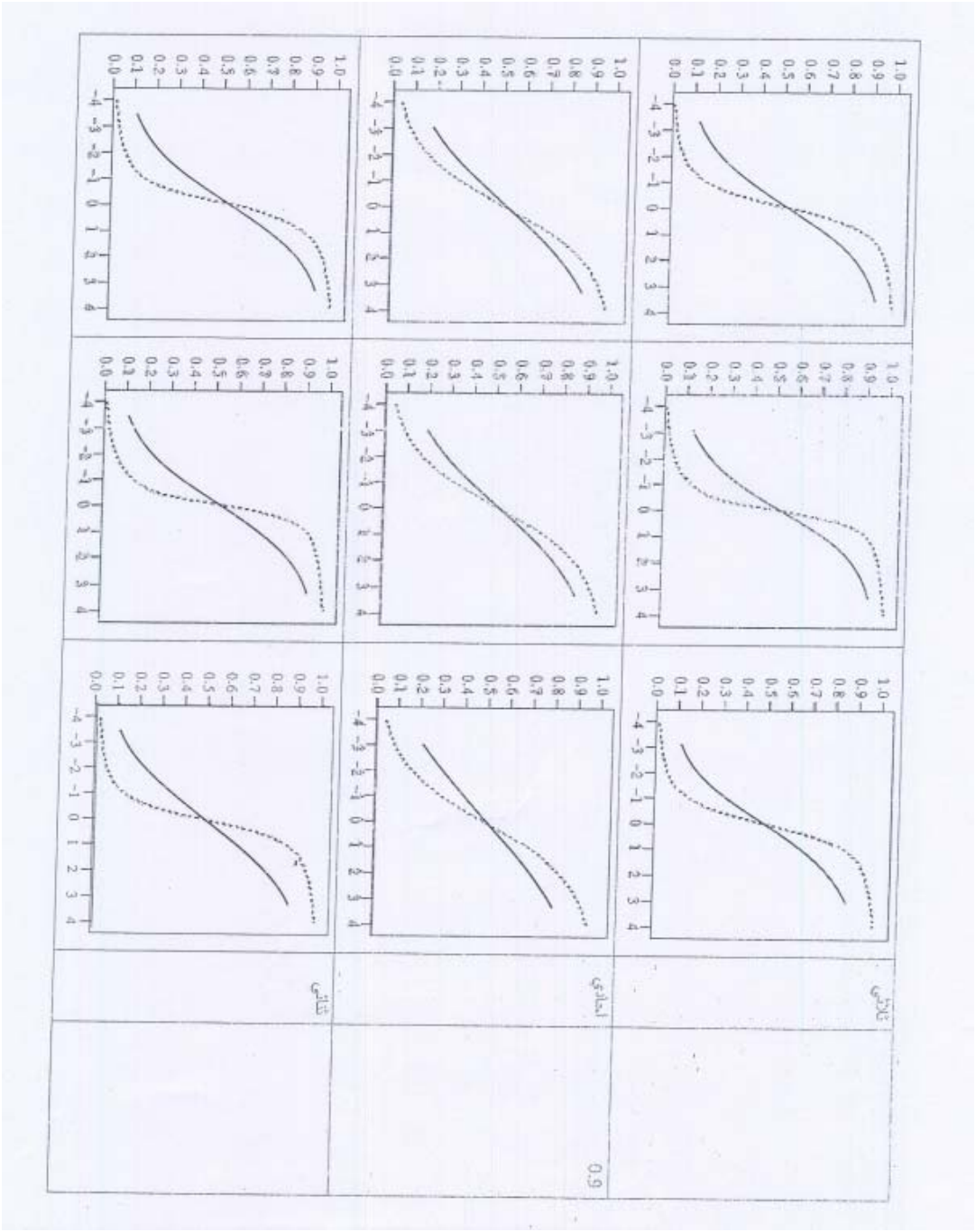

ثلاثي البعد
ثنائي البعد
أحادي البعد

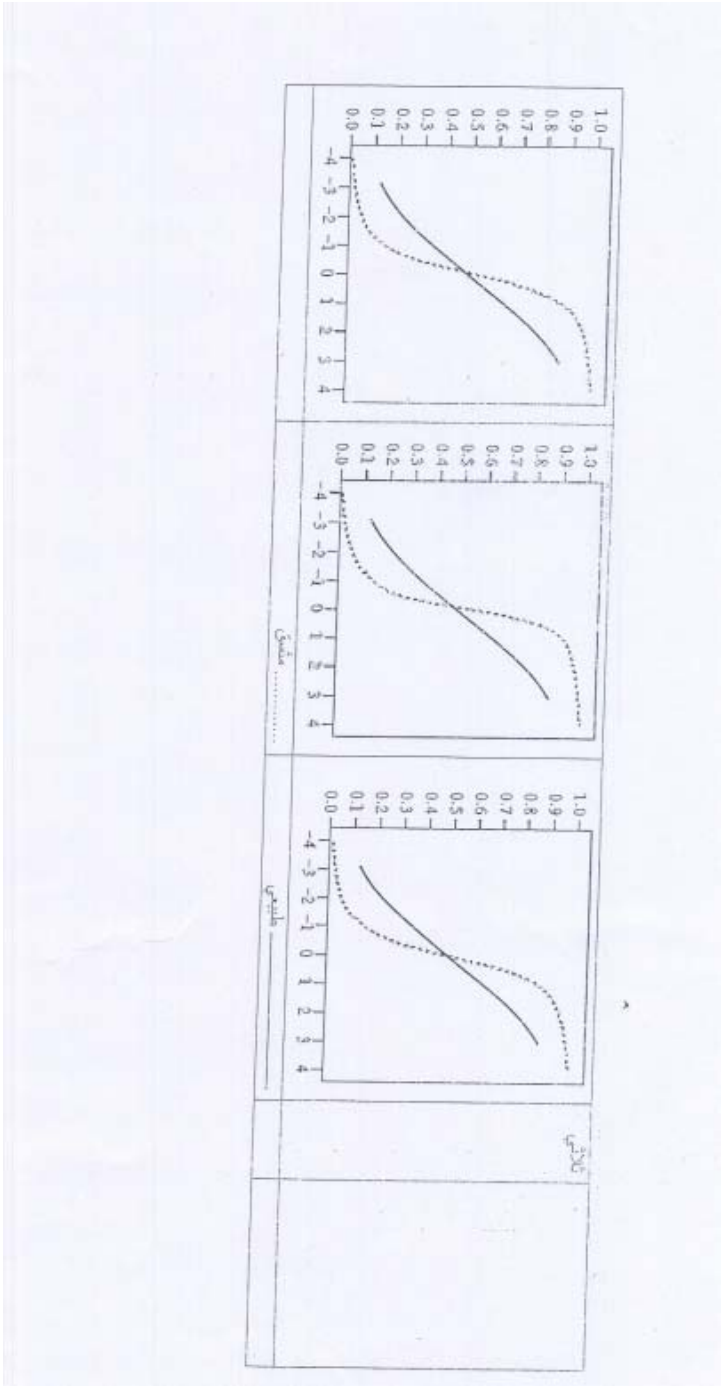

ملحق 2: التمثيلات البيانية للعلاقة بين علامات القدرة وتقديرات العلامات الحقيقية باختلاف قوة الارتباط بين الابعاد (0.2, 0.9)
10
20
30
طبيعي
أحادي
ثنائي


ثلاثي
أحادي
مشترك
ثنائي


قوة الارتباط بين الإيعاز 0.9
قوة الارتباط بين الإيعاز 0.2
الخلفي


ملحق 3: التمثيلات البيانية للعلاقة بين علامات القدرة وتقديرات العلامات الحقيقية باختلاف توزيع القدرة المفحوصين (طبيعي، مشتق)
10
20
30
0.2
الجلدي
الثاني


ثلاثي
احادي
ثنائي
0.9


مشرق
طبيعي
تلاشي

المراجع

ابوشندي، يوسف، 2010، اثر استخدام معايرات أحادية البعد لبيانات متعددة الأبعاد على التقديرات المختلفة لنظرية استجابة الفقرة: دراسة محاكاة. مجلة جامعة دمشق (مقبول للنشر).

ابوشندي، يوسف، 2008، أثر تعدد الأبعاد للاختبار والعلاقة بينها على تقديرات معالم فقراته: دراسة محاكاة. رسالة دكتوراه، جامعة اليرموك، إربد، الأردن.

حجازين، نائل، 2007، أثر تعدد الأبعاد في تقدير معالم فقرات ثنائية التدريج ومتعددة التدريج وفي تقدير معالم القدرة باستخدام برامج حاسوبية تفترض أحادية البعد أو التعدد في الأبعاد. رسالة دكتوراه، جامعة عمان العربية، عمان، الأردن.

ذيابات، لؤي، 2007، أثر التغير في طول الاختبار على خصائص توزيع العلامة الحقيقية حسب النموذج الثلاثي المعلم. رسالة ماجستير غير منشورة، جامعة اليرموك، إربد، الأردن.

Ackerman, T. A. 1989. Unidimensional IRT calibration of compensatory and noncompensatory multidimensional items. *Applied Psychological Measurement, 13*(2), 113-127.

Ackerman,T.A. 1994.Using multidimensional item response theory to understand what items and tests are measuring. *Applied measurement in education .7*(4),255-278.

Adams, R. J., Wilson, M., & Wang, W. C. 1997. The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21*(1), 1-23.

Ansley, T. N., & Forsyth, R. A. 1985. An examination of the characteristics of unidimensional IRT parameter estimates derived from tow-dimensional data. *Applied Psychological Measurement, 9*(1), 37-48.

Batley, R., & Boss, M. W. 1993. The Effects on parameter estimation of correlated dimensions and a distribution-restricted trait in a multidimensional item response model. *Applied Psychological Measurement, 17*(2), 131-141.

Bock, R., Thissen, D& Zimowski,M. 1997. IRT Estimation of Domain scores. *Journal of educational measurement. 34*(3). 197 – 211.

Bolt, D., & Lall, V. 2003. Estimation of compensatory and noncompensatory multidimensional item response models using markov chain Monte carlo. *Applied Psychological Measurement, 27*(6), 395-414.

Camilli, G., Wang, M & Fesq, J. 1992. The effects of dimensionality on True score conversion tables for the law school admission test. LSAC research report series. *(ERIC Document Reproduction Service No. ED 468754).*

Cook, L., Dorans, N., Eignor, D & Petersen,N. 1983. An assessment of the relationship between the assumption of unidimensionality and the quality of IRT True – score Equating. *(ERIC Document Reproduction Service No. ED 235190).*

De Ayala, R. J. 1993. *The Influence of multidimensionality on the graded response model*. Paper presented at the annual meeting of the National Council on Measurement in Education. April, 1993, Atlanta, GA.

De Champlain, A. 1995. Assessing the effect of multidimensionality on IRT True score equating for subgroups of examinees. (*ERIC Document Reproduction Service No. ED 382643).*

Embretson, S., & Reise, S. 2000. *Item response theory for psychologists.* New Jersey: Lawrence Erlbaum Associates,Inc.

Hambleton, R. K., & Swaminithan, H. 1985. *Item response theory: Principles and Applications*. Norwell, MA: Klumer Academic publishers.

Kiresci, L., & Hsu, T. 2001. Robustness of item parameter estimation programs to assumptions of unidimensionality and normality. *Applied Psychological Measurement, 25*(2), 146–162.

Knol, D. L., & Berger, M. P. 1991. Empirical comparison between factor analysis and multidimensional item response models. *Multivariate Behavioral Research, 26*(3), 457-477.

Lord, F.M. 1980. *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Meara, K., Robin, F., & Sireci, S. G. 2000. Using multidimensional scaling to assess the dimensionality of dichotomous item data. *Multivariate Behavioral Research, 35*(2), 229-259.

Mckinley, R., & Way, w. 1992. The feasibility of modeling secondary TOEFEL Ability dimensions using multidimensional IRT models *(ERIC Document Reproduction Service No. ED 385583)* Pommerich, M& Nicewander, W. 1998. Estimating average domain score. *(ERIC Document Reproduction Service No. ED 425175).*

Reckase, M. D. 1997. The past and future of multidimensional item response theory. *Applied Psychological Measurement, 21*(1), 25–36.

Reckase, M. D. 1985. The difficulty of test items that measure more than one ability. *Applied Psychological*

Wang, S. & Wang, N. 2003.The effects of multidimensional polytomous response data on unidimensional many-FACT Rash model parameter estimates. *(ERIC Document Reproduction Service NO.ED 478079)*.

Wang, X. B., Harris, V., & Rosses, L. 2002. *Effects of multidimensionality on IRT item characteristics and true score estimates: implications for computerized test assembly.* Low school Admission council, Newton, PA.

Way, W. D., Ansley, T. N., & Forsyth, R. A. 1988. The comparative effects of compensatory and noncompensatory two-dimensional data on unidimensional IRT estimation. *Applied Psychological Measurement, 12*(3), 239-252.

Measurement, *9*(4), 401-412.

Reckase, M. D., & McKinley, R. L. 1991. The discriminating power of items that measure more than one dimension. *Applied Psychological Measurement, 15*(4), 361–373.

Torre, J.D., & Patz, R. J. 2002. A multidimensional item response theory approach to simultaneous ability estimation *(ERIC Document Reproduction ServiceNo. ED 464142)*

Van Abswoude, A. A., Van der Ark, A., & Sijtsma, K. 2004. A comparative study of test data dimensionality assessment procedures under nonparametric IRT models. *Applied Psychological Measurement, 28*(1), 3–24.

Robustness Domain Score of the Change in Test of Dimensionality and the Correlation between Dimensions and Ability Distribution: A Generated Data

Yousef Abdelkader Abu shindi *

ABSTRACT

This study aimed at investigateing the effects of the test of dimensionality Domain Score calibration. Also. It investigated the effects of the correlation between dimensions (0.2, 0.9) and ability distribution (Normal, Uniform) on Domain Score calibration. To achieve the purpose of the study, responses for four groups (1000 examinee for each group) were generated on seven tests (3 unidimensional, 3 bidimensional, and 1multidimensional).

The results indicate that there were differences in Domain score calibration due to the test of dimensionality (unidimensional, bidimensional and multidimensional) and ability distribution. Results showed no differences due to correlation between dimensions.

Keywords: Domain Score**,** Item response theory, Unidimensional, Multidimensional, Ability parameter, Correlation between Dimensions.

***** Faculty of education, Zarqa University, Jordan. Received on 30/11/2010 and Accepted for Publication on 5/10/2011.