أثر طريقة تقدير معالم الفقرة وقدرات الأفراد على قيم معالم الفقرة، والخصائص السيكومترية للاختبار، في ضوء تغيُّر حجم العينة

د. نضال كمال الشريفين

كلية التربية - جامعة اليرموك
الأردن

الملخص

هدفت الدراسة إلى الكشف عن أثر طريقة تقدير معالم الفقرة وقدرات الأفراد على قيم معالم الفقرة، والخصائص السيكومترية للاختبار، في ضوء تغيّر حجم العينة. لتحقيق هدف الدراسة تم بناء اختبار تحصيلي في الفيزياء من نوع الاختيار من أربعة بدائل، تكوّن بصورته النهائية من (٣٣) فقرة. طبق الاختبار على عينة الدراسة المكونة من (١٠٠٠) طالبٍ وطالبة من طلبة الصف الثاني الثانوي العلمي، وحللت النتائج وفق النموذج الثلاثي المعلمة باستخدام البرمجية (Bilog-Mg). وبينت النتائج وجود فروق ذات دلالة إحصائية (α = ٠,٠٥) في متوسطات الأخطاء المعيارية لتقديرات معالم الفقرات تعزى للتفاعل بين طريقة التقدير وحجم العينة، في حين لم تظهر فروق ذات دلالة إحصائية تعزى لمتغير حجم العينة أو طريقة التقدير، كما أشارت النتائج إلى وجود فروق ذات دلالة إحصائية (α = ٠,٠٥) في متوسطات الأخطاء المعيارية لتقديرات القدرة للأفراد تعزى لمتغير حجم العينة، وللتفاعل بين طريقة التقدير وحجم العينة، في حين لم تظهر فروق ذات دلالة إحصائية تعزى لطريقة التقدير. كما بيّنت النتائج عدم وجود فروق دالة إحصائياً بين معاملات الثبات المقدرة وفق نظرية الاستجابة للفقرة عند أحجام العينة المختلفة (١٠٠، ٥٠٠، ١٠٠٠). وأشارت النتائج إلى أن دقة تقديرات معلمة القدرة تزداد في حالة عينة الأفراد ذوي القدرة العالية وعينة الأفراد ذوي القدرة المتدنية عند استخدام طريقة بييز التوقع (EAP)، في حين تزداد الدقة عند مستويات الأفراد ذوي القدرة المتوسطة باستخدام طريقة الأرجحية العظمى (MLE) بغض النظر عن حجم العينة.

المقدمة

تعد نظرية الاستجابة للفقرة Item Response Theory (IRT) ثورة في القياس النفسي والتربوي، اتضحت فائدتها في التغلب على كثير من مشكلات القياس التي عجزت النظرية الكلاسيكية عن مواجهتها (علام، ٢٠٠٥). وتقوم نظرية الاستجابة للفقرة على افتراض مفاده أنه يمكن التنبؤ بأداء الأفراد، أو تفسير أدائهم في اختبار نفسي أو تربوي معين، في ضوء خاصية أو خصائص مميزة لهذا الأداء تسمى السمات (Traits)؛ إذ تحاول ربط العلاقة القائمة بين مستوى سمة معينة لدى الفرد التي يقيسها اختبار معين، واستجابته لفقرة من فقرات الاختبار عن طريق منحنى خصائص الفقرة (Item Characteristic Curve (ICC))، حيث إن هناك علاقة منتظمة بين مستويات السمة المقيسة لدى الأفراد المختلفين، وبين احتمالات الإجابة الصحيحة عن فقرات مختلفة، التي توصف بأنها دالة متزايدة وتيرياً (Monotonically Increasing)، بمعنى أنه كلما زادت قدرة الفرد، زاد احتمال إجابته الصحيحة عن الفقرة (Hambleton, 1989). ويعتمد منحنى الخصائص للفقرة (ICC) على ثلاثة معالم: الصعوبة (b)، والتمييز (a)، والتخمين (c).

وتستند نظرية الاستجابة للفقرة على افتراضين أساسيين: الافتراض الأول: أحادية البعد (Unidimensionality)؛ وهذا الافتراض يعني أن هناك قدرة واحدة، أو سمة واحدة تفسر أداء الفرد على الاختبار، ونظراً لوجود عوامل أخرى شخصية، وعقلية مثل: الدافعية، والقلق، والطموح، ومفهوم الذات...الخ، وعوامل أخرى لها علاقة بظروف تطبيق الاختبار، فإن هذا الافتراض لا يتحقق بشكل مؤكد، ومع ذلك لا بد من وجود عامل أساسي في الاختبار، وهو ما يشار إليه بالقدرة التي يقيسها الاختبار، والافتراض الثاني: الاستقلال الموضعي (Local Independence)، ويعني هذا الافتراض أن تكون استجابات الفرد على الفقرات المختلفة في الاختبار مستقلة استقلالاً إحصائياً (Statistically Independent) عند مستوى قدرة معين. وحتى يكون هذا الافتراض صحيحاً، يجب أن تؤثر استجابة

المفحوص على فقرة ما في الاختبار سلباً أوايجاباً على استجابته على الفقرات الأخرى في الاختبار (Hambleton, Swaminathan & Rogers, 1991).

وانبثقت عن هذه النظرية مجموعة من النماذج التي تعرف بنماذج السمات الكامنة (Latent Trait Models)، وتعد النماذج السكونية أحادية البعد الأكثر شيوعاً في تصميم الاختبارات والمقاييس التربوية والنفسية وبنائها، وهي الأكثر ملاءمة للفقرات ثنائية التدريج، ويعد عـدد المعالم التي توصف الفقرة بها الفارق الأساسي بين نماذج السمات الكامنة أحادية البعد (De Gruijter & Van Der kamp, 2005؛ علام، ٢٠٠٦).

ويعد النموذج اللوجستي ثلاثي المعلمة (Parameter Logistic (PLM_3)) (Model Three-المستخدم لأغراض تحليل بيانات هذه الدراسة أقل النماذج أحادية البعد تشدداً؛ إذ يسمح بأن تختلف فقرات الاختبار في صعوبتها وتمييزها، وهو ما يلاحظ عادة في بناء الاختبارات. إذ من الصعب إيجاد مجموعة من الفقرات التي تميز بدرجة واحدة بين مستويات السمة، أو القدرة التي يقيسها اختبار معين، وهذا الافتراض الذي اعتمده النموذج الأحادي المعلمة (One Parameter Logistic Model or Rasch Model). كذلك يفترض تأثر الإجابات بعامل التخمين وهو المعلم الثالث الذي تميز به عن النموذج الثنائي المعلمة (Two -Parameter Logistic Model)، الذي أضافه لورد، وأطلق عليه معلم الخط التقاربي الأدنى (Lower Asymptote Line)، أو معلم التخمين (Guessing Parameter)، وهذا المعلم يحدد احتمال أن يجيب فرد يفترض أن مستوى قدرته متدنٍ جداً، ومع هذا يجيب إجابة صحيحة عن فقرات الاختبار عن طريق التخمين، والحقيقة أن إضافة هذا المعلم الثالث يكون بمثابة تعويض عدم مطابقة المنحنيات المميزة للفقرات عند النهاية السفلى لمتصل القدرة؛ نتيجة تأثر إجابات هؤلاء الأفراد بعامل التمييز (Lord, 1980; Hambleton, 1989). ويوصف النموذج اللوجستي الثلاثي المعلمة بالمعادلة الرياضية الآتية:

$$P_i(\theta) = c_i + \frac{1 - c_i}{1 + e^{Dai(\theta - bi)}} (i = 1, 2, ..., n)$$

حيث $p_i(\theta)$: احتمال أن يجيب المفحوص ذو القدرة (θ) عن الفقرة (i)

إجابة صحيحة، D: ثابت التدريج ويساوي تقريباً 1.7، b_i: معلم الصعوبة للفقرة (i) وهي النقطة التي تقع على متصل القدرة عندما يكون ميل منحنى خصائص الفقرة أكبر ما يمكن، a_i: معلم تمييز الفقرة، وهو ميل منحنى خصائص الفقرة عند انعطاف المنحنى، c_i: خط التقارب السفلي لمنحنى خصائص الفقرة الذي يمثل احتمالية إجابة المفحوص بأقل قدرة إجابة صحيحة على الفقرة (معلم التخمين للفقرة). ويعتمد احتمال إجابة الفرد عن الفقرة إجابة صحيحة في نماذج نظرية الاستجابة للفقرة، على قدرة الفرد (θ)، والمعالم المتعلقة بالفقرة متمثلة بمعلم الصعوبة (b)، ومعلم التمييز (a)، ومعلم التخمين (c)، ولكن هذه القيمة الاحتمالية، وهذه المعالم تكون غير معلومة، بينما إجابات الأفراد عن الفقرة تكون معلومة، لذلك فإن أساليب تقدير المعالم تهدف لتحديد قيمة (θ) لكل فرد، وكذلك قيم معالم الفقرة من إجابات الأفراد عليها.

أساليب تقدير معالم الفقرات وقدرات الأفراد

هناك أسلوبان لتقدير معالم الفقـــرات والأفراد، إما بشكل مشترك (Joint)؛ أي يتم تقدير معالم الفقرات والأفراد معاً؛ أي يتم تقدير معالم الفقرات أولاً بعد حذف معالم قدرات الأفراد عن طريق تكامل اقتران الاحتمالية الهامشي (Marginal)، ثم إيجاد معالم تقديرات القدرة للأفراد. فالأسلوب الأول: يعتمد على تقديرات الأرجحية العظمى Maximum (Likelihood)، بينما يعتمد الأسلوب الثاني: يعتمد على نظرية بييز (Bayes Theory).

وهناك ثلاثة أساليب رئيسة تعتمد عليها طريقة الأرجحية العظمى (ML) لتقدير معالم فقرات النماذج أحادية البعد (De Gruijter & Van Der kamp, 2005; Hambleton, 1989)، وتختلف هذه الأساليب في تعاملها مع مستويات القدرة غير المعلومة، وهي:

- تقديرات طريقة الأرجحية العظمى المشتركة (Joint Maximum Likelihood Estimation): يمكن تطبيق هذه الطريقة في النماذج اللوجستية الأحادية

والثنائية والثلاثية المعلم، ويتم وفقاً لهذه الطريقة تقدير معالم القدرة والفقرة في آن معاً. إذ تعمل هذه الطريقة على نمذجة احتمالات الاستجابات باستخدام تقديرات مبدئية لمستويات قدرة الأفراد.

- تقديرات الأرجحية العظمى الشرطية (Conditional Maximum Likelihood Estimation): وتقوم هذه الطريقة بفصل المعالم الإحصائية للمفحوصين في أثناء عملية التدريج، وتطبق على النموذج اللوجستي الأحادي المعلمة، حيث يكون اقتران الاحتمالية (Likelihood) مشروطاً بعدد الإجابات الصحيحة للأفراد على الفقرات الاختيارية، لذلك يمكن الحصول على تقديرات الأرجحية العظمى الشرطية لمعالم صعوبة الفقرات بغض النظر عن معالم القدرة.
- تقديرات الأرجحية العظمى الهامشية (Marginal Maximum Likelihood Estimation): ويمكن تطبيق هذه الطريقة على النماذج اللوجستية الأحادية والثنائية والثلاثية المعلم، حيث يتم وفقاً لهذه الطريقة إيجاد قيمة (θ) التي تجعل قيمة (θ) أكبر مايمكن من خلال معادلة رياضية، عن طريق إيجاد المشتقة الأولى لتلك المعادلة ومساواتها بالصفر، حيث يتم وفقاً لهذه الطريقة إيجاد اقتران الاحتمالية الهامشي (Marginal Likelihood Function) لمعالم الفقرة، من خلال تكامل اقتران الكثافة (Density Likelihood Function)) على معالم القدرة ثم إيجاد تقديرات معالم الفقرات من خلال المعادلة الرياضية الآتية:

$$L(a,b,c) = \prod \int_{\infty}^{\infty} g(\theta_a) L(\theta_a; a,b,c) d\theta_a.$$

أما عن أساليب تقدير معالم القدرة للأفراد فيتم تقديرها بأسلوبين: يعتمد الأسلوب الأول على تقديرات الأرجحية العظمى (ML)، بينما يعتمد الأسلوب الثاني على تقديرات منهج بييز (Bayesian Estimation). ويعتمد برنامج (Bilog - Mg_3) النسخة المطورة من خلال شركة البرامج العلمية الدولية (Scientific Software International, Inc) الذي اعتمده الباحث في تحليل

البيانات، في تقدير معالم الفقرات طريقة الأرجحية العظمى الهامشية (Zimowski, Muraki, Mislevy & Bock, 2003). وتتميز هذه الطريقة بأنه يمكن استخدامها في تقدير معالم جميع النماذج أحادية البعد، وكذلك النماذج متعددة الأبعاد، كما تتميز هذه الطريقة بالفاعلية (Efficiency)، سواء كانت فقرات الاختبار قليلة أو كثيرة. والقيم التقديرية للأخطاء المعيارية الناتجة تتميز بالدقة، عن طريق إعادة متعاقبة لعمليات التقدير، ويمكننا الحصول على قيم تقديرية لمعالم الأفراد الذين أجابوا إجابة صحيحة أو خاطئة عن جميع الفقرات، وتعطي تقديرات للعلامة الكلية، وبالتالي لا يوجد فقدان للمعلومات عائد لحذف استجابة بعض أفراد الدراسة، إضافة إلى أن التقديرات الناتجة تكون متسقة، وتقترب من القيم الحقيقية بزيادة حجم العينة (علام، ٢٠٠٥)، وتفترض هذه الطريقة التوزيع الطبيعي لتقديرات القدرة للأفراد، وتتضمن طريقتين لحل معادلة الأرجحية العظمى الهامشية: الطريقة الأولى: وتتضمن مرحلتين إحداهما: مرحلة التوقع (Expectation)، والأخرى مرحلة التعظيم (Maximization)، ويرمز لهما اختصاراً (EM)، وهذا البرنامج يعتمد على تكرار الخطوات. أما الطريقة الثانية: تتضمن أسلوب نيوتن- جاوس (Newton-Gauss)، وهذا الطريقة تقوم بالبحث بطريقة متكررة، حيث يتم فيها تحسين القيم التقديرية لهذه المعالم بالتتابع، بحيث تحقق القيم النهائية محك التقارب (Zimowski, Muraki, Mislevy & Bock, 2003).

كما يعتمد برنامج (Bilog - Mg_3) طريقتي الأرجحية العظمى (ML) وبييز (BE) في تقديرات القدرة للأفراد -كما ذكر مسبقاً - ويتم تقدير تقديرات القدرة للأفراد عن طريق الأرجحية العظمى (ML)، عندما تكون معالم الفقرات معروفة عن طريق الأرجحية العظمى الهامشية. ويتم تقدير معالم قدرة الأفراد (θ) في هذه الحالة عن طريق تعظيم اقتران دالة الأرجحية والتي تعطى بالعلاقة الآتية:

$$LogL_i(\theta) = \sum_{j=1}^{n} \{x_{ij} Log_e p_j(\theta) + (1 - x_{ij}) Log_e [1 - p_j(\theta)]\}$$

حيث إن: $p_j(\theta)$ احتمالية أن يجيب المفحوص ذو القدرة (θ) عن الفقرة (i) إجابة صحيحة، i: رقم الفقرة، n: عدد فقرات الاختبار.

ومن ثم إيجاد تقديرات القدرة للأفراد (θ) التي تمثل أكبر ترجيح عن طريق إيجاد المشتقة الأولى لاقتران الترجيح ومساواته بالصفر والمعطى بالعلاقة الآتية:

$$\frac{\partial dLogLi(\theta)}{\partial \theta} = \sum_{j=1}^{n} \frac{xi_j - p_j(\theta)}{p_j(\theta)[1 - p_j(\theta)]} \cdot \frac{\partial p_j(\theta)}{\partial \theta} = 0$$

وبعد ذلك يتم القيام بالتدوير المتعاقب (Iteration) حسب طريقة نيوتن رافسون (Newton-Raphson Method)، من أجل الحصول على تقديرات ثابتة لمعلم القدرة من خلال معادلة فشر (Fisher -Scoring Solution). والخطأ المعياري المحسوب وفق هذه الطريقة يتم من خلال حساب الجذر التربيعي لمعكوس دالة المعلومات (Zimowski et al., 2003).

أما الأسلوب الثاني المعتمد في تقدير قدرات الأفراد، والمعتمد كذلك في برنامج (Bilog - Mg_3) فهو طريقة بييز (Bayesian Estimation): حيث تستخدم هذه الطريقة في العادة عندما لا نستطيع تطبيق طريقة الأرجحية العظمى، وتتميز هذه الطريقة باستخدامها لمعلومات سابقة عن العينة تسمى المعلومات الأولية (Priori Information)، والمتوفرة من خبرات سابقة. وفي منهج بييز يتم اعتبار كمية ذات تباين معروف، من خلال وصف التوزيع الاحتمالي لـ (θ)، يطلق عليه التوزيع القبلي .(Priori distribution) ويعتمد تحديد هذا التوزيع الاحتمالي على معرفة البحث والخبرة السابقة عن (θ)، وذلك قبل أن يتم جمع البيانات؛ ولهذا السبب يطلق عليه التوزيع القبلي، ويتم مراجعة التوزيع القبلي استناداً للمعلومات التي توفرها العينة، حيث يطلق على التوزيع الجديد اسم التوزيع البعدي Posteriori distribution (Zeng, 1997; Lord, 1986; De Gruijter & Van Der kamp, 2005). ويوصف التوزيع الاحتمالي البعدي لـ θ بالمعادلة الرياضية الآتية:

$$(\theta_k/x) = \frac{p(x/\theta_k)g(\theta_k)}{p(x)} = \frac{p(x/\theta_k)g(\theta_k)}{\sum_{h=1}^{q} \sum p(x/\theta_h)g(\theta_h)}, k = 1, ..., q$$

حيث إن: $g(\theta_k)$: التوزيع الاحتمالي القبلي لـ (θ_k)، $p(x/\theta) = L(x/\theta)$: هو اقتران الأرجحية (Likelihood Function) للدرجات الملاحظة $x = (x_1,.....x_n)$ إذا أعطيت (θ).

وفي منهج بييز يتم تقدير قدرات الأفراد بأسلوبين الأول: يطلق عليه توقع الاقتران البعدي (Expected A Posteriori, EAP)، أو تقدير بييز (Bayesian Estimation)، أما الأسلوب الثاني فيطلق عليه طريقة بييز الشرطية (Bayes Modal Estimation)، أو تعظيم الاقتران البعدي (Maximum A Posteriori, MAP)، وتعمل طريقة بييز على إيجاد تقديرات القدرة للأفراد بالاستناد إلى تقديرات معالم الفقرات، من خلال طريقة الأرجحية العظمى الهامشية. ويحسب توقع الاقتران البعدي (EAP) من خلال الوسط الحسابي للتوزيع البعدي لـ (θ)، دون اللجوء إلى تقريب متتابع كما في طريقة الأرجحية العظمى (ML).

وتقاس درجة الدقة من خلال الخطأ المعياري المحسوب من خلال الانحراف المعياري للتوزيع البعدي لـ (θ) (Posterior Standard Deviation) (PSD)، وتقديرات توقع الاقتران (EAP) لها أقل معدل خطأ معياري من التقديرات الأخرى بما فيها الأرجحية العظمى، وتعظيم الاقتران Mislevy & Bock, 2003). أما طريقة بييز الشرطية، أو تعظيم الاقتران (MAP)، فهي مشابهة لإجراءات طريقة الأرجحية العظمى، حيث يتم القيام بالتقريب المتتابع لإعطاء تقديرات ثابتة لمعلم القدرة، ويتم تقدير (θ) عن طريق تعظيم (θ) من خلال معادلة الترجيح الآتية:

$$P(\theta/xi) = \sum_{j=1}^{n} (x_{ij} Log_e(\theta) + (1 + x_{ij}) Log_e(1 - p_j(\theta)) + [Log_e g(\theta)]$$

حيث: $g(\theta)$: اقتران الكثافة ذو التوزيع المتصل لـ (θ). وبعد ذلك يتم إيجاد قيمة القدرة (θ) التي تمثل أكبر ترجيح من خلال تفاضل المعادلة السابقة ومساواتها بالصفر، ومعادلة الترجيح هذه تشبه طريقة الأرجحية العظمى، إذ يتم تقديرها بعد ذلك عن طريق معادلة فشر، ومما يؤخذ على هذه الطريقة أنها تعطي أخطاء معيارية أعلى من طريقة بييز التوقع (EAP)، إلا أنها أسهل

في حسابها من طريقة توقع الاقتران البعدي، ويعد تباين الخطأ معياراً لدقة القياس، ويتم حساب الخطأ المعياري في التقدير من خلال الجذر التربيعي لمعكوس دالة المعلومات (Mislevy & Bock, 2003; Lord, 1986).

وتختلف عملية تقدير معالم الفقرات وفق الأرجحية العظمى الهامشية، وفق الطريقة المستخدمة في تقدير معالم قدرات الأفراد. فعند اختيار طريقة الأرجحية العظمى؛ لو افترضنا أن $\delta = (\delta_1, \delta_2, ..., \delta_n)$ تمثل متجه معالم الفقرات للاختبار، و$Xi = (x_{i1}, x_{i2}, ...x_{in})$ يمثل متجه درجات الأفراد الملاحظة، و$g(\theta)$ يمثل اقتران الكثافة، فإن اقتران الاحتمالية الهامشي للدرجات الملاحظة للمتجه Xi يعطى بالعلاقة الآتية:

$$fx_{i\backslash\delta}(X_i \backslash \delta) = \int fx_{i\backslash\theta,\delta}(X_i \backslash \theta, \delta) g(\theta) d\theta$$

أما اقتران الاحتمالية الهامشي لدرجات المفحوصين للمتجه X فتعطى بالعلاقة الآتية:

$$f_{x\backslash\delta}(X \backslash \delta) = \prod f_{xi/\delta}(X_i \backslash \delta)$$

وتعمل الأرجحية العظمى الهامشية على إيجاد تقديرات معالم الفقرات $\hat{\delta}$ وذلك عن طريق تعظيم اقتران الأرجحية العظمى في المعادلة السابقة، كما في المعادلة الآتية:

$$L_1 = Log f_{x\backslash\delta}(x \backslash \delta)$$

وتعمل الأرجحية العظمى بشكل جيد في إيجاد تقديرات أكثر دقة لمعالم الفقرات إذا كان حجم العينة كبيراً، أما إذا كان حجم العينة صغيراً، فإن التقديرات تكون أقل دقة، وخاصة معلم التخمين الخاص بالنموذج الثلاثي المعلمة (Birinbaum, 1986; Gao & Chen, 2005).

أما في حال اختيار طريقة بييز في تقديرات القدرة للأفراد، فإن متجه معالم الفقرات δ يتم تقديره مع الأخذ بالاعتبار التوزيع القبلي للعينة، ويتم تعظيم اقتران التوزيع البعدي لإيجاد تقديرات بييز لمعالم الفقرات $\hat{\delta}$، ووفق

نظرية بييز التوزيع البعدي لـ δ إذا أعطي متجه X لدرجات المفحوصين يعطى بالعلاقة $f_{x\backslash\delta}(x \backslash \delta)$.

وتعمل الأرجحية العظمى الهامشية وفق نظرية بييز على إيجاد تقديرات معالم الفقرات δ، وذلك عن طريق تعظيم اقتران الأرجحية العظمى في المعالجة السابقة كما في المعادلة الآتية:

$$L_2 = Logf_{x\backslash\delta}(x \backslash \delta) + \log_{\delta}(\delta)$$

ويلاحظ من المعادلتين L_1 و L_2، أن $L_2 = L_1 + \log_{\delta}(\delta)$؛ هذا يعني أنه بالإضافة إلى تقديرات الأرجحية العظمى الهامشية، فإن تقديرات بييز تأخذ بالاعتبار معلومات أولية عن المعالم، وتسمى هذه الطريقة Marginal Bayes Model Estimation (Gao & Chen, 2005)، أو Maximum Marginal A Posteriori Estimation of Item Parameters (Zimowski, et al., 2003). وتعتمد تقديرات بييز الهامشية (MBME) على التوزيع القبلي أو المعلومات الأولية، فإذا كان التوزيع القبلي منتظماً غير غني بالمعلومات فإن تقديرات (MMLE) وتقديرات (MBME) تقودنا للنتائج والتقديرات نفسها. أما إذا كان التوزيع القبلي توزيعاً طبيعياً فإن طريقة بييز أكثر دقةً، كذلك تعتمد على حجم العينة وطول الاختبار، فإذا كان حجم العينة، وطول الاختبار كبيرين، فإن كلا الطريقتين تقودنا للنتائج نفسها، أما إذا كان التوزيع القبلي (Informative)، وذلك عند تقليل طول الاختبار وحجم العينة ستكون مساهمة L_2 كبيرة، وبالتالي فإن تقديرات (MBME) تختلف عن تقديرات (MMLE) (Gao &Chen, 2005). أو أن تكون الفقرات صعبة جداً، أو سهلة جداً، وهذا يقودنا إلى تقديرات غير دقيقة، لذلك تستخدم هذه الطريقة بشرط أن تكون حجم عينة المفحوصين أقل أو تساوي ٢٥٠ مفحوصاً (Zimowski et al., 2003).

وقد وجهت لطريقة بييز بعض الانتقادات تتمثل في طبيعة المعلومات الأولية المتوفرة من خبرات سابقة لدى الباحثين، إلا أن طريقة بييز يتوقع فيها تحسين دقة التقدير وخاصة إذا كان حجم العينة صغيراً، في الحالة التي يكون

فيها تقدير الأرجحية العظمى لا يعمل بشكل جيد(Miselvy, 1986; Gao &Chen, 2005) . وقد أشار العديد من الباحثين (Garre & Vermunt,2006; Gao & Chen, 2005; Lewy & Nielsen, 2002) أن استخدام إحصاء بييز في تقدير معالم نماذج السمات الكامنة يعوزه كثير من البحوث والدراسات حتى تكتمل صورته، فهو يتطلب أساليب جديدة لتحديد المعلومات القبلية المتعلقة بمعالم هذه النماذج تحديدً كمياً، كما يتطلب تصميم برامج خاصة بالحاسوب لإجراء العمليات الرياضية المعقدة اللازمة لإجراء هذا النوع من التقدير. بينما يشير علماء القياس أن منهج بييز يقدم حلولاً لمشكلات التقدير لمعالم نماذج السمات الكامنة عندما يكون حجم عينة الأفراد صغيراً، وعدد فقرات الاختبار قليلاً؛ ففي مثل هذه الحالات يصبح للمعتقدات القبلية أهمية كبيرة، أما إذا كان حجم العينة كبيراً، فإن هذه المعتقدات تبدو قليلة الأهمية في مثل هذه التقديرات (علام، 2005).

وتعتمد عملية تقدير معالم الفقرات وتقديرات القدرة للأفراد على عدة عوامل من أهمها: اختيار الطريقة المناسبة للتقدير، فقد أشار بعض الباحثين أن طريقة بييز تعمل بشكل أفضل من طريقة الأرجحية العظمى مع الاختبارات القصيرة، وحجم عينة أقل. كما تؤدي إلى نتائج أقل دقة في حالة استخدام العينات المتجانسة (Homogeneous Samples)، وتؤدي لنتائج أقل دقة في حالة العينات غير المتجانسة (Heterogeneous Sample) (De Gruijter & Van Der kamp, 2005). كما تبين وجود آراء مختلفة حول أيهما أكثر دقة في تقدير معالم الفقرات والأفراد وفق ظروف مختلفة، لذا تأتي هذه الدراسة لتقدم دليلاً عملياً على أكثر طرق التقدير دقة في تقدير معالم الفقرات والأفراد باختلاف حجم العينة.

الخطأ المعياري للتقدير ودالة المعلومات والفعالية النسبية وعلاقتها بالثبات

بعد عملية تقدير معالم الفقرات وتقديرات القدرة للأفراد، فإنه يجب التحقق من مدى دقة التقدير، فقد بين لورد (Lord, 1980) وجود معايير

مختلفة للكشف عنها من أبرزها: الخطأ المعيـــاري في التقدير (S.E.E) (Standard Error of Estimation)، ودالة المعلومات للفقرات والاختبار (Information Function)، وهما موضع اهتمام الباحث في هذه الدراسة، والفعالية النسبية للاختبار (Relative Efficiency). ويعرف الخطأ المعياري في القياس حسب النظرية الكلاسيكية، والمعطى العلاقة الآتية:

$$SE_m = \sigma_x \sqrt{1 - P_{xx}}$$

حيث تمثل P_{xx}: معامل الثبات حسب النظرية الكلاسيكية.

σ_{xx} : الانحراف المعياري للدرجات الخام.

والخطأ المعياري في القياس حسب النظرية الكلاسيكية يناظر الخطأ المعياري في التقدير حسب النظرية الحديثة في القياس، وحيث إن مربع الخطأ المعياري في التقدير يرتبط بالثبات، فإن ذلك يشير إلى أنه إذا قل مقدار الخطأ المعياري، فإن ذلك يــؤدي إلى زيادة الثبات، وزيادة دقة الاختبار وثباته (Reeve, 2004).

ومن المؤشرات الأخرى التي يستدل منها على معامل الثبات للاختبار في النظرية الحديثة في القياس هي دالة المعلومات للاختبار (Information Function)، ومنحنى دالة المعلومات للاختبار ينتج من تكوين منحنيات الفقرات فوق بعضها بعضا وفق العلاقة: $I(\theta) = \sum_j Ig(\theta)$

حيث $(I(\theta))$: هي كمية المعلومات للاختبار عند مستوى القدرة، $\sum_g Ig(\theta)$ مجموع دوال المعلومات لفقرات ذلك الاختبار عند مستوى القدرة نفسه. ولذلك فإن زيادة عدد الفقرات يعطي خطأ معيارياً صغيراً (θ) S.E، ونقصان قيمة الخطأ المعياري في تقدير القدرة عند مستوى القدرة يؤدي إلى زيادة كمية المعلومات للاختبار وفق العلاقة الآتية:

$$I(\theta) = \frac{1}{\sqrt{S.E(\theta)}}$$

ونقصان قيمة الخطأ المعياري في تقدير القدرة يؤدي إلى زيادة قيمة

معامل الثبات شأنه في ذلك شأن النظرية الكلاسيكية. وما يميز هذه الطريقة عن الطرق الكلاسيكية أن تقدير الثبات في النظرية الكلاسيكية مرتبط بالعينة، وهذه خاصية سيئة لتقدير الثبات، كما أنها تؤدي لتقدير جماعي للأخطاء الفردية في علامات الاختبار، وهو ما يسمى بالخطأ المعياري في القياس، بينما في نظرية الاستجابة للفقرة تزودنا بتقدير للخطأ المعياري للقياس عند كل مستوى من مستويات القدرة، ويمكن باستخدامها تحديد مدى مساهمة كل فقرة في تحديد دقة الاختبار (Hambelton & Swaminathan, 1985).

ويشير كروكر والجينا (Crocker & Algina, 1987) أن دالة المعلومات لها فوائد متعددة مثل مقارنة اختبارات مختلفة تقيس السمة نفسها، أو مقارنة طريقتين مختلفتين لتصحيح الاختبار نفسه، بالإضافة إلى مقارنة استراتيجيات اختبار مختلفة مثل المقارنة بين الاختبار التقليدي والاختبار التكيفي وذلك من خلال كمية المعلومات التي يعطيها كل اختبار. ويمكن الاستفادة من دالة المعلومات للاختبار في حساب الفعالية النسبية للاختبار مقارنة باختبار آخر. وقد عرف لورد (Lord, 1980) الفعالية النسبية للاختبار Y بالنسبة للاختبار X بأنها النسبة بين دالة المعلومات للاختبارين، ويعبر عنها كالآتي:

$$RE\{Y,X\} = \frac{I\{\theta,Y\}}{I\{\theta,X\}}$$

حيث $I\{\theta,Y\}$، $I\{\theta,X\}$: هي دالة المعلومات للاختبارين X,Y على الترتيب.

ودرجات الاختبارين X,Y يمكن أن تكون الدرجات على اختبارين مختلفين عند مستوى القدرة نفسه، أو نحصل على هذه الدرجات من الاختبار نفسه بطريقتين. وتعرف الفعالية النسبية فقط عندما نأخذ النسبة بين دالة المعلومات للاختبارين X,Y عند مستوى القدرة نفسه، وتختلف الفعالية النسبية لاختبارين باختلاف مستوى القدرة.

ويرى تريفسان وساكس وميشيل (Trevisan, Sax & Michael, 1991) أن

التنبؤ بدرجات الأفراد على الاختبار يكون أكثر دقة إذا ربطنا ذلك بمستويات القدرة، حيث وجدوا أن الاتجاه المفضل لعدد البدائل هو أربعة بدائل، يليه ثلاثة بدائل، ثم خمسة بدائل للطلبة ذوي القدرة المنخفضة، أما بالنسبة لذوي القدرة المرتفعة وذوي القدرة المتوسطة فلا يوجد تفضيل لعدد معين من البدائل، حيث لم توجد فروق دالة إحصائياً، وكان العدد المثالي للبدائل عبر مجموعات القدرة هو أربعة بدائل، وذلك باستخدام النموذج اللوجستي ثلاثي المعلمة.

ويتم استخراج الثبات المقدر وفق نظرية الاستجابة للفقرة الخاص بتقديرات القدرة للأفراد عن طريق الثبات الامبريقي (التجريبي) (Empirical Reliability)، الذي يعتمد بشكل أساسي على الجذر التربيعي لمربعات الأخطاء المعيارية، ويتم تقديره من خلال تباين الدرجات الحقيقية والتباين الخطأ، الذي يعتمد على كيفية تقدير درجات القدرة للمفحوصين في العينة. ويستخرج الثبات الامبريقي بطريقتين: الأولى تعتمد على درجات الأرجحية العظمى (Maximum Likelihood Scores)، والثانية تعتمد على درجات بييز (Bayesen).

أما بالنسبة للطريقة المعتمدة وهي طريقة الأرجحية العظمى (ML)، فإن تقدير تباين الخطأ (Error Variance) يحسب من خلال مقلوب الوسط الحسابي لدالة المعلومات، وتباين الدرجة (Score Variance) هو تباين الدرجات المستخرج من خلال الأرجحية العظمى في العينة، وتباين الدرجات الحقيقية (True Score Variance) يقدم من خلال طرح تباين الخطأ من تباين الدرجة، حيث يستخرج الثبات الإمبريقي من خلال قسمة التباين الحقيقي على تباين الدرجة. أما بالنسبة للثبات المستخرج وفق طريقة توقع الاقتران (EAP) فيتم من خلال الوسط الحسابي لتباين التوزيع البعدي لتقديرات القدرة للأفراد ولجميع المفحوصين في العينة، وتباين الدرجة الحقيقية يقدر مباشرة من تباين الوسط الحسابي للتوزيع البعدي في العينة. حيث يستخرج الثبات الامبيريقي في هذه الحالة من خلال قسمة تباين الدرجة الحقيقية على مجموع تباين الدرجة الحقيقية وتباين الخطأ. أما في حالة تعظيم الاقتران (MAP)، يقدر تباين

الخطأ من خلال مقلوب أفضل دالة معلومات للتوزيع البعدي لجميع المفحوصين في العينة، وتباين الدرجة الحقيقية يقدر من خلال الوسط الحسابي لتباين التوزيع البعدي في مرحلة التعظيم، حيث يستخرج الثبات الامبريقي في هذه الحالة من خلال قسمة تباين الدرجة الحقيقية على مجموع تباين الدرجة الحقيقية وتباين الخطأ .(Zimowski et al., 2003)

الدراسات السابقة

يتناول هذا الجزء مراجعة للدراسات ذات الصلة بمشكلة البحث، وقد تم عرض هذه الدراسات وفق التسلسل الزمني والمنطقي، فقد عرضت الدراسات الأجنبية أولاً؛ نظراً لعمق التجربة في هذا المجال، ومن ثم عرضت الدراسات العربية ذات الصلة بمشكلة الدراسة، والاستعانة ببعض الدراسات حول سبب اختيار النموذج ثلاثي المعلمة.

فقد تباينت وجهات النظر حول الطريقة الفضلى في دقة تقدير معالم الفقرات وتقديرات القدرة للأفراد، ولا زال هناك جدل واضح. فمن الدراسات الأجنبية ذات الصلة بمشكلة الدراسة، دراسة روسو وريكيس (Rosso & Reckase, 1981) التي هدفت إلى مقارنة طريقتي التقدير، وهما طريقة الأرجحية العظمى MLE، وطريقة بييز EAP في الاختبارات التكيفية. أشارت نتائج الدراسة إلى عدم وجود فروق في تقديرات الثبات في الاختبار التكيفي، كما بينت النتائج أن استخدام طريقة الارجحية العظمى قد أعطت كمية معلومات أقصى على الاختبار ككل من طريقة بييز، أما الاختلاف الرئيس بين الطريقتين فيكمن بوجود اختلاف في تقديرات القدرة بينهما، إذ تعد طريقة الأرجحية العظمى هي الطريقة الفضلى.

أما دراسة مكينلي وريكيس (Mckinley & Reckase, 1981) التي هدفت إلى المقارنة بين طريقتي بييز وطريقة الأرجحية العظمى في إجراء الاختبارات التكيفية، وذلك باستخدام التباين البعدي لتوزيع تقدير القدرة في طريقة بييز

واستخدام طريقة المعلومات القصوى في طريقة الأرجحية العظمى. بينت النتائج أن عدد الفقرات المثالية في حالة استخدام طريقة بييز هو (14) فقرة، في حين أن عدد الفقرات المثالي في حالة طريقة الأرجحية العظمى هو (12) فقرة، وأن طريقة بييز هي الأكثر ملاءمة في حالة استخدام النموذج الثلاثي المعلمة من طريقة الأرجحية العظمى. كما لم تظهر أي فروق بين الطريقتين في تقديرات الثبات. وبينت النتائج كذلك أن طريقة الأرجحية العظمى باستخدام أقصى معلومات هي الطريقة الأنسب في الاختبارات التكيفية، وذلك في حالة حجوم العينات الكبيرة. وتوافقت هذه النتائج مع سلسلة من الدراسات قام بها سواميناثان وجيفورد (Swaminathan & Gifford, 1982,1985,1986) حول المقارنة بين فعالية طريقتي الأرجحية العظمى، وبييز من خلال توليد بيانات باستخدام نماذج نظرية الاستجابة للفقرة (الأحادي، والثنائي، والثلاثي) المعلمة، التي أشارت إلى أن طريقة بييز أكثر دقة في تقدير قدرات الأفراد بمعيار الوسط الحسابي لمربعات الفروق بين القيمة الحقيقية والقيمة المقدرة (MSD).

وفي دراسة قام بها وانغ وفيسبول (Wang & Vispoel, 1998) هدفت إلى تقويم النتائج التي يتم الحصول عليها من استخدام أربع طرق في تقدير القدرة وهي: طريقة الأرجحية العظمى، وثلاث طرق مرتبطة بأسلوب بييز هي: طريقة أوينز (Owen's Method)، وتوقع التوزيع البعدي (EAP)، وطريقة تعظيم التوزيع البعدي: (MAP)، إذ تم استخدام أسلوب المحاكاة لاختبار تكيفي، أظهرت النتائج اختلافات واضحة بين طريقة الأرجحية العظمى، وطريقة بييز، من بينها أن طريقة الأرجحية العظمى تعطي أخطاء عالية في تقدير معلمة القدرة مقارنة بطرق بييز، وفيما يتعلق بطرق بييز، فإن طريقة أوينز كانت أسوأ طريقة من حيث دقة التقديرات التي تقدمها.

وفي دراسة أخرى قام بها لندن (Linden, 1998) هدفت إلى فحص طرق تقدير المعتمدة على أسلوب بييز في انتقاء الفقرات في الاختبارات التكيفية، من حيث دقة التقديرات التي تقدمها، إذ اقترحت هذه الدراسة خمس طرق

لتحديد معايير انتقاء الفقرات للاختبارات التكيفية التي تعتمد على التوزيع البعدي الحقيقي (True Posterior)، حيث تم استخدام مجمع فقرات مؤلف من (300) فقرة، وتم أخذ فقرات بأعداد (30، 20، 10، 5)، وتم معايرتها باستخدام النموذج اللوجستي ذي المعلمتين، إذ توصلت الدراسة إلى أن طرق انتقاء الفقرات تبعاً لأسلوب بييز تتفوق على طريقة الأرجحية العظمى في تقدير معلمة القدرة خصوصاً عند استخدام الاختبارات القصيرة، واعتماد معيار متوسط مربعات الأخطاء (Mean-Squared Error) محكاً لدقة التقديرات. وقد أوصت الدراسة بإجراء دراسات مماثلة لكن بتناول النموذج الثلاثي المعلمة.

كما قام وانغ فيزبول (Wang & Vispole,1998) بدراسة هدفت التعرف إلى خصائص طرق تقدير القدرة (MLE, OWEN, MAP, EAP) في الاختبارات التكيفية المحوسبة، حيث تم توليد ثلاثة بنوك من الفقرات المختلفة في التمييز، وتم استخدام طريقة أقصى معلومات في اختيار الفقرات، وبينت النتائج اختلافاً واضحاً بين طريقة MLE، والطرق البييزية، حيث قدمت MLE تحيزاً أقل، وأخطاء معيارية أعلى وتحديداً في الاختبارات القصيرة، وكان الجذر التربيعي لمتوسط مربعات الأخطاء أعلى مع ثقة وكفاءة أقل من الطرق البييزية، أما الطرق البييزية فقد أظهرت تقارباً في الأخطاء المعيارية عند منتصف توزيع القدرة مع فروق بسيطة عند الأطراف، كما بينت النتائج أن طريقة EAP هي الأفضل، وطريقة OWEN هي الأقل دقة في مؤ شرات أخرى.

وفي دراسة مشابهة قام جن ودود (Chen & Dodd, 1998) بدراسة هدفت إلى مقارنة طريقة EAP وطريقة MLE في تقدير القدرة في الاختبارات التكيفية المحوسبة، وذلك باستخدام نموذج التقدير الجزئي (Partial Credit)، حيث تم توليد مجموعتين من البيانات مختلفة في توزيع القدرة. احداهما ملتوية نحو اليسار، والأخرى ذات توزيع طبيعي (500 فرد، 30 فقرة)، واستخدمت طريقة أقصى معلومات لاختيار الفقرات، بحيث يتوقف الاختبار عند خطأ معياري يساوي .0.28 وأظهرت النتائج عدم وجود اختلاف بين طرق

تقدير القدرة ضمن المجموعة نفسها، وبين مجموعتي البيانات، وأن متوسط الأخطاء المعيارية كانت متقاربة جداً.

وفي دراسة قام بها بان وهانسون ووانغ وهاريس وكوينغ (Ban, Hanson, Wang, Qing & Harris, 2000) كان الغرض منها مقارنة وتقويم خمسة طرق لمعايرة فقرات اختبار تكيفي (Computerized Adaptive Testing: CAT) وهي: طريقة الأرجحية العظمى الهامشية لدورة تقدير واحدة (One Estimation Method Cycles)، وطريقة الأرجحية العظمى الهامشية لدورتي تقدير، وطريقة ستوكنغA (Stocking's Method)، وطريقة ستوكنغ B، وطريقة بييز المستخدمة في برنامج Bilog، حيث تم استخدام أحجام عينات مقدارها (300، 1000، 3000). وقد أظهرت نتائج الدراسة أن طريقة الأرجحية العظمى الهامشية لدورتي تقدير تقدم تقديرات أفضل بمعيار الخطأ المعياري في التقدير (Standard Error of Estimation: SE)، تحت الشروط المختلفة لحجم العينة، وكذلك فإن طريقة ستوكنغ B عملت بشكل جيد، ولكنها تحتاج إلى فقرات جذع مشترك (Anchor Items)، مما يؤدي إلى زيادة طول الاختبار، وأما طريقة بييز، فلم تعمل بصورة مناسبة عندما يكون حجم العينة صغيراً.

في حين أشارت نتائج دراسة جيمس ودانييل والين وينغ شن (James, Daniel, Allen, & Young-Shun, 2000) أن كلا من طريقتي الأرجحية العظمى الهامشية، وطريقة ماركوف المرتبطة بأسلوب بييز (Markov Chain Monte Carlo) متشابهة في تقديراتها وجيدة، ولكن التقديرات تكون سيئة إذا كانت الفقرات صعبة جداً أو سهلة جداً، وأن التقديرات تتحسن بزيادة عدد الفقرات من (20-40) فقرة، واتفقت إلى حد ما مع نتائج دراسة قام بها كيم (Kim, 2001) التي هدفت إلى مقارنة دقة التقديرات التي تم الحصول عليها باستخدام طريقة الأرجحية العظمى الشرطية، وطريقة الأرجحية العظمى الهاشمية، وطريقة الأرجحية العظمى المشتركة، وطريقة بييز التي تعتمد على توقع التوزيع البعدي، وذلك من خلال استخدام أسلوب المحاكاة، إذ توصلت

الدراسة إلى أن نتائج تقدير معلمات الفقرة ومعلمة القدرة كانت متشابهة للطرق الأربعة.

كما أجرى وانغ ووانغ (Wang & Wang, 2001) دراسة لتقويم الفاعلية النسبية لطريقة WLE، ومقارنتها بطرق التقدير MAP, EAP, MLE، وفق نموذج التقدير الجزئي المعمم، وتم استخدام طريقة أقصى معلومات في اختيار الفقرات. أظهرت نتائج الدراسة أن طريقة WLE هي أكثر دقة من طريقة MLE، وذلك عندما يكون الاختبار ثابتاً في طوله، وأن طريقتيWLE, MLE لها تحيز أقل وخطأ معياري أكبر من طريقتي MAP, EAP. ولم تظهر أي فروق ذات دلالة إحصائية في متوسطات تقديرات القدرة، ومن حيث الفاعلية تميزت طريقتي EAP, MAP عن الطرق الأخرى. في حين أظهرت نتائج دراسة ليوي ونيلسون (Lewy & Nielsen, 2002) أن تقديرات بييز تتمتع بثبات أفضل من طريقة الأرجحية العظمى؛ لأنها أقل تحيزا، وأقل تباين خطأ، وذلك بالاعتماد على محك (MSE) مربعات الأخطاء المعيارية.

وفي دراسة أخرى قام بها وانغ ووانغ (Wang & Wang, 2002) هدفت إلى المقارنة بين طريقة WLE ومقارنتها بطرق التقدير MAP, EAP, MLE، وفق نموذج التقدير الجزئي المعمم (GPCM,Generalized Partial Credit Model)، ونموذج الاستجابات المتدرجة (Graded Response Model,GRM). بينت النتائج أن طريقة WLE لها أقل تحيز وأقل خطأ معياري على تقدير القدرة في النموذجين GPCM, GRM، وأن طريقتي MLE, WLE أظهرتا تحيزاً أقل من طريقتي بييز وللنموذجين. كما أظهرت الطريقتان MLE, WLE تحيزاً أقل من طريقتي بييز، وذلك عند مستويات القدرة المتطرفة، وأن لهما أقل خطأ معياري لمتوسط المربعات RMSE، وأن الخطأ المعياري والتحيز يقل في جميع الطرق المستخدمة بزيادة طول الاختبار وثباته، وبزيادة عدد المتغيرات المستقلة الاختلافات.

أما دراسة جاو وجن (Gao & Chen, 2005) فقد هدفت إلى المقارنة بين

طريقتي الأرجحية العظمى الهامشية، وطريقة بييز الهامشية، وذلك باستخدام أسلوب تعظيم الاقتران البعدي (MAP)، في دقة التقدير لمعالم الفقرات، وذلك باستخدام النموذج الثلاثي المعلمة. ولتحقيق ذلك الهدف، تم توليد بيانات بأحجام (100، 500، 2000)، وفقرات عددها(10، 30، 60) فقرة. وباستخدام محك (RMSD) (Root of Mean Squared Error Difference) أظهرت النتائج أن دقة تقديرات الأرجحية العظمى وطريقة بييز لأحجام العينات (500، 2000) كانت متشابهة، كذلك عند اختلاف عدد الفقرات، إلا أن طريقة بييز كانت أفضل في تقديراتها عند حجم العينة (100). وكانت دراسة أثر التوزيعات القبلية المختلفة في تقديرات بييز هي الأهم في تلك الدراسة، حيث تمت دراسة أربعة توزيعات مختلفة؛ لفحص مواطن القوة والضعف في تقديرات بييز بالنسبة للتغير في التوزيعات الحقيقية والقبلية، حيث أظهرت النتائج كذلك أن تقديرات بييز تتأثر بشكل بسيط إذا كان حجم عينة المفحوصين ليس صغيراً.

ومن جهة أخرى قام جلاس (Glass, 2005) بدراسة هدفت إلى معرفة أثر حجم العينة وعدد الفقرات في دقة تقدير معلم قدرة الفرد وفق طريقة بييز التي تعتمد أسلوب تعظيم الاقتران، وذلك من خلال توليد بيانات ثنائية التدريج بأحجام عينات مقدارها (500، 1000، 2000) وفق النموذج الثلاثي المعلمة، وفقرات بعدد (200، 440) فقرة تم توليدها من بنك للأسئلة، وقد تم اعتماد الوسط الحسابي للأخطاء المعيارية لمعلم قدرة الفرد عند مستويات مختارة من القدرة موزعة على أطراف متصل القدرة (2، 1، 0، 1-،2-) لوجيت، عند أحجام العينات المختلفة والفقرات. وقد أظهرت نتائج الدراسة أنه بزيادة عدد الفقرات عند أحجام العينات (500، 2000) تقل الأخطاء المعيارية وهذا بدوره يؤدي إلى الزيادة في دقة التقدير. أما عند زيادة عدد الفقرات عند حجم العينة (1000) تقل دقة التقدير.

وعلى صعيد آخر اهتمت دراسة ميشيل وحافظ ومارك (Micheal, Hafidz, & Mark, 2006) بمقارنة نتائج المعادلة العمودية للتقديرات المقدرة

باستخدام نموذج راش المقدرة وفق طريقة الأرجحية العظمى(MLE) المستخرجة من برنامج (Bilog-Mg$_3$)، والعظمى المشتركة (JML) المستخرجة من برنامج (Winsteps)، بحيث تمت مقارنة معالم الصعوبة للفقرة، وتقديرات القدرة للأفراد من خلال توليد بيانات ثنائية التدريج اعتماداً على توزيعين لقدرة الفرد (طبيعي، ملتو التواءاً سالباً)، وقد تم اعتماد الوسط الحسابي، والوسيط، والانحراف المعياري، وحجم الأثر كمعيار لدقة التقدير. أظهرت نتائج الدراسة أن طريقة الأرجحية العظمى تتفوق على طريقة الأرجحية العظمى المشتركة في دقة التقديرات لمعالم الفقرات.

كما اتفقت مع نتائج الأبحاث التي أجراها جير وفيرمنت (Garee & Vermunt, 2006) المتعلقة بدقة تقدير المعالم، وذلك من خلال استخدام توزيعات قبلية مختلفة لمعلمة القدرة، بحيث تم التحقق الامبريقي من نتائج الاشتقاقات الرياضية الخاصة بطريقتي الأرجحية العظمى وبييز بأسلوب تعظيم الاقتران، من خلال تطبيق بيانات حقيقية بأحجام عينات مقدارها (100، 1000)، وفقرات بعدد (9،5)، دلت نتائج الدراسة أن تقديرات المعالم والأخطاء المعيارية التي نحصل عليها خلال طريقة بييز، تتمتع بثبات أفضل من تقديرات المعالم والأخطاء المعيارية التي يتم الحصول عليها من خلال طريقة الأرجحية العظمى، خاصة عند تقدير القدرات على أطراف متصل القدرة، كذلك عند حجم العينة (1000)، وعدد الفقرات (9)؛ أي بزيادة عدد الفقرات وحجم العينة. وأكدت ذلك دراسة روندال (Rondall, 2007) بأن الأخطاء المعيارية لتقديرات القدرة للأفراد المقدرة وفق طريقة بييز هي أقل؛ أي أنها أكثر فاعلية من التقديرات الناتجة وفق الأرجحية العظمى.

ومن الدراسات العربية ذات الصلة بموضوع البحث، دراسة الدرابيع (2001) التي بحثت في فعالية النموذج اللوغاريتمي ذي المعلمة الواحدة " نموذج راش" في دقة تقدير قدرة الفرد، ومعامل صعوبة الفقرة باختلاف حجم العينة وطول الاختبار. أشارت نتائج الدراسة إلى وجود فروق ذات دلالة إحصائية

لتفاعل كل من حجم العينة وطول الاختبار على دقة تقدير قدرة الفرد، وإلى وجود فروق ذات دلالة إحصائية في دقة تقدير قدرة الفرد ترجع إلى طول الاختبار وحده، في حين لم تظهر النتائج وجود فروق ذات دلالة إحصائية لحجم العينة. وفيما يتعلق بدقة تقدير معامل صعوبة الفقرة، فقد أشارت النتائج إلى وجود فروق ذات دلالة إحصائية لتفاعل كل من حجم العينة، وطول الاختبار، وكذلك وجود فروق ذات دلالة لكل منهما على حدة.

وفي دراسة قام بها العبابنه (2004) هدفت إلى استقصاء أثر حجم العينة، وطريقة انتقائها، وعدد الفقرات، وطريقة انتقائها على دقة التقديرات لمعالم الفقرات، من خلال اختبار قدرة عقلية باستخدام نظرية الاستجابة للفقرة، حيث تم بناء مقياس قدرة عقلية مكون من 71 فقرة. طبقت على عينة مكونة من 1000 مفحوص، وتم استخدام النموذج اللوجستي ثلاثي المعلمة في معايرة فقرات الاختبار. وكان من أبرز النتائج التي تم التوصل إليها أن الدقة في تقدير معلمة التمييز تزداد بزيادة تباين قدرات المفحوصين، كما تزداد الدقة في تقدير معلمتي القدرة والصعوبة عندما يكون مدى القدرة متوافقاً مع مدى الصعوبة للفقرات، وكذلك تزداد الدقة في تقدير معلمة التخمين عند استخدام عينة من ذوي القدرات المتدنية في معايرة الفقرات.

كما قام مسلم (2010) بدراسة هدفت إلى اقتراح أسلوب في اختيار الفقرات، وتقدير القدرة في الاختبارات التكيفية المحوسبة وفحص فاعليته. ولتحقيق أغراض الدراسة تم توليد قدرات 2000 فرد من خلال توزيع طبيعي. وبالاعتماد على معلم القدرة، تم توليد بنك فقرات مكون من 500 فقرة تلائم النموذج اللوجستي ثلاثي المعلمة. وتم الحصول على قيم القدرة لكل مفحوص باستخدام الطرق MLE, EAP, MAP، بينت نتائج الدراسة أن معاملات الارتباط بين القدرات المقدرة وفق الطريقة المقترحة وطرق التقدير آنفة الذكر، كانت موجبة وعالية وذات دلالة إحصائية، وأشارت كذلك إلى أن قيم التحيز والجذر التربيعي لمتوسط مربعات الأخطاء المعيارية لتقديرات القدرة كانت

متقاربة مع الطرق المذكورة. كما بينت النتائج أن مؤشرات الدقة في القياس من تحيز، وخطأ معياري للتقدير، والجذر التربيعي لمتوسط المربعات للأخطاء للطريقة المقترحة كانت قريبة من تقديرات طريقة الأرجحية العظمى، والطرق البييزية عندما يكون عدد الفقرات قليلاً، وأنها قد تميزت في التحيز على الأرجحية العظمى وبخاصة عند مستويات القدرة العالية والمتوسطة، بينما تميزت الطرق الأخرى في حالة زيادة عدد الفقرات.

وأجرى بني عواد (2010) دراسة هدفت إلى مقارنة طرق التعامل مع البيانات المفقودة في تقدير معالم الفقرات وقدرات الأفراد، حيث تمت معالجة البيانات المفقودة من خلال استخدام ثماني طرق مختلفة. ولتحقيق أغراض الدراسة، تم تطبيق اختبار للقدرة العقلية على عينة مكونة من 1600 طالباً وطالبة من طلبة الصف الثامن في مدارس مديرية التربية والتعليم لمنطقة اربد الأولى. وأظهرت النتائج أن الطرق التعويضية تؤثر على معالم الفقرة وقدرات الأفراد باختلاف طول الاختبار، كما أظهرت النتائج وجود فروق ذات دلالة إحصائية بين المتوسطات الحسابية للأخطاء المعيارية لتقدير معلمة القدرة للأفراد، وكذلك معالم الصعوبة والتمييز والتخمين تعزى للطريقة، وطول الاختبار، والتفاعل بينهما.

في ضوء ما تقدم عرضه من الدراسات السابقة يمكن ملاحظة: أن معظم الدراسات الأجنبية التي تم القيام بها قد تمت من خلال بيانات محاكاة باستخدام البرامج الإحصائية، وليست باستخدام بيانات فعلية، وأن معظمها قد استخدم الإحصاءات الوصفية، ولم تستخدم أساليب الإحصاء الاستدلالي، كما أن تلك الدراسات لم تحسم الجدل حول الطريقة الفضلى في تقدير معالم الفقرات وتقديرات القدرة للأفراد، وطريقة تقدير الثبات، والفعالية النسبية، ولا يوجد اتفاق بين تلك الدراسات حول هذه النقاط تحديداً، ولم تتطرق معظمها إلى أثر حجم العينة كمتغير يؤثر في تلك التقديرات، وبالتالي لم تقدم إجابة حول أثر التفاعل بين طريقة التقدير لمعالم الفقرات وتقديرات القدرة

للأفراد، كما أنه - في حدود علم الباحث - لا توجد دراسات عربية حول الموضوع، لذا ونظرا لقلة الدراسات العربية التي عالجت مثل هذا النوع من المشكلات البحثية، فإنه يؤمل من هذه الدراسة أن تضيف معلومات جديدة حول الطريقة الأكفأ في تقدير المعالم المختلفة باختلاف حجم العينة، حيث إن الباحث لم يجد دراسة تناولت أثر طريقة تقدير معالم الفقرات، وقدرات الأفراد، وتفاعلها مع حجم العينة في خصائص الاختبار وخصائص فقراته السيكومترية، لذا جاءت هذه الدراسة لتسهم في تقديم إضافة معرفية جديدة حول هذا الموضوع.

أما عن سبب اختيار الباحث النموذج اللوجستي ثلاثي المعلمة لتحقيق غرض هذه الدراسة، فهو أن هذا النموذج يعد من أكثر النماذج واقعية، كما أشارت لذلك بعض الدراسات المشار إليها آنفاً، وكذلك دراسة هامبلتون وتروب (Hambleton & Traub, 1971) التي بينت أن النموذج الثلاثي المعلمة هو الطريقة الفضلى في تقدير القدرة للفرد؛ نظراً لاحتوائه على معلمة التخمين الخاصة بالأفراد ذوي القدرة المتدنية، كما أن هذا النموذج يشتمل إضافة إلى المعالم التي يشتملها النموذج ذو المعلمتين على معلمة التخمين للفقرة، لتقدير احتمالية أن يجيب الفرد إجابة صحيحة عليها، كما أن هذا النموذج لم يحظ باهتمام كبير في الدراسات العربية في هذا المجال، إلا أن بعض الباحثين أشاروا بأن النموذج ثلاثي المعلمة مقارنة بالنماذج الأخرى يتطلب عددا كبيراً من المفحوصين، وعدداً كبيراً من فقرات الاختبار، فمثلاً أشارت دراسة هامبلتون وتروب(Hambleton & Traub, 1973)، إلى أن دقة التقدير لقدرة الفرد تزداد بزيادة عدد الفقرات، متفقة بذلك مع دراسة سوامينثان وجفورد(Swaminathan & Gifford, 1979) بأن دقة التقدير لقدرة الفرد تزداد بزيادة طول الاختبار، وكذلك اتفقت مع دراسة هامبلتون وكوك(Hambleton & Cook, 1980) التي توصلت إلى أنه بزيادة حجم العينة، وعدد الفقرات، تزداد دقة التقدير لقدرة الفرد باستخدام النموذج الثلاثي المعلمة. لذا، وفي ضوء ما تقدم فقد ارتأى

الباحث استخدام النموذج اللوجستي الثلاثي المعلمة لتحقيق الدقة في أهداف الدراسة.

أهداف الدراسة

سعت الدراسة إلى تحقيق الأهداف الآتية:

* الكشف عن أثر طريقة التقدير (طريقة الأرجحية العظمى، وطريقة بييز باستخدام أسلوب توقع الاقتران البعدي، وطريقة بييز باستخدام أسلوب تعظيم الاقتران البعدي) على دقة تقدير معالم الفقرات في ضوء تغير حجم العينة.
* الكشف عن أثر طريقة التقدير (طريقة الأرجحية العظمى، وطريقة بييز باستخدام أسلوب توقع الاقتران البعدي، وطريقة بييز باستخدام أسلوب تعظيم الاقتران البعدي) على دقة تقدير القدرة للأفراد في ضوء تغير حجم العينة.
* الكشف عن أثر طريقة التقدير (طريقة الأرجحية العظمى، وطريقة بييز باستخدام أسلوب توقع الاقتران البعدي، وطريقة بييز باستخدام أسلوب تعظيم الاقتران البعدي) على ثبات الاختبار في ضوء تغير حجم العينة.
* الكشف عن مدى اختلاف الفعالية النسبية باختلاف طريقة التقدير (طريقة الأرجحية العظمى، وطريقة بييز باستخدام أسلوب توقع الاقتران البعدي، وطريقة بييز باستخدام أسلوب تعظيم الاقتران البعدي) في ضوء تغير حجم العينة.

مشكلة الدراسة

إن عدد مطوري الاختبارات الذين يستخدمون نماذج نظرية الاستجابة للفقرة يزداد بشكل مطرد، حيث إن مطوري الاختبارات يعودون بشكل متزايد إلى نظرية الاستجابة للفقرة لمعالجة مشكلات تتعلق ببناء الاختبارات، واختيار

الفقرات، والكشف عن تحيز الفقرات، والقياس التكيفي، ومعادلة درجات الاختبار ذلك أن النظرية الكلاسيكية في القياس تعاني من جوانب قصور أشار إليها هامبلتون وسوامينثان (Hambelton & Swaminthan, 1985)، ومنها: اختلاف معالم الفقرات وهي الصعوبة، والتمييز، والتخمين باختلاف عينة الأفراد المستخدمة في استخراج هذه المعالم، كما أن النظرية الكلاسيكية لا تزودنا بتقدير دقيق لقدرات الأفراد ذوي القدرات العالية، والأفراد ذوي القدرات المتدنية، وأن الأداء على الفقرة أو مجموعة جزئية من الفقرات في النظرية الكلاسيكية لا يقدم معلومات كافية عن مستوى أداء الفرد، ناهيك عن أن مفهوم الثبات في النظرية الكلاسيكية يقوم على أساس إمكانية تطوير صور متوازية، والذي يصعب تحقيقه في الواقع. لذا ظهرت نظرية الاستجابة للفقرة كنظرية بديلة، نظراً للمزايا العديدة التي تتمتع بها إضافة إلى الافتراضات القوية التي اعتمدت عليها، والتي يفترض من خلالها تلافي عيوب النظرية الكلاسيكية.

وفي الحقيقة فإن هناك مشكلة تظهر عند استخدام نظرية الاستجابة للفقرة تتعلق بدقة تقدير معالم الفقرة والقدرة للأفراد، حيث تفترض هذه النظرية أن إجابة الأفراد على الفقرات تتحدد بقدرة عقلية غير ملاحظة تدعى السمة الكامنة، وهناك عوامل تؤثر على استقرار ودقة التقدير لتلك المعالم؛ وبمعنى آخر فان هناك عوامل تهم القائمين ببناء الاختبارات والمقاييس، وتساعدهم في انتقاء فقرات اختبارات تتميز بالدقة وتقدم معلومات أفضل عن الأفراد، ومن تلك العوامل طريقة التقدير المستخدمة في تقدير المعالم، وحجم العينة والتفاعل بينهما. لذا تعد عملية تقدير معالم الفقرات والأفراد وفق نماذج نظرية الاستجابة للفقرة أحادية البعد من الخطوات الأساسية والهامة لتطبيق هذه النظرية، كما أن استخدام الطريقة المناسبة لعملية التقدير يتوقف على عوامل عدة، من أهمها حجم العينة، وعدد الفقرات، وتوزيع قدرات الأفراد في عينة الدراسة. إذ تباينت وجهات النظر حول استخدام الطريقة الفضلى ما بين مؤيد ومعارض. فمنهم من يرى أن التقديرات التي يتم الحصول عليها

متشابهة باختلاف الطريقة المستخدمة، وخاصة عندما يكون التوزيع منتظماً، ومنهم من يرى أنها تختلف باختلاف حجم العينة، أو عدد الفقرات، أو توزيع قدرات الأفراد. كذلك، فإن الدراسات السابقة استندت في تقييم نتائجها باستخدام الطرق الإحصائية الوصفية التي تقدم بيانات وصفية فقط مثل الأوساط الحسابية للأخطاء المعيارية، والانحراف المعياري، ومربعات الأخطاء المعيارية، والفعالية النسبية، لذا فقد بات من الضروري أن توظف هذه البيانات من أجل اختبار فرضيات بحثية؛ لمعرفة دلالة الفروق للخصائص السيكومترية لكل من الفقرات، والأفراد، والاختبار باختلاف العوامل التطبيقية للاختبار كاختلاف حجم العينة، كما أن معظم تلك الدراسات استخدمت أسلوب المحاكاة في توليد البيانات لإجراء تلك الدراسات، وكان من توصياتها ضرورة استخدام بيانات فعلية وتصاميم تجريبية. ومن هنا تنشأ الحاجة إلى مثل هذه الدراسة التي تهدف إلى معرفة أثر طريقة التقدير لمعالم الفقرات والأفراد على الخصائص السيكومترية للاختبار، وخصائص فقراته السيكومترية باختلاف حجم العينة، من أجل تقديم أدلة إمبريقية، ومن خلال بيانات فعلية عن أدق الطرق المستخدمة، التي تساهم بها طريقتي الأرجحية العظمى، وطريقة بييز باستخدام أسلوبي تعظيم الاقتران البعدي، وتوقع الاقتران البعدي، وذلك باختلاف حجم العينة. في تقدير معالم الفقرات والأفراد وباستخدام النموذج اللوجستي الثلاثي المعلمة الذي يعد من أكثر النماذج واقعية. وبالتحديد فإن هذه الدراسة هدفت للإجابة عن الأسئلة الآتية:

١ - ما أثر طريقة التقدير (طريقة الأرجحية العظمى، وطريقة بييز باستخدام أسلوب توقع الاقتران البعدي، وطريقة بييز باستخدام أسلوب تعظيم الاقتران البعدي) وحجم العينة على دقة تقدير معالم الفقرات؟

٢ - ما أثر طريقة التقدير (طريقة الأرجحية العظمى، وطريقة بييز باستخدام أسلوب توقع الاقتران البعدي، وطريقة بييز باستخدام أسلوب تعظيم الاقتران البعدي) وحجم العينة على دقة تقدير القدرة للأفراد؟

٣ - ما أثر طريقة التقدير (طريقة الأرجحية العظمى، وطريقة بييز باستخدام أسلوب توقع الاقتران البعدي، وطريقة بييز باستخدام أسلوب تعظيم الاقتران البعدي) وحجم العينة على ثبات الاختبار ؟

٤ - هل تختلف الفعالية النسبية باختلاف طريقة التقدير (طريقة الأرجحية العظمى، وطريقة بييز باستخدام أسلوب توقع الاقتران البعدي، وطريقة بييز باستخدام أسلوب تعظيم الاقتران البعدي) وحجم العينة؟

أهمية الدراسة

تتمثل أهمية هذه الدراسة في جانبين: الأول نظري يتوقع من خلال هذه الدراسة أن تكشف عن الطريقة الفضلى في تقدير معالم الفقرات والأفراد في ظروف مختلفة، ومن خلال بيانات فعلية وليست بيانات محاكاة. والجانب الثاني تطبيقي يتوقع أن توفر هذه الدراسة أدلة إمبريقية من خلال أساليب الإحصاء الاستدلالي؛ بمعنى محاولة التوصل إلى تبريرات عملية تقدم إلى الباحثين يمكن اعتمادها بناءً على أساس تجريبي، ومن خلال بيانات فعلية حول الطريقة الفضلى التي يمكن استخدامها من قبل الباحثين في دراسات أخرى لتقدير معالم الفقرات وقدرات الأفراد عند استخدام نظرية الاستجابة للفقرة، كما يمكن أن يستفيد مطورو الاختبارات الذين يلجأون إلى نظرية الاستجابة للفقرة لحل المشكلات المتعلقة ببناء الاختبارات لاختيار الطريقة الأنسب عند تقدير معالم الفقرات وهي الصعوبة، والتمييز، والتخمين، وفق النموذج الثلاثي المعلمة.

تعريف المصطلحات

طريقة الأرجحية العظمى: طريقة تقدير معالم تقديرات القدرة للأفراد من خلال إجراءات تعظيم الاحتمالية للمعلمة المراد تقديرها.

طريقة بييز: طريقة تقدير معالم قدرات الأفراد، عن طريق استخدام معلومات أولية عن العينة متوافرة من خلال خبرات سابقة. وتتم وفق طريقتين هما: طريقة بييز باستخدام أسلوب توقع الاقتران البعدي، وطريقة بييز

باستخدام أسلوب تعظيم الاقتران البعدي. هذا وقد سبق أن تم تقديم توضيح تفصيلي لكل من هاتين الطريقتين في مقدمة البحث.

معالم الفقرات: وهي معالم الصعوبة، والتمييز، والتخمين، وفق النموذج الثلاثي المعلمة.

معالم الأفراد: وهي معالم قدرات الأفراد، وهي مقدار ما يمتلكه الفرد من السمة التي يتم قياسها تبعاً لنظرية الاستجابة للفقرة.

تقدير المعالم للفقرات والأفراد: هي عملية التعبير الكمي عن المعالم.

دقة التقدير: هو تعبير يشير لجودة التقدير التي يميزها الاحتمالية الكبيرة في أن التقدير قريب من القيمة الحقيقية، وذلك باستخدام معدل مربعات الأخطاء أو الأخطاء المعيارية في التقدير.

الفعالية النسبية للاختبار: هي نسبة دالة المعلومات للاختبار إلى دالة المعلومات لاختبار آخر عند مستوى القدرة نفسه.

حدود الدراسة

- اقتصرت عينة الدراسة على طلبة الصف الثاني الثانوي العلمي في مديرية التربية والتعليم لمنطقة إربد الأولى.
- اقتصرت الدراسة على الفصلين الأول والثاني من وحدة الكهرباء والمغناطيسية لمبحث الفيزياء للصف الثاني الثانوي العلمي في الفصل الثاني للعام الدراسي ٢٠٠٩/٢٠١٠م.
- نتائج الدراسة تتحدد بمدى صدق الاختبار وثباته الذي أعده الباحث.

منهجية الدراسة وإجراءاتها

مجتمع الدراسة وعينتها: تكون مجتمع الدراسة من طلاب وطالبات الصف الثاني الثانوي العلمي للعام الدراسي 2009–2010 التابعين لمدارس مديرية التربية والتعليم لمنطقة إربد الأولى، ويبلغ عددهم (10980) طالباً وطالبة. تكونت

عينة الدراسة من (1012) طالباً وطالبة موزعين على (23) شعبة، اختيروا بالطريقة العشوائية العنقودية، حيث كانت وحدة الاختيار هي المجموعة. فقد تم حصر عدد المدارس التي يتواجد فيها طلبة الصف الثاني الثانوي العلمي من الذكور والإناث، ومن هذه المدارس تم اختيار ست مدارس من الذكور، وخمس مدارس للإناث باستخدام الطريقة العشوائية، بما يتناسب وعدد المدارس.

أداة الدراسة: تـم بناء اختبار تحصيلي في مبحث الفيزياء لطلبة الصف الثاني الثانوي العلمي في موضوعات الكهرباء السكونية، والتيار الكهربائي، والدارات الكهربائية من الوحدة الأولى، من نوع الاختيار من متعدد من أربعة بدائل. وقد اقتضى بناء أسئلة الاختبار ووضعه في الصيغة النهائية الاسترشاد بالأسس العامة المتبعة في بناء الاختبارات التحصيلية التي أوردها (Gronlund & Linn,1990) كما يلي:

أولا: تحديد الغرض من الاختبار، وهو قياس تحصيل طلبة الصف الثاني الثانوي العلمي في موضوعي الكهرباء السكونية، والتيار الكهربائي، والدارات الكهربائية من الوحدة الأولى في كتاب الفيزياء.

ثانياً: تحديد الموضوعات الداخلة في الاختبار، وذلك من خلال تحليل المحتوى لموضوعي الكهرباء السكونية والتيار الكهربائي، والدارات الكهربائية من الوحدة الأولى في كتاب الفيزياء، ثم صياغة النتـاجات التعليمية (أهداف المحتوى) موضوع الاختبار، بعد ذلك قام الباحث بإعداد لائحة (جدول مواصفات) للاختبار تم فيه ربط مستويات النتاجات التعليمية بمحتوى المادة الدراسية موضوع الاختبار.

ثالثاً: صياغة فقرات الاختبار من نوع الاختيار من متعدد، لكل فقرة منها أربعة بدائل تمثل احداها الإجابة الصحيحة، وبلغ عدد الفقرات بصورتها الأولية (50) فقرة؛ لقياس كل هدف من أهداف المحتوى، وقد تطلب قياس بعضها أكثر من فقرة واحدة، وروعيت فيها الأسس الفنية في كتابة هذا النوع من الفقرات، ومطابقتها للهدف الذي يقيسه من حيث المحتوى

والمستوى المعرفي، كما روعي ألّا تعتمد إجابة إحدى هذه الفقرات على إجابة الفقرات الأخرى، وهذا ما تتطلبه نظرية الاستجابة للفقرة.

رابعاً: التحقق من صدق المحتوى، بعرض الصورة الأولية للاختبار، وجدول المواصفات، وتحليل المحتوى، والنتاجات التعليمية (أهداف المحتوى)، على مجموعة من المتخصصين (المحكمين) من أصحاب الخبرة والكفاية؛ من أجل أخذ وجهات نظرهم في صدق الفقرة الاختبارية في قياس الهدف المحدد، ومدى تمثيل عينة الفقرات لأهداف المحتوى، وكذلك لمستوى الأهداف وفق جدول المواصفات المعد لهذه الغاية، واقتراح ما يرونه مناسباً من خلال استبانة معدة لهذا الغرض. وبناءً على آراء المحكمين تم حذف بعض الفقرات التي أجمع المحكمون أن هناك فقرات تؤدي غرضها، أو أن نسبة الاتفاق حولها تقل عن 80٪، وبلغ عددها (9) فقرات. وتم الأخذ بالاقتراحات والملاحظات القيَمة التي قدمها المحكمون التي ساعدت في إعادة النظر في بعض الأهداف، ومراجعة عدد من الفقرات التي تم تغيير الكلمات الغامضة فيها وتبديلها، بحيث يكون لها مدلول واضح؛ بالإضافة إلى تصويب الأخطاء اللغوية والنحوية والإملائية والمطبعية فيها، كما تم تغيير عدد من البدائل لبعض الفقرات التي أشار بعض المحكمون أنها بدائل ضعيفة، وبعد ذلك تم إعادة كتابة فقرات الاختبار، وبلغ عددها في صورتها بعد تحكيمها (41) فقرة.

خامساً: التجريب الأولي للاختبار: تم تطبيق الاختبار بصورته الأولية على عينة استطلاعية مكونة من (50) طالباً وطالبة من طلبة المدرسة النموذجية في جامعة اليرموك موزعين على شعبتين؛ شعبة للذكور وشعبة للإناث تم اختيارها بشكل عشوائي من شعب الصف الثاني الثانوي العلمي في المدرسة؛ وذلك بهدف تحديد الزمن الكافي للأداء على الاختبار، والتأكد من وضوح الصياغة اللغوية للفقرات، ومفاتيح الإجابة، والكشف على الفقرات الغامضة، أوالعبارات التي كثر التساؤل حولها، وبالتالي تعديل الفقرات أو حذفها في ضوء ذلك. وتم تصحيح أوراق الاختبار يدوياً، وإجراء عملية التقدير لبيانات التجريب الأولي باستخدام برنامج

التحليل الإحصائي (SPSS)، فقد تم حساب معامل الصعوبة لكل فقرة من خلال إيجاد نسبة الطلبة الذين أجابوا عن الفقرة إجابة صحيحة من بين الطلبة الذين حاولوا الإجابة عن هذه الفقرة (عوده، 2010)، كما تم حساب معاملات التمييز للفقرات، عن طريق حساب معامل الارتباط بين الدرجات على الفقر ة، والدرجات على الاختبار ككل؛ أي بعد حذف الفقرة نفسها من الاختبار (Corrected item Total Correlation). والجدول رقم (1) يبين معاملات الصعوبة، ومعاملات التمييز لفقرات الاختبار للعينة الاستطلاعية.

الجدول رقم (١)

معاملات الصعوبة ومعاملات التمييز لفقرات الاختبار للعينة الاستطلاعية

رقم الفقرة	معامل الصعوبة	معامل التمييز	رقم الفقرة	معامل الصعوبة	معامل التمييز	رقم الفقرة	معامل الصعوبة	معامل التمييز
١	٠,٥١	٠,٦٢	١٥	٠,٨٣	٠,٣٠	٢٩	٠,٢٧	٠,٣٦
٢	٠,٥١	٠,٧٣	١٦	٠,٤٢	٠,٣٠	٣٠	٠,٣٣	٠,٤٤
٣	٠,٤٨	٠,٥٦	١٧	٠,٣٦	٠,٥٨	٣١	٠,٣٣	-٠,٣١
٤	٠,٤٩	٠,٦٢	١٨	٠,٨٧	٠,٦١	٣٢	٠,٣٩	٠,٢٥
٥	٠,٥٩	٠,٧١	١٩	٠,٧٨	٠,٣٠	٣٣	٠,٤٥	٠,٣١
٦	٠,٢٥	٠,٣٣	٢٠	٠,٥٤	٠,٦٠	٣٤	٠,٣١	٠,٤٦
٧	٠,٥٩	٠,٧٠	٢١	٠,٤٨	٠,٦٧	٣٥	٠,٤٦	٠,٤٤
٨	٠,٤٧	٠,٥٤	٢٢	٠,٣٦	٠,٣٤	٣٦	٠,٣٨	٠,٤٦
٩	٠,٧٢	-٠,٢٠	٢٣	٠,٥٢	٠,٥٥	٣٧	٠,٢٠	٠,٢٨
١٠	٠,٥٦	-٠,١٩	٢٤	٠,٦٦	٠,٥١	٣٨	٠,٣٣	٠,٤٩
١١	٠,٥٧	٠,٣٩	٢٥	٠,٣١	٠,٣٥	٣٩	٠,٤٩	٠,٥٢
١٢	٠,٦١	٠,١٧	٢٦	٠,٥٧	٠,٣٤	٤٠	٠,٤٣	٠,٣٤
١٣	٠,٦٠	٠,٥٣	٢٧	٠,٣١	٠,١٨	٤١	٠,٤٤	٠,٦٤
١٤	٠,٤٤	٠,٦١	٢٨	٠,٣١	٠,٤٦			

يتضح من الجدول رقم (١) أن معاملات الصعوبة في الاختبار تراوحت

بين (٠,٢٠-٠,٨٧) بوسط قدره (٠,٤٨)، أما معاملات التمييز فقد تراوحت بين (-٠,٣١ و٠,٧٣) بوسط قدره (٠,٤١)، وتم حذف الفقرات (٩،١٠،٢٧،٣١)؛ وذلك نظراً لتدني معـاملات التمييز لها، حيث كانت معاملات التمييز لهذه الفقرات (-٠,٢٠، -٠,١٩، ٠,١٨، -٠,٣١) على التوالي. أما معاملات الصعوبة لهذه الفقرات فكانت على التوالي (٠,٧٢، ٠,٥٣، ٠,٣١، ٠,٣٣)، ويلاحظ أن جميعها مقبولة. كما تم حذف الفقرتين (١٨، ١٥) نظراً لارتفاع معامل الصعوبة، حيث كانت معاملات الصعوبة لهما على التوالي (٠,٨٣، ٠,٨٧)؛ بمعنى أنه تم حذف جميع الفقرات التي صعوبتها أقل من ٠,٢٠ وكذلك التي زادت صعوبتها على ٠,٨٠ (عوده، ٢٠١٠)، وبعد حذف هذه الفقرات أصبح عدد الفقرات (٣٥) فقرة. وتم حساب معامل ارتباط بيرسون بين علامات الطلبة في العينة الاستطلاعية على الفقرات المتبقية، ومعدل الطالب في مبحث الفيزياء في الصف الأول الثانوي؛ وذلك للتأكد من الصدق المرتبط بمحك، وبلغت (٠,٨٨)، كما تم حساب معامل كرونباخ ألفا التي بلغت (٠,٩٢). وهما قيمتان مرتفعتان ومناسبتان لإجراء مثل هذه الدراسة.

سادساً: إعداد ورقة تعليمات للاختبار وورقة الإجابة النموذجية.

سابعاً: إجراءات جمع البيانات: تم تطبيق الاختبار على كل أفراد عينة الدراسة، ثم تم جمع استجابات الطلبة على الاختبار، وتمت مراجعة أوراق الطلبة للكشف عن الاستجابات ذات النمطية الواحدة، كذلك تم رصد أعداد الغائبين عن الاختبار - لظروف صحية -، وكان عددهم (١٢) طالب وطالبة.

إجراءات الدراسة

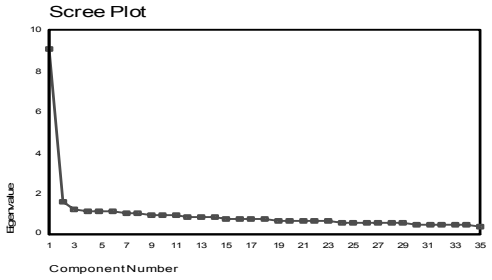
١ - تم بناء الاختبار التحصيلي المكون من (٣٥) فقرة، ليصبح بصورته النهائية بعد المعايرة (٣٣) فقرة.

٢ - تم تحديد المدارس المشمولة في العينة.

٣ - تم التحدث مع طلبة الشعب التي وقع عليها الاختيار حول الدراسة، وأهميتها، وتعليمات الإجابة على الاختبار، وضرورة الحرص والجدية في الإجابة.

٤ - تم تطبيق الاختبار، وجمع استجابات الطلبة على الاختبار، وبلغ عددها (١٠٠٠) ورقة.

٥. تم التحقق من افتراضات النموذج ثلاثي المعلمة، حيث تم إدخال البيانات إلى برنامج (Bilog-Mg$_3$) وللتحقق من افتراض أحادية البعد: تم استخدام التحليل العاملي (Factor Analysis) باستخدام أسلوب المكونات الأساسية (Principal Component Analysis)، وذلك لاستجابات الطلبة على الاختبار. وتم إيجاد قيم الجذور الكامنة (EigenValue)، ونسب التباين المفسّر (Explained Variance)، لكل عامل من العوامل، وجرى التدوير للعوامل باستخدام طريقة التدوير المتعامد (Varimax Rotation)، التي كانت قيمة الجذر الكامن لها أكبر من واحد، إذ بلغ حاصل قسمة الجذر الكامن للعامل الأول إلى العامل الثاني (٤,٠٥٩)، ونسبة التباين المفسر للعامل الأول ٢٠,٦٨٪. باعتماد قيمة محك الجذرالكامن كمؤشر على أحادية البعد، فقد ذكر لورد (As cited in Albanese & Forsyth, 1984) أن الفقرات، تكون أحادية البعد، إذا كان حاصل قسمة الجذر الكامن للعامل الأول إلى العامل الثاني كبيرة وتزيد على (٢). والمحك الآخر هو نسبة التباين المفسّر، إذ بلغت نسبة التباين المفسّر للعامل ٢٠,٦٨٪ من التباين الكلي، مما يؤكد وجود عامل طاغ في الاختبار، وهو ما يتماشى مع ما اقترحه ريكاس (Rechase) عام ١٩٨٥، من أنه:إذا أمكن أن يفسّر العامل الأول ٢٠٪ من التباين المفسّر على الأقل، فإن ذلك يعد مؤشراً كافياً لأحادية البعد (Hambelton & Swaminathau,1985; Hattie, 1985). كما تم تمثيل الجذور الكامنة للعوامل جميعها بيانياً لبيانات الاختبار، وهو ما يعرف باختبار فرز العوامل (Scree plot)، حيث يلاحظ وجود عامل سائد على بقية العوامل الأخرى في نموذج الاختبار كما هوموضح في الشكل رقم (١).



الشكل رقم (١)
التمثيل البياني لقيم الجذور الكامنة للعوامل المكونة للاختبار

ويلاحظ من الشكل رقم (١) أن الخط المنحني بدأ يغير ميله بشكل مفاجئ بين النقطتين المناظرتين للعاملين الثاني والثالث، ويلاحظ أن قيم الجذور الكامنة للعوامل بدءا من الثالث أصبحت متقاربة، مما يعد أساساً لاعتماد العوامل الثلاثة الأولى، واستبعاد بقية العوامل. ومما هو جدير بالقول أن الاستدلال على عدد العوامل التي يمكن أن تعتمد دلالتها من مخطط سكري لا يكون بالدقة الكافية؛ لأنه يعتمد على تقدير تقريبي للموقع الذي يتم فيه التحول في ميل الخط المار بالنقاط المناظرة للعوامل. كما تم حساب معاملات ثبات الاتساق الداخلي باستخدام (كرونباخ α) للاختبار باعتباره مؤشراً آخر على أحادية البعد، وفق ما أشار إليه هيتي (Hattie, 1985)، الذي عده أكثر مؤشرات أحادية البعد استخداماً. فمعامل كرونباخ ألفا عبارة عن الحد الأدنى لنسبة تباين الاختبار الذي يعزى للعوامل المشتركة بين الفقرات؛ أي أنه تقدير للتباين الذي يعزى لجميع العوامل المشتركة بين الفقرات، وأن وجود قيمة عالية لمعامل ألفا دليل على وجود عامل مشترك، ومن هنا تأتي علاقة ألفا بالبعدية. حيث بلغت قيمة معامل كرونباخ α (0.89) وهي قيمة عالية، ويعد ذلك مؤشراً على أحادية البعد. كما تم حساب معامل الارتباط بين الأداء على الفقرة والأداء على الاختبار الكلي بعد حذف الفقرة (Corrected Item Total Correlation)، وذلك ضمن افتراض بأن الارتباطات المرتفعة (التي تزيد على 0.2) إذا تحققت بين معظم الفقرات أو جميعها، فإن ذلك مؤشر على أحادية البعد

(Hattie, 1985). وتراوحت قيم معاملات الارتباط للاختبار بين (0.427 و0.631) وجميعها كانت ذات دلالة إحصائية عند مستوى الدلالة ($\alpha = 0.01$)، مما يشير إلى أن هذه الفقرات جميعها تتشارك في قياس بعد واحد تعبر عنه العلامة الكلية. أما افتراض الاستقلال الموضعي (Local Independence) فقد عدَ مكافئا لافتراض أحادية البعد وفق ما أشار إليه هامبلتون وسواميناثان (Hambleton & Swaminathan, 1985)، وكذلك هامبلتون وروجرز (Hambleton & Rogers,1991)، حيث عدا هذين الافتراضين متلازمين، بمعنى أنه إذا تحقق الافتراض الأول، فإن الافتراض الثاني يتحقق أيضا. وفيما يتعلق بالتحرر من السرعة في الأداء (Speedeness): فقد أنهى جميع الطلبة الإجابة عن فقرات الاختبار ضمن الوقت المحدد للإجابة، بمعنى أن إخفاق الأفراد في إجابة فقرات الاختبار عائد إلى انخفاض قدرتهم، وليس تأثير عامل السرعة على إجاباتهم.

٦ - معايرة فقرات الاختبار للنموذج الثلاثي المعلمة، حيث تبين وجود فقرتين غير مطابقتين للنموذج المستخدم، والفقرتان هما (3، 4)، وبالتالي فإن عدد الفقرات الكلي الذي تم استخدامه لأغراض هذه الدراسة (33) فقرة.

٧ - تقسيم عينة الدراسة بأحجام ذات عينات (100، 500، 1000) بحيث تراعي أن تكون من مستويات قدرة مختلفة؛ أي تكون موزعة على متصل القدرة.

نتائج الدراسة ومناقشتها

فيما يتعلق بالإجابة عن السؤال الأول: ما أثر طريقة التقدير (طريقة الأرجحية العظمى، وطريقة بييز باستخدام أسلوب توقع الاقتران البعدي، وطريقة بييز باستخدام أسلوب تعظيم الاقتران البعدي) وحجم العينة على دقة تقدير معالم الفقرات؟ فقد تم تقدير قيم معالم الصعوبة والتمييز والتخمين لكل فقرة من فقرات الاختبار بصورته النهائية (٣٣) فقرة؛ أي بعد استبعاد الاستجابات الخاصة بالفقرتين غير المطابقتين للنموذج والأخطاء المعيارية لها باستخدام طريقة الأرجحية العظمى، وطريقة بييز بأسلوبي توقع الاقتران، وتعظيم الاقتران،

وذلك عند أحجام العينات (١٠٠، ٥٠٠، ١٠٠٠). والجدول رقم (٢) يبين الإحصائيات الوصفية الخاصة بالأخطاء المعيارية لكل معلم من معالم الفقرات: الصعوبة، والتمييز، والتخمين، تبعاً لطريقة التقدير (طريقة الأرجحية العظمى، وطريقة بييز باستخدام أسلوب توقع الاقتران البعدي، وطريقة بييز باستخدام أسلوب تعظيم الاقتران البعدي) عند أحجام العينات (١٠٠، ٥٠٠، ١٠٠٠).

الجدول رقم (٢)

الإحصائيات الوصفية الخاصة بالأخطاء المعيارية للتقدير لمعالم الفقرات: الصعوبة، التمييز، التخمين، تبعا لطريقة التقدير وحجم العينة

حجم العينة	الإحصائيات الوصفية	طريقة التقدير								
		الأرجحية العظمى			تعظيم الاقتران			توقع الاقتران		
		الصعوبة	التمييز	التخمين	الصعوبة	التمييز	التخمين	الصعوبة	التمييز	التخمين
١٠٠	الوسط الحسابي	٠,٣٨	٠,٣٦	٠,٣٤	٠,٢٨	٠,٢٨	٠,٢٢	٠,٢١	٠,٢٧	٠,١٨
	أعلى قيمة	٠,٤٦	٠,٤٨	٠,٥١	٠,٤١	٠,٣٦	٠,٣٩	٠,٣٤	٠,٣٦	٠,٢٦
	أدنى قيمة	٠,٢٦	٠,١٩	٠,١٨	٠,١٩	٠,١٨	٠,١٤	٠,١٣	٠,١١	٠,٠٩
	الانحراف المعياري	٠,٤٠	٠,٣٣	٠,٣٠	٠,١٩	٠,١٨	٠,١٧	٠,١٦	٠,١٨	٠,١٥
	RMS	٠,٢١	٠,١٩	٠,٢١	٠,٢٣	٠,٢٠	٠,٢٤	٠,١٨	٠,١٧	٠,١٩
	(RMS) التباين	٠,١٠	٠,١٤	٠,١٥	٠,١١	٠,١٦	٠,١٤	٠,١٠	٠,١٢	٠,٠٨
٥٠٠	الوسط الحسابي	٠,٣٤	٠,٣٢	٠,٢٧	٠,٢٩	٠,٢٤	٠,٢٥	٠,٢٧	٠,٢٣	٠,٩
	أعلى قيمة	٠,٥٢	٠,٤٥	٠,٤١	٠,٣٩	٠,٤٠	٠,٤٥	٠,٤١	٠,٣٨	٠,٣٢
	أدنى قيمة	٠,١٤	٠,١١	٠,١٣	٠,٠٩	٠,١٤	٠,٢٠	٠,١٦	٠,٠٨	٠,٠٩
	الانحراف المعياري	٠,٣١	٠,٤٢	٠,٣٦	٠,٢٤	٠,٢٣	٠,١٩	٠,٢١	٠,١٨	٠,١٤
	RMS	٠,٢٠	٠,٢٢	٠,٢٠	٠,١٦	٠,١٦	٠,١٨	٠,١٧	٠,١٥	٠,١٠
	(RMS) التباين	٠,١٧	٠,١٩	٠,١٧	٠,١١	٠,١٠	٠,٠٩	٠,٠٩	٠,٠٨	٠,٠٧

تابع/ الجدول رقم (٢)

الإحصائيات الوصفية الخاصة بالأخطاء المعيارية للتقدير لمعالم الفقرات: الصعوبة، التمييز، التخمين، تبعا لطريقة التقدير وحجم العينة

حجم العينة	الإحصائيات الوصفية	طريقة التقدير								
		الأرجحية العظمى			تعظيم الاقتران			توقع الاقتران		
		الصعوبة	التمييز	التخمين	الصعوبة	التمييز	التخمين	الصعوبة	التمييز	التخمين
١٠٠٠	الوسط الحسابي	٠,٢٦	٠,٢٢	٠,٢٠	٠,٢٥	٠,٢٢	٠,١٩	٠,٢٤	٠,٢١	٠,١٩
	أعلى قيمة	٠,٤١	٠,٣٤	٠,٣٦	٠,٣٨	٠,٣٣	٠,٢٨	٠,٣٧	٠,٣٩	٠,٢٦
	أدنى قيمة	٠,١٧	٠,١٨	٠,١٢	٠,١٥	٠,١٠	٠,١٢	٠,١٤	٠,١٢	٠,٠٩
	الانحراف المعياري	٠,١٦	٠,١٥	٠,١١	٠,١٢	٠,١٢	٠,١١	٠,١٣	٠,١٤	٠,٠٩
	RMS	٠,١٦	٠,١٣	٠,١٠	٠,١٠	٠,١١	٠,٠٩	٠,١١	٠,١٢	٠,٠٨
	(RMS) التباين	٠,٠٩	٠,٠٨	٠,٠٧	٠,٠٨	٠,١٠	٠,٠٧	٠,٠٩	٠,١٠	٠,٠٦

ويلاحظ من الجدول رقم (٢) أن الوسط الحسابي للأخطاء المعيارية لتقديرات معالم الفقرات الملاحظ كان أعلى، وذلك عند استخدام طريقة الأرجحية العظمى، وذلك عند أحجام العينات (١٠٠، ٥٠٠)، تليها طريقة تعظيم الاقتران، ثم طريقة توقع الاقتران؛ أي أن طريقة توقع الاقتران كانت الأكثر دقة في تقديرات معالم الفقرة الثلاثة : الصعوبة، والتمييز، والتخمين. كما أن أعلى قيمة للأخطاء المعيارية والانحراف المعياري لها كانت وفق طريقة الأرجحية العظمى؛ وهذا يعني أن طريقة الأرجحية العظمى تعمل على المغالاة في تقدير معالم الفقرات عند حجم العينة ١٠٠، وبشكل أقل عند حجم العينة ٥٠٠، أما عند حجم العينة ١٠٠٠ فقد كان هناك تطابق بين طرق التقدير المستخدمة لتقدير معالم الفقرات، وبالتالي زيادة الأخطاء المعيارية.

وللإجابة عن السؤال الأول، الخاص بتعرّف دلالة الفروق بين متوسطات

الأخطاء المعيارية لتقدير معلمة الصعوبة للفقرات عند أحجام عينات مختلفة، وطريقة التقدير والتفاعل بينهما، تم استخدام التصميم العاملي للقياسات المتكررة على أحد العاملين فقط؛ لأن التحليل تم على العينة نفسها بثلاثة أحجام مختلفة (متغير غير مكرر عليه القياس)، ولكن اختلفت طريقة التقدير (متغير مكرر عليه القياس). لذا تم استخدام تحليل التباين الثنائي × في حالة القياسات المتكررة على أحد العاملين فقط. كما هو موضح في الجدول رقم (٣).

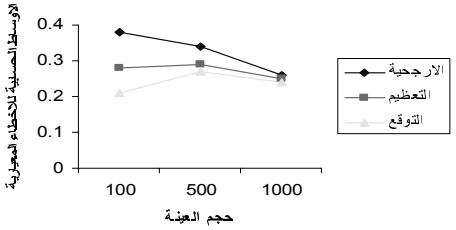
الجدول رقم (٣)

نتائج تحليل التباين للقياسات المتكررة لأثر متغيري حجم العينة، وطريقة التقدير في تقدير معلمة الصعوبة للفقرات

مصدر التباين	مجموع المربعات	درجات الحرية	وسط المربعات	قيمة F	الدلالة الإحصائية
بين الأفراد (حجم العينة)	٠,٤١٢	٢	٠,٢٠٦	٠,٧٧٤	٠,٤٣٥
الخطأ	٤٢٣,٢٥٦	١٥٩٠	٠,٢٦٦		
داخل الأفراد (طريقة التقدير)	٠,٧٥٩	١	٠,٧٥٩	٢,٧٩٠	٠,٠٦٦٢
التفاعل	١,٨٧٦	٢	٠,٩٣٨	٣,٤٤٩	٠,٠٣٥٦❖
الخطأ	٤٣٢,١٢٤	١٥٩٠	٠,٢٧٢		

❖ ذات دلالة إحصائية عند مستوى الدلالة (α =٠,٠٥)

ويلاحظ من الجدول رقم (٣) عدم وجود فروق ذات دلالة إحصائية تعزى لأثر متغيري: طريقة التقدير، وحجم العينة، ووجود أثر التفاعل بين حجم العينة وطريقة التقدير على تقديرات معلمة الصعوبة للفقرات؛ أي أن هناك تأثيراً مشتركا لكل من العاملين، ويمكن تمثيل هذا التفاعل بيانيا كما في الشكل رقم (٢) وذلك بالاستعانة بالأوساط الحسابية في الجدول.



الشكل رقم (٢)

التمثيل البياني للتفاعل بين حجم العينة وطريقة التقدير في تقدير معلم الصعوبة للفقرات

ويلاحظ أن الأخطاء المعيارية الملاحظة كانت أعلى وفق طريقة الأرجحية العظمى، تليها طريقة تعظيم الاقتران، ثم طريقة توقع الاقتران وذلك عند أحجام العينات المختلفة؛ أي أن طريقة توقع الاقتران كانت الأكثر دقة في تقديرات معالم الفقرة. كما يتبين من الشكل رقم (١) أن هناك تفاوتا واضحا بين قيم الأخطاء المعيارية في حالة حجم العينة ١٠٠، بينما تقلص هذا الفرق عندما كان حجم العينة ٥٠٠، وتلاشى تقريباً عندما كان حجم العينة ١٠٠٠؛ أي أن طريقة الأرجحية العظمى في التقدير كانت أقل الطرق دقة في حالة حجم العينة ١٠٠؛ وتصبح الطرق الثلاث متماثلة في دقتها في حالة زيادة حجم العينة.

وللتعرف إلى دلالة الفروق بين متوسطات الأخطاء المعيارية لتقدير معلمة التمييز للفقرات عند أحجام عينات مختلفة، وطريقة التقدير، والتفاعل بينهما، فقد تم استخدام التصميم العاملي للقياسات المتكررة على أحد العاملين فقط؛ لأن التقدير تم على العينة نفسها بثلاث أحجام مختلفة (متغير غير مكرر عليه القياس)، ولكن اختلفت طريقة التقدير (متغير مكرر عليه القياس). لذا تم استخدام تحليل التباين الثنائي ٣×٣ في حالة القياسات المتكررة على أحد العاملين فقط. كما هو موضح في الجدول رقم (٤).

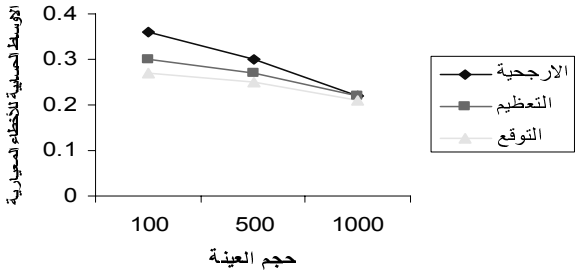
الجدول رقم (٤)

نتائج تحليل التباين للقياسات المتكررة لأثر متغيري حجم العينة، وطريقة التقدير في تقدير معلمة التمييز للفقرات

مصدر التباين	مجموع المربعات	درجات الحرية	وسط المربعات	قيمة F	الدلالة الإحصائية
بين الأفراد (حجم العينة)	٠,٣٦٦	٢	٠,١٨٣	٠,٨٤٧	٠,٣٦٧
الخطأ	٣٩٨,٥٥٠	١٥٩٠	٠,٢١٦		
داخل الأفراد (طريقة التقدير)	٠,٦٥٦	١	٠,٦٥٦	٢,٥٣٣	٠,٠٧٤١
التفاعل	١,٨٩٢	٢	٠,٩٤٦	٣,٦٥٣	٠,٠٣٦٧❖
الخطأ	٤١٢,٥٩١	١٥٩٠	٠,٢٥٩		

❖ ذات دلالة إحصائية عند مستوى الدلالة (α =٠,٠٥)

ويلاحظ من الجدول رقم (٤) عدم وجود فروق ذات دلالة إحصائية تعزى لأثر متغيري: طريقة التقدير، وحجم العينة، ووجود أثر التفاعل بين حجم العينة وطريقة التقدير على تقديرات معلمة التمييز للفقرات؛ أي أن هناك تأثيراً مشتركاً لكل من العاملين، ويمكن تمثيل هذا التفاعل بيانياً كما في الشكل رقم (٣) وذلك بالاستعانة بالأوساط الحسابية في الجدول رقم (٢).



الشكل رقم (٣)

التمثيل البياني للتفاعل بين حجم العينة وطريقة التقدير في تقدير معلم التمييز للفقرات

ويلاحظ أن الأخطاء المعيارية الملاحظة كانت أعلى وفق طريقة الأرجحية العظمى، تليها طريقة تعظيم الاقتران، ثم طريقة توقع الاقتران وذلك عند أحجام

العينات المختلفة؛ أي أن طريقة توقع الاقتران كانت الأكثر دقة في تقديرات القدرة، كما يتبين من الشكل ٣ أن هناك تفاوتاً واضحاً بين قيم الأخطاء المعيارية في حالة حجم العينة ١٠٠، بينما تقلص هذا الفرق عندما كان حجم العينة ٥٠٠، وتلاشى تقريبا عندما كان حجم العينة ١٠٠٠؛ أي أن طريقة الأرجحية العظمى في تقدير معلمة التمييز كانت أقل الطرق المستخدمة دقة في حالة حجم العينة ١٠٠، وتصبح الطرق الثلاث متماثلة في دقتها في حالة زيادة حجم العينة.

وللتعرف إلى دلالة الفروق بين متوسطات الأخطاء المعيارية لتقدير معلمة التخمين للفقرات عند أحجام عينات مختلفة، وطريقة التقدير، والتفاعل بينهما، فقد تم استخدام التصميم العاملي للقياسات المتكررة على أحد العاملين فقط؛ لأن التحليل تم على العينة نفسها عند ثلاثة أحجام مختلفة (متغير غير مكرر عليه القياس)، ولكن اختلفت طريقة التقدير (متغير مكرر عليه القياس)، لذا تم استخدام تحليل التباين الثنائي ٣×٣ في حالة القياسات المتكررة على أحد العاملين فقط. كما هو موضح في الجدول رقم (٥).

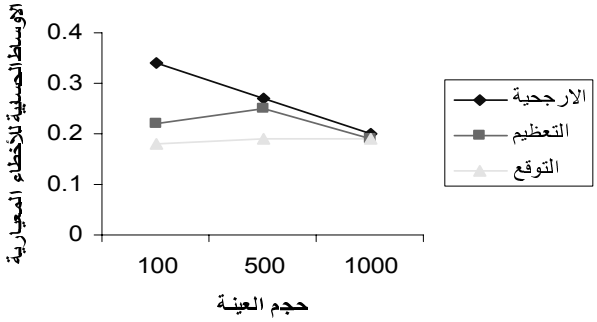
الجدول رقم (٥)

نتائج تحليل التباين للقياسات المتكررة لأثر متغيري حجم العينة، وطريقة التقدير، لمعلمة التخمين للفقرات

مصدر التباين	مجموع المربعات	درجات الحرية	وسط المربعات	قيمة F	الدلالة الإحصائية
بين الأفراد (حجم العينة)	٠,٣١٠	٢	٠,١٥٥	٠,٧٦٧	٠,٤١٢
الخطأ	٣٢٠,٦١٨	١٥٩٠	٠,٢٠٢		
داخل الأفراد (طريقة التقدير)	٠,٥٧٨	١	٠,٥٧٨	٢,٧٦٦	٠,٠٦٩٨
التفاعل	١,٦٥٨	٢	٠,٨٢٩	٣,٩٦٧	٠,٠٣٧٦❖
الخطأ	٣٣٢,٦٥٥	١٥٩٠	٠,٢٠٩		

❖ ذات دلالة إحصائية عند مستوى الدلالة (α =٠,٠٥)

ويلاحظ من الجدول رقم (٥) عدم وجود فروق ذات دلالة إحصائية تعزى لأثر متغيري: طريقة التقدير، وحجم العينة، ووجود أثر التفاعل بين حجم العينة وطريقة التقدير على تقديرات معلمة التخمين للفقرات؛ أي أن هناك تأثيرا مشتركا لكل من العاملين، ويمكن تمثيل التفاعل بيانياً كما في الشكل رقم (٤) وذلك بالاستعانة بالأوساط الحسابية في الجدول رقم (٢).



الشكل رقم (٤)

التمثيل البياني للتفاعل بين حجم العينة وطريقة التقدير في تقدير معلم التخمين للفقرات

ويلاحظ أن الأخطاء المعيارية الملاحظة كانت أعلى وفق طريقة الأرجحية العظمى، تليها طريقة تعظيم الاقتران، ثم طريقة توقع الاقتران وذلك عند أحجام العينات المختلفة؛ أي أن طريقة توقع الاقتران كانت الأكثر دقة في تقديرات القدرة، كما يتبين من الشكل ٤ أن هناك تفاوتاً واضحاً بين قيم الأخطاء المعيارية في حالة حجم العينة ١٠٠، بينما تقلص هذا الفرق عندما كان حجم العينة ٥٠٠، وتلاشى تقريبا عندما كان حجم العينة ١٠٠٠؛ أي أن طريقة الأرجحية العظمى في تقدير معلمة التمييز كانت أقل الطرق المستخدمة دقة في حالة حجم العينة ١٠٠؛ وتصبح الطرق الثلاث متماثلة في دقتها في حالة زيادة حجم العينة. وتتفق هذه النتيجة مع ما توصل إليه جاو وجن (Gao & Chen, 2005) من أن التقديرات كانت متشابهة وفق طريقتي الأرجحية العظمى الهامشية، وطريقة بييز الهامشية عند أحجام العينات الكبيرة، وأن التقديرات كانت مختلفة عند حجم العينة الصغير. كما اتفقت مع نتائج دراسة جيمس ودانييل والين وينغ شن (James, Daniel, Allen, & Young-Shun, 2000) التي

أشارت أن كلا من طريقتي الأرجحية العظمى الهامشية وطريقة ماركوف المرتبطة بأسلوب بييز (Markov) (Chain Monte Carlo) متشابهة في تقديراتها وجيدة، كما تتشابه في بعض جوانبها مع ما توصل إليه كيم (Kim, 2001) من أن دقة التقديرات التي تم الحصول عليها باستخدام طريقة الأرجحية العظمى، وطريقة الأرجحية العظمى الهاشمية، وطريقة بييز التي تعتمد على توقع التوزيع البعدي، وذلك من خلال استخدام أسلوب المحاكاة كانت متشابهة. كما تتفق مع نتائج دراسة جيمس وآخرون (James et al., 2000) التي بينت أن كلاً من طريقتي الأرجحية العظمى الهامشية وطريقة ماركوف (Markov) المرتبطة بأسلوب بييز متشابهة في تقديراتها إلى حد كبير.

أما فيما يتعلق بالإجابة عن السؤال الثاني: ما أثر طريقة التقدير (طريقة الأرجحية العظمى، وطريقة بييز باستخدام أسلوب توقع الاقتران البعدي، وطريقة بييز باستخدام أسلوب تعظيم الاقتران البعدي)، وحجم العينة على دقة تقدير القدرات للأفراد، تم تقدير قيم معالم تقديرات القدرة للأفراد والأخطاء المعيارية لها باستخدام طريقة الأرجحية العظمى، وطريقة بييز بأسلوبي توقع الاقتران، وتعظيم الاقتران، وذلك عند أحجام العينات (١٠٠،٥٠٠،١٠٠٠). وقبل ذلك تم حذف فرد عند معايرة العينة ذات الحجم (٥٠٠) وذلك عند استخدام طريقة الأرجحية العظمى، كما تم حذف (٦) أفراد عند معايرة العينة ذات الحجم (١٠٠٠)، وذلك عند استخدام طريقة الأرجحية العظمى كذلك؛ والسبب في ذلك لأن مقدار الخطأ في تقدير القدرة لهما كبير، وهذا ما دلت عليه القيمة (٩٩٩,٠٠٠❖) في برنامج (Bilog -Mg_3)؛ وهذا يعني أن مقدار المعلومات التي يعطيها لهذه القدرات ضئيل جداً، في حين لم يتواجد ذلك عند استخدام طريقتي بييز (التوقع، والتعظيم)، حيث تراوحت قيم الأخطاء المعيارية لهؤلاء الأفراد المحذوفين عند استخدام طريقة بييز التوقع ما بين (٠,٣١٢٤) و(٠,٨٨٤٩)، أما بالنسبة للأخطاء المعيارية المقابلة لها عند استخدام طريقة بييز التعظيم فقد تراوحت ما بين (٠,٣٥٨) و(٠,٦٣٥١). ويبين الجدول رقم (٦) الإحصائيات الوصفية الخاصة بالأخطاء المعيارية لمعلم تقديرات القدرة للأفراد وفق طرق التقدير الثلاث الخاصة بتقديرات القدرة للأفراد في برنامج (Bilog-Mg_3).

الجدول رقم (٦)

الإحصائيات الوصفية الخاصة بالأخطاء المعيارية لمعلم القدرة للأفراد وفق طريقة التقدير وحجم العينة

حجم العينة	الإحصائيات الوصفية	طريقة التقدير		
		الأرجحية العظمى	تعظيم الاقتران	توقع الاقتران
١٠٠	الوسط الحسابي	٠,٥٥	٠,٤٠	٠,٣٢
	أعلى قيمة	٠,٥٥	٠,٩٥	٠,٦٤
	أدنى قيمة	٠,٢١	٠,٢٨	٠,٠٦
	الانحراف المعياري	٠,٨٧	٠,١٦	٠,١٨
	RMS	٠,٢٨	٠,٢٣	٠,٣٣
	التباين (RMS)	٠,٠٩	٠,٠٨	٠,١١
٥٠٠	الوسط الحسابي	٠,٦٤	٠,٤١	٠,٣٤
	أعلى قيمة	١٢,٠٠	٠,٩٤	١,٧٤
	أدنى قيمة	٠,٠٩	٠,١٢	٠,٠٤
	الانحراف المعياري	١,٢٢	٠,١٦٤	٠,١٩
	RMS	٠,٢٨	٠,٣٠	٠,٣٧
	التباين (RMS)	٠,٠٩	٠,٠٨	٠,١٣
١٠٠٠	الوسط الحسابي	٠,٥٣	٠,٤٢	٠,٣٩
	أعلى قيمة	٧,٠٣	٠,٩٨	٠,٩٥
	أدنى قيمة	٠,٢١	٠,٣١٤	٧٠,٠٨
	الانحراف المعياري	٠,٧٣	٠,١٤	٠,١٥
	RMS	٠,٣٢	٠,٣٥	٠,٣٩
	التباين (RMS)	٠,١٢	٠,١٢	٠,١٥

ويلاحظ من الجدول رقم (٦) أن الوسط الحسابي للأخطاء المعيارية لتقديرات القدرة للأفراد الملاحظ كان أعلى وذلك عند التقدير باستخدام طريقة الأرجحية العظمى، وذلك عند أحجام العينات المختلفة، تليها طريقة تعظيم الاقتران، ثم طريقة توقع الاقتران؛ أي أن طريقة توقع الاقتران كانت الأكثر دقة في تقديرات القدرة للأفراد. كما أن أعلى قيمة للأخطاء المعيارية والانحراف المعياري لها كانت وفق طريقة الأرجحية العظمى؛ وهذا يعني أن طريقة الأرجحية العظمى تعمل على المغالاة في تقدير قدرات الأفراد، وبالتالي زيادة الأخطاء المعيارية. وهذا ينسجم مع ما توصل إليه روندال ونيلسن وليوي ولندن (Rondall, 2007; Lewy & Nielsen, 2002; Linden, 1998) من أن تقديرات القدرة للأفراد وفق طريقة بييز تتمتع بثبات أفضل. كما ينسجم مع دراسة وانغ وفيسبول (Wang & Vispoel, 1998) التي أشارت إلى أن طريقة الأرجحية العظمى تعطي أخطاء عالية في تقدير قدرات الأفراد مقارنة بطريقة بييز التي تعتمد أسلوبي توقع الاقتران وتعظيم الاقتران. إلا أنه يتناقض مع ما توصل إليه وبان وهانسون وونغ وهاريس وكوينج (Ban et al., 2000)، من أنه تزداد دقة التقديرات لمعلمة القدرة للمفحوصين عند استخدام طريقة الأرجحية العظمى مقارنة بدقة التقديرات معلمة القدرة عند استخدام طريقة بييز التي تعتمد أسلوب توقع الاقتران البعدي. كما يتناقض مع ما توصل إليه كيم (Kim, 2001) من أن التقديرات كانت متشابهة عند استخدام طريقتي الأرجحية العظمى وبييز. كما يلاحظ من الجدول رقم (٦) أن الانحراف المعياري للأخطاء المعيارية وفق طريقة تعظيم الاقتران كانت الأقل قيمة وذلك عند أحجام العينات المختلفة؛ وهذا يعني أن هذه الطريقة هي الفضلى من حيث التجانس في قيم الأخطاء المعيارية، وعدم تقديم أخطاء معيارية شاذة، تليها طريقة توقع الاقتران، ثم الأرجحية العظمى، إلا أنه باستخدام محك الجذر التربيعي لمربعات الأخطاء المعيارية لتقديرات القدرة للأفراد (RMS) فقد كان التقدير باستخدام الأرجحية العظمى له قيمة أقل، تليها تعظيم الاقتران، ثم توقع الاقتران؛ حيث أنه بازدياد الأخطاء المعيارية تقل قيمة الجذر التربيعي لمربعاتها.

وللإجابة عن السؤال الثاني للدراسة، وللتعرف الى دلالة الفروق بين متوسطات التقديرات لقدرات الأفراد عند أحجام عينات مختلفة وطريقة التقدير والتفاعل بينهما، فقد تم استخدام التصميم العاملي للقياسات المتكررة على أحد العاملين فقط؛ لأن التحليل تم على العينة نفسها بثلاثة أحجام مختلفة (متغير غير مكرر عليه القياس)، ولكن اختلفت طريقة التقدير (متغير مكرر عليه القياس). لذا تم استخدام تحليل التباين الثنائي ٣×٣ في حالة القياسات المتكررة على أحد العاملين فقط. كما هو موضح في الجدول رقم (٧):

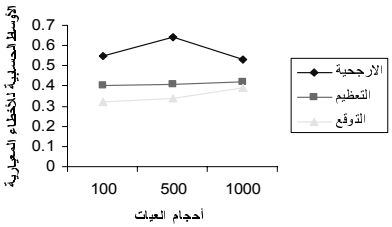
الجدول رقم (٧)
نتائج تحليل التباين للقياسات المتكررة لأثر متغيري حجم العينة، وطريقة التقدير في تقدير معالم القدرة للأفراد

مصدر التباين	مجموع المربعات	درجات الحرية	وسط المربعات	قيمة F	الدلالة الإحصائية
بين الأفراد (حجم العينة)	٠,٦٢٨	٢	٠,٣١٤	٠,٨٣٣	٠,٤٣٥❖
الخطأ	٥٩٨,٨٥٧	١٥٩٠	٠,٣٧٧		
داخل الأفراد (طريقة التقدير)	١٥,٦٣٠	١	١٥,٦٣٠	٣٩,٧٨٧	٠,٠٠٠❖
التفاعل	٣,٧١٠	٢	١,٨٥٥	٤,٧٢١	٠,٠٠٩
الخطأ	٦٢٤,٦٣٤	١٥٩٠	٠,٣٩٣		

❖ ذات دلالة إحصائية عند مستوى الدلالة (α =٠,٠٥)

يلاحظ من الجدول رقم (٧) وجود فروق ذات دلالة إحصائية تعزى لأثر متغيرات طريقة التقدير، والتفاعل بين حجم العينة وطريقة التقدير، وعدم فروق دالة إحصائياً تعزى لأثر متغير حجم العينة، وللكشف عن مواقع الفروق بين متوسطات الأخطاء المعيارية للتقدير، تم استخدام اختبار شفيه للمقارنات البعدية، حيث تبين وجود فرق ذي دلالة إحصائية بين طريقة الأرجحية العظمى وطريقة بييز التوقع ولصالح طريقة الأرجحية العظمى، وكذلك تبين وجود فرق ذي دلالة إحصائية بين طريقة الأرجحية العظمى، وطريقة بييز التعظيم، وكذلك عدم وجود

فرق بين طريقة بييز التعظيم، وبييز التوقع، ولتوضيح أثر التفاعل بين حجم العينة وطريقة التقدير في تقدير معلمة القدرة للأفراد؛ أي أن هناك تأثيرا مشتركا لكل من العاملين، فقد تم تمثيل هذا التفاعل بيانيا كما في الشكل رقم (٥) وذلك بالاستعانة بالأوساط الحسابية في الجدول رقم (٦). كما يوضح الشكل رقم (٥) التفاعل بين حجم العينة وطريقة التقدير في دقة تقدير قدرات الأفراد. ويلاحظ أن الأخطاء المعيارية كانت أعلى وفق طريقة الأرجحية العظمى، تليها طريقة تعظيم الاقتران، ثم طريقة توقع الاقتران وذلك عند أحجام العينات المختلفة؛ أي أن طريقة توقع الاقتران كانت الأكثر دقة في تقديرات القدرة للأفراد.



الشكل رقم (٥)
التفاعل بين حجم العينة وطريقة التقدير في دقة تقدير قدرات الأفراد

يلاحظ من الشكل رقم (٥) أن الأخطاء المعيارية الملاحظة كانت أعلى وفق طريقة الأرجحية العظمى، تليها طريقة تعظيم الاقتران، ثم طريقة توقع الاقتران وذلك عند أحجام العينات المختلفة؛ أي أن طريقة توقع الاقتران كانت الأكثر دقة في تقديرات القدرة للأفراد. ويلاحظ من الشكل كذلك، أن التقديرات كانت أكثر دقة عند حجم العينة ١٠٠ وفق طريقة التقدير باستخدام توقع وتعظيم الاقتران، ثم عند حجم العينة (٥٠٠، ١٠٠٠) بالترتيب؛ وهذا يشير إلى قيمة قليلة لمتوسطات الأخطاء المعيارية، تشير إلى دقة أعلى في التقدير؛ أي أن هذه الطريقة تزداد الدقة فيها عند أحجام العينات القليلة. وهذا يتناقض مع ما توصل إليه بان وآخرون (Ban et al., 2000)، من أن طريقة بييز لم تعمل بصورة مناسبة مع حجم العينات الصغيرة. وينسجم مع ما توصل إليه ميسلفي (Miselvy, 1986) من أن طريقة بييز يتوقع فيها التحسن، وخاصة إذا كان حجم العينة صغيراً. أما عند

استخدام طريقة الأرجحية العظمى، وهي الأقل دقة في التقدير لقدرات الأفراد، فقد كانت أقل دقة عند حجم العينة ٥٠٠، في حين أنها كانت أكثر دقة عند حجم العينة ١٠٠٠، وأقل عند حجم العينة (١٠٠).

فيما يتعلق بالإجابة عن السؤال الثالث: ما أثر طريقة التقدير (طريقة الأرجحية العظمى، وطريقة بييز باستخدام أسلوب توقع الاقتران البعدي، وطريقة بييز باستخدام أسلوب تعظيم الاقتران البعدي) وحجم العينة على ثبات الاختبار؟ تم تقدير الثبات الامبريقي (الثبات التجريبي) الذي يعتمد على متوسطات الأخطاء المعيارية لتقديرات القدرة للمفحوصين، ويظهر من خلال المرحلة الثالثة من مراحل التحليل في برنامج (Bilog-Mg3)، عند أحجام العينات (100، 500، 1000)، كما تم تقدير الثبات كذلك وفق النظرية الكلاسيكية باستخدام معادلة كرونباخ ألفا. ويبين الجدول رقم (٨) قيم معاملات الثبات الكلاسيكي والتجريبي المقدرة وفق طرق التقدير الثلاثة عند أحجام العينات المختلفة.

الجدول رقم (٨)

قيم معاملات الثبات الكلاسيكي (كرونباخ الفا) والتجريبي المقدرة وفق طرق التقدير واختلاف حجم العينة

أحجام العينات	معامل كرونباخ ألفا وفق النظرية الكلاسيكية	معاملات الثبات التجريبي وفق طرق التقدير:		
		الأرجحية العظمى	تعظيم الاقتران	توقع الاقتران
١٠٠	٠,٩١	٠,٨٩	٠,٩١	٠,٩٣
٥٠٠	٠,٩٢	٠,٩٣	٠,٩٢	٠,٩١
١٠٠٠	٠,٩٣	٠,٩٢	٠,٩١	٠,٩٠

ويلاحظ من الجدول رقم (٨)، أن قيم معاملات الثبات الملاحظة، والمقدرة وفق طريقة الأرجحية العظمى كانت الأعلى، وذلك عند أحجام العينات (٥٠٠، ١٠٠٠)، تليها طريقة تعظيم الاقتران، ثم طريقة توقع الاقتران، أما عند

حجم العينة (١٠٠) فقد كانت طريقة توقع الاقتران هي الأعلى تليها طريقة تعظيم الاقتران، ثم طريقة الأرجحية العظمى. كما كانت معاملات الثبات المقدرة باستخدام تعظيم الاقتران هي الأقرب للقيم المستخرجة باستخدام النظرية الكلاسيكية. وجدير بالذكر أن الثبات المستخرج وفق طريقة الأرجحية العظمى يعتمد على تباين الخطأ الذي يقدر من خلال الوسط الحسابي لدالة المعلومات، لذلك فإن أعلى قيمة للثبات الامبريقي كانت لصالح الأرجحية العظمى. أما الثبات المستخرج وفق طريقة بييز التعظيم فيقدر تباين الخطأ من خلال مقلوب أفضل دالة معلومات للتوزيع البعدي. في حين أن الثبات المستخرج وفق طريقة بييز التوقع، فيقدر تباين الخطأ من خلال الوسط الحسابي لتباين التوزيع البعدي لتقديرات القدرة للأفراد ولجميع المفحوصين، لذلك يعتمد الثبات المقدر بطريقتي بييز على التوزيع البعدي.

ولمعرفة دلالة الفروق بين قيم كرونباخ ألفا الثلاثة عند أحجام العينات الثلاثة (١٠٠، ٥٠٠، ١٠٠٠) فقد استخدم الباحث الإحصائي (M) المقترح من هاكستين وولن (Hakstin and Whalen, 1976) الذي يتبع توزيع كاي تربيع بدرجات حرية تساوي (عدد المعاملات -١) وتحسب قيمة M من الصيغة الرياضية الآتية:

$$M = \frac{(J-1)(9n-11)^2}{18J(n-1)} \left[k - \frac{[\sum_{k=1}^{3}(1-r_k)^{-\frac{1}{3}}]^2}{\sum_{k=1}^{3}(1-r_k)^{-\frac{2}{3}}} \right]$$

حيث تمثل الرموز: k عدد الطرق، n: عدد أفراد المجموعة في الطريقة المستخدمة، r_k: معامل ثبات الطريقة، J : عدد فقرات الاختبار في الطريقة المستخدمة.

وكشفت نتائج التقديرات الخاصة بقيم كرونباخ ألفا عند أحجام العينات المختلفة (١٠٠، ٥٠٠، ١٠٠٠) عن عدم وجود فروق بين معاملات ثبات كرونباخ ألفا، حيث كانت قيمة الإحصائي المحسوبة (٣,٦٩) وهي أقل من القيمة الحرجة لكاي تربيع بدرجات حرية (٢) عند مستوى الدلالة الإحصائية ($\alpha = 0.05$). كما قام الباحث برسم منحنيات المعلومات للاختبار تبعاً لطرق التقدير الثلاثة: طريقة

الأرجحية العظمى، وطريقة بييز باستخدام أسلوب توقع الاقتران البعدي، وطريقة بييز باستخدام أسلوب تعظيم الاقتران البعدي، وباختلاف أحجام العينات (١٠٠، ٥٠٠، ١٠٠٠) والناتجة من برنامج (BILOG-MG)، وعند المقارنة بين منحنيات دوال المعلومات لطرق التقدير كما يظهر في الملحق رقم (١) (مع مراعاة مقياس الرسم للمحور العمودي: كمية المعلومات، والخطأ المعياري)، أن كمية المعلومات التي يقدمها الاختبار متماثلة إلى حد كبير جداً باختلاف الطريقة، وثبات حجم العينة، وأما في حالة اختلاف حجم العينة وثبات الطريقة، فيلاحظ أن كمية المعلومات التي يقدمها الاختبار كانت أكبر ما يمكن عندما كان حجم العينة ٥٠٠، يليها عند حجم العينة ١٠٠٠، بينما كانت كمية المعلومات أقل ما يمكن عند حجم العينة ١٠٠، فربما كان السبب في ذلك التوافق في مدى القدرة للمفحوصين مع مدى الصعوبة للفقرات، فقد أشارت دراسة عبابنه (٢٠٠٤) أن الدقة في تقدير معلمي القدرة والصعوبة تزداد عندما يكون مدى القدرة للمفحوصين متوافقاً مع مدى الصعوبة للفقرات، فحجم العينة القليل، وكذلك الكبير ربما يؤدي إلى زيادة احتمال ابتعاد مدى القدرة للمفحوصين عن مدى الصعوبة للفقرات، كما أن هذه النتيجة أظهرت تطابقاً مع قيم معاملات الثبات التي تم تقديرها وفق طرق التقدير الثلاث؛ حيث كانت قيم معاملات الثبات وفق طرق التقدير الثلاث هي الأعلى عند حجم العينة ٥٠٠، وتتفق هذه النتيجة مع دراسة الدرابيع (٢٠٠١)، ومع دراسة (Mislevy,1986). كما أن أقصى معلومات لتقدير معلم القدرة كان أكبر بقليل من قيمة القدرة المقابلة لصعوبة الفقرة، وهذا يتفق مع ما أشارت إليه ستوكنغ (Stocking,1990) في حالة استخدام النموذج الثلاثي المعلمة.

وللإجابة عن السؤال الرابع: هل تختلف الفعالية النسبية باختلاف طريقة التقدير (طريقة الأرجحية العظمى، وطريقة بييز باستخدام أسلوب توقع الاقتران البعدي، وطريقة بييز باستخدام أسلوب تعظيم الاقتران البعدي) وحجم العينة؟ تم إيجاد الأخطاء المعيارية عن مستويات مختارة من القدرة بحيث تغطي متصل القدرة، وذلك عن أحجام العينات المختلفة، ثم تم إيجاد الفعالية النسبية بالنسبة إلى طريقة الأرجحية العظمى التقدير المستخدمة،

وعند أحجام عينات مختلفة. ويبين الجدول رقم (٨) الأخطاء المعيارية والفعالية النسبية لتقديرات القدرة بالنسبة لطريقة الأرجحية العظمى، وطريقة بييز باستخدام أسلوب توقع الاقتران البعدي، وطريقة بييز باستخدام أسلوب تعظيم الاقتران البعدي تبعا لحجم العينة (١٠٠٠،٥٠٠،١٠٠).

الجدول رقم (٩)

الأخطاء المعيارية (SE) والفعالية النسبية (RE) لتقديرات باختلاف طريقة التقدير وحجم العينة

حجم العينة	مستوى القدرة	SE_{ML}	SE_{MB}	SE_{EB}	$RE_{MB/ML}$	$RE_{EB/ML}$
١٠٠	١,٤	٠,٣٨٦	٠,٤٢٨	٠,٣٢٩٨	١,١٠٨	٠,٨٥٤
	٠,٩٣	٠,٢٧٦	٠,٣٤٧	٠,٥٤١	١,٢٥٧	١,٩٥٤
	٠,٦٧	٠,٢٣٩	٠,٣٢١	٠,٣١٣	١,٣٤٣	١,٣٠٨
	٠,٢١	٠,٢٠٦	٠,٢٧٦	٠,٢٣٧	١,٣٤٢	١,١٤٩
	-٠,٢٠	٠,٢٢٩	٠,٣٠٤	٠,٥٢٥	١,٣٧٣	٠,٦٨٧
	-٠,٨٣	٠,٤١٥	٠,٤٥٦	٠,٥١٩	١,٠٩٨	١,٢٥٢
	-١,٣	٣,١٢	٠,٦٢٥	٠,٥٥٤	٠,٢٠٠٥	٠,١٧٧
	-٢,١	٢,٠١	٠,٧٤٢	٠,١٥٧	٠,٣٦٨	٠,٢٦١
	الوسط الحسابي	٠,٨٦	٠,٤٣٧	٠,٣٩٧	١,٠١	٠,٩٥٥
٥٠٠	١,٤	٠,٣٥٧	٠,٤٠٧	٠,٢٩٧	١,١٤	٠,٨٣
	٠,٩٣	٠,٢٧٨	٠,٣٤٦	٠,٦٣٨٥	١,٢٤	٢,٢٩
	٠,٦٧	٠,٢٣٦	٠,٢٠٥	٠,٣٩٨٨	٠,٨٦	١,٦٩
	٠,٢١	٠,٢٠٩	٠,٤٧١	٠,٥٨٥	٢,٢٥	٠,٥٣٥
	-٠,٢٠	٠,٢٦٣	٠,٣١٩	٠,٣٧١	١,٢١	١,٤١
	-٠,٨٣	٠,٥١١	٠,٥١٨	٠,٥٢٣	١,٠١	١,٠٢
	-١,٣	٣,٠٢	٠,٧١٣	٠,٥٥٢	٠,٢٣٦	٠,١٨
	-٢,١	٤,٢٠٦	٠,٨٤٤٧	٠,١١٢	٠,٢٠	٠,١٣٩
	الوسط الحسابي	١,١٣٥	٠,٤٧٧	٠,٤٣٧	١,٠١٨	١,١٠١

تابع/ الجدول رقم (٩)

الأخطاء المعيارية (SE) والفعالية النسبية (RE) لتقديرات باختلاف طريقة التقدير وحجم العينة

حجم العينة	مستوى القدرة	SE_{ML}	SE_{MB}	SE_{EB}	$RE_{MB/ML}$	$RE_{EB/ML}$
١٠٠٠	١,٤	٠,٤١٩	٠,٤٤٧	٠,٤٠٠	١,٠٧	٠,٩٥
	٠,٩٣	٠,٣٢٧	٠,٣٨٥	٠,٢٥١	١,١٧	٠,٧٧
	٠,٦٧	٠,٢٧٨	٠,٣٤٧	٠,٤١٣	١,٢٥	١,٤٩
	٠,٢١	٠,٢٣٥	٠,٣١٣	٠,٥٧١	١,٣٣	٠,٦٧
	-٠,٢٠	٠,٢٨٤	٠,٣٥٤	٠,٤٣٨	١,٢٧	١,٧٦
	-٠,٨٣	٠,٤٦٣	٠,٤٩٢	٠,٥٣٨	١,٠٦	١,١٦
	١,٣	٢,٠٠	٠,٧٣٠	٠,٥٧٤	٠,٣٧	٠,٢٨٧
	-٢,١	٢,١٨	٠,٨٥٦	٠,١٥٨	٠,٣٩	٠,٢٦
	الوسط الحسابي	٠,٧٧٣	٠,٤٩١	٠,٤١٨	٠,٩٨٨	٠,٩١٨

يلاحظ من الجدول رقم (٩) أن تقديرات القدرة باستخدام طريقة الأرجحية العظمى (ML) هي أكثر دقة عند مستويات القدرة المتوسطة، بينما تقديرات القدرة باستخدام طريقة بييز بأسلوب التعظيم (MB) هي أكثر دقة عند مستويات القدرة المتدنية، بيد أن طريقة بييز بأسلوب التوقع (EB) كانت أكثر دقة عند مستويات القدرة المتطرفة (المرتفعة جدا، والمتدنية جدا). وهذا يتفق مع ما توصل إليه (Chen & Chao, 2005) من أن طريقة بييز هي أكثر دقة عند أطراف متصل القدرة. وهذا يشير إلى أن طريقة بييز تكون مناسبة لتقدير القدرة عند أطراف متصل القدرة بمعنى أن القدرات العالية والقدرات المتدنية يتم تقديرها بصورة أفضل باستخدام طريقة بييز وهذا يتفق مع ما توصل إليه كذلك جاري وفيرمنت (Garre & Vermunt, 2006). أي أن تقديرات معلمة القدرة تكون أكثر دقة عند مستويات القدرة المتطرفة (ذوي القدرات المتدنية وذوي القدرات العالية) عند استخدام طريقة بييز التوقع، فيما تكون تقديرات

معلمة القدرة أكثر دقة عند مستويات القدرة الأخرى باستخدام طريقة الأرجحية العظمى بغض النظر عن حجم العينة.

الخلاصة والتوصيات

تبين للباحث من خلال نتائج الدراسة بأنه لا توجد فروق دالة إحصائياً في دقة التقديرات لمعالم الفقرات: الصعوبة، والتمييز، والتخمين تعزى لطريقة التقدير، وحجم العينة، إلا أن هناك تأثيراً مشتركاً لكل من العاملين، وأن طريقة بييز بأسلوب توقع الاقتران كانت الأكثر دقة في تقدير معالم الفقرة عند أحجام العينات الصغيرة، وتصبح الطرق الثلاث متماثلة في دقة تقديرها في حالة زيادة حجم العينة. وبينت النتائج وجود فروق ذات دلالة إحصائية في دقة تقدير معالم القدرة تعزى لطريقة التقدير، والتفاعل بين حجم العينة وطريقة التقدير، وعدم وجود فروق ذات دلالة تعزى لحجم العينة، وكانت طريقة بييز توقع الاقتران هي الأكثر دقة في تقدير معلمة القدرة للأفراد، وعند أحجام العينات المختلفة. كما كشفت نتائج التحليل عن عدم وجود فروق ذات دلالة إحصائية بين معاملات ثبات كرونباخ ألفا. وأشارت نتائج الدراسة كذلك أن كمية المعلومات التي يقدمها الاختبار متماثلة إلى حد كبير جداً باختلاف الطريقة، وثبات حجم العينة، وأما في حالة اختلاف حجم العينة، وثبات الطريقة فيلاحظ أن كمية المعلومات التي يقدمها الاختبار كانت أكبر ما يمكن عندما كان حجم العينة ٥٠٠، يليها عند حجم العينة ١٠٠٠، بينما كانت كمية المعلومات أقل ما يمكن عند حجم العينة ١٠٠، وهذا يشير إلى أن طريقة بييز التوقع تكون مناسبة لتقدير القدرة عند أطراف متصل القدرة؛ بمعنى أن القدرات العالية والقدرات المتدنية يتم تقديرها بصورة أفضل باستخدام طريقة بييز التوقع.

واعتماداً على ذلك فإن الباحث يوصي مطوري الاختبارات بإمكانية استخدام طريقة بييز التوقع في تقدير معالم الفقرات، وتحديداً عندما تكون عينة المفحوصين قليلة، حيث تكون دقة التقدير لمعالم الصعوبة، والتمييز، والتخمين عالية باستخدام طريقة بييز التوقع مما ينعكس إيجابا على جودة

التقديرات، وربما تساهم هذه التوصية في زيادة تمكين باني الاختبارات التكيفية (المفصلة) (Adaptive Testing)، كما يوصي الباحث مطوري الاختبارات، والجهات الحريصة على دقة التقدير لقدرات المفحوصين بإمكانية استخدام طريقة بييز التوقع في تقدير قدرات المفحوصين، إذ تبين من نتائج هذه الدراسة بأنها الأكثر دقة في التقدير عندما استخدم النموذج ثلاثي المعلمة في عملية التحليل، حيث اتصفت تلك النتائج بنوع من الاستقرار، وكانت دقة التقدير أكثر وضوحاً في حالة العينات صغيرة الحجم. كما يمكن أن تقدم هذه الدراسة دليلاً امبريقياً بأن تقدير الثبات باستخدام النموذج الثلاثي المعلمة تبقى قيمته إلى حد بعيد مستقرة، حتى وإن اختلف حجم العينة، وهذا يخالف ما هو شائع بين بعض الباحثين بأن زيادة عدد الأفراد تعمل على زيادة قيمة الثبات. فعلى سبيل المثال، لا أهمية لأفراد يجيبون على فقرات الاختبار أو المقياس بطريقة عشوائية، إذ أن برامج التحليل وفق نظرية الاستجابة للفقرة تستبعد مثل هذه الفئة وتعدهم أفراداً غير مطابقين للنموذج، كما هو الحال في هذه الدراسة، ولذلك فالجدية والحرص والمصداقية مهمة في الإجابة على فقرات الاختبار أو المقياس عند تقدير الثبات أكثر من زيادة حجم العينة، وربما تعد هذه النقطة تحديداً ميزة مهمة لنظرية الاستجابة للفقرة في حين تعتبر مأخذاً على النظرية الكلاسيكية التي لا تستبعد أي فرد عند تقدير الثبات للإخبار أو المقياس، كما يوصي الباحث بإجراء دراسات أخرى مماثلة لكن باستخدام بيانات محاكاة مولدة باستخدام البرامج الإحصائية، ودراسة المقارنة بين الطرق الثلاث في دقة التقدير لمعالم الفقرات والأفراد تحت ظروف خاصة من توزيعات مستويات القدرة (طبيعي، منتظم، ملتو التواءً سالباً، ملتو التواءً موجباً). كذلك أخذ حجم العينة وعدد الفقرات كمتغيرين مستقلين، واستخدام أساليب الإحصاء الاستدلالي في الكشف عن وجود فروق دالة إحصائياً، كما يمكن إعادة الدراسة نفسها باستخدام نماذج أخرى من نماذج نظرية الاستجابة للفقرة كالنموذجين الأحادي المعلمة والثنائي المعلمة، وكذلك استخدام مقاييس التقدير والاختبارات بأشكالها المختلفة للوصول إلى نتائج أكثر قابلية للتعميم.

Effect of the Method of Estimating Item Parameters and Abilities of Persons on the Item Parameters Values and the Psychometric Properties of Test in Light of Sample Size Change

Dr. Nedal K. Al-Shraifin
Faculty of Education -Yarmouk University
H.K.J

Abstract

This study aims at investigating effect of the method of estimating item parameters and abilities of persons on the item parameters values and the psychometric properties of test in light of sample size change. To achieve the study objective, a 33 - item achievement test of physics was constructed, and administerd on the study sample of (1012) male and female twelvth grade students. Data obtained were analyzed according to the three parameter logistic model using Bilog- Mg programs. The study results revealed statistically significant difference ($\alpha = 0.05$) among the standard error means of item parameters estimations due to interaction of method of estimate and sample size. However, no statistically significant difference ($\alpha = 0.05$) were revealed due to: sample size, method of estimate. Results of the study also revealed statistically significant differences ($\alpha = 0.05$) among the standard error means of person abilities estimations due to sample size as well as interaction of method of estimate and sample size.. However, no statistically significant differences ($\alpha = 0.05$) were revealed for the estimated reliability coefficients according to item response theory for samples of (100, 500, 1000). Also, results of the study indicated that ability parameter accuracy estimation increases in both high ability level and low ability level samples when using (EAP), Whereas accuracy of estimation for persons with medium ability level when using (MLE) regardless of sample size.

المراجع

١ - بني عواد، علي محمد (٢٠١٠). مقارنة طرق التعامل مع البيانات المفقودة في تقدير معالم الفقرات وقدرات الأفراد. رسالة دكتوراه غير منشورة، الأردن: جامعة اليرموك.

٢ - الدرابيع، ماهر (٢٠٠١). فعالية النموذج اللوغاريتمي ذي المعلمة الواحدة " نموذج راش" في دقة تقدير قدرة الفرد ومعامل صعوبة الفقرة باختلاف حجم العينة وطول الاختبار. **مجلة دراسات- العلوم الإنسانية**، ٢٨(١)، ١٩٧-٢٠٨.

٣ - عبابنه، عماد (٢٠٠٤). أثر حجم العينة وطريقة انتقائها وعدد الفقرات وطريقة انتقائها على دقة تقدير معالم الفقرة، والقدرة لاختبار قدرة عقلية باستخدام نظرية الاستجابة للفقرة. رسالة دكتوراه غير منشورة، جامعة عمان العربية للدراسات العليا، عمان.

٤ - علام، صلاح الدين (٢٠٠٥). **نماذج الاستجابة للمفردات الاختيارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي** (الطبعة الأولى). القاهرة: دار الفكر العربي.

٥ - علام، صلاح الدين (٢٠٠٦). **القياس والتقويم التربوي والنفسي: أساسياته وتطبيقاته وتوجهاته المعاصرة** (الطبعة الثانية). القاهرة: دار الفكر العربي.

٦ - عوده، أحمد (٢٠١٠). **القياس والتقويم في العملية التدريسية** (الطبعة الرابعة). اربد: دار الأمل للنشر والتوزيع.

٧ - مسلم، فؤاد عبد الرزاق (٢٠١٠). أسلوب مقترح لاختيار الفقرات وتقدير القدرة في الاختبارات التكيفية المحوسبة. رسالة دكتوراه غير منشورة، جامعة اليرموك، الأردن.

8 - Albanese, M. A.; & Forsyth, R. A. (1984). The one- two and modified

two- parameter latent trait models: An empirical study of relative fit. **Education and Psychological Measurement, 44**, 229-246.

9 - Ban. J., Hanson. B., Wang. T., Qing. Y. & Harris. D. (2000). A Comparative study of online pretest item calibration/scaling methods in CAT. Papers Presented at the Annual Meeting of the American Educational Research Association, New Orleans. [on-line]. Available: <http://eric.ed.gov>.

10 - Birnbaum, A. (1986). Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord & M. R. Novick (Eds.), **Statistical theories of mental test scores.** Reading,MA: Addison-Wesley.

11 - Chen, S. H. L. & Dodd, B. (1998). Comparison of Maximum Likelihood estimation and expected a posteriori estimation in CAT using the partial credit model. **Educational and Psychological Measurement, 58**, 569-595.

12 - Crocker, L.& Algina, J. (1987**). Introduction to classical and modern test theory**. New York: Harcourt Jovanovich.

13 - De Gruijter, D. N. & Van der kamp, L. J. (2005). Statistical test Theory for education and Psychology. Retrieved October, 2005 from: www.leidenuniv,n1/ griuijterdnme.

14 - Gao, F. & Chen, L. (2005). Bayesian or Non-Bayesian: A Comparison Study of Item Parameter Estimation in the Three-Parameter Logistic Model. **Applied Measurement in Education, 18**(4), 351-380.

15 - Garre. G. & Vermunt. K. (2006). Avoiding boundary estimation in latent class analysis by Bayesian posterior mode estimation**. Behaviormentrika**, **33**, 256-271.

16 - Glass, G. (2005). The impact of item parameter estimation of computerized adaptive testing with item cloning. Law School Admission Conceit Computerized Testing Report 02-06 November.

17 - Gronlund, N., & Linn, R. (1990). **Measurement and Evaluation in Teaching**. New York: Macmillan Publishing Company.

18 - Hakstin, A. R. & Whalen, T. E. (1976). A k- sample significance test for independent alpha coefficients. **Psychometrica, 41**(2), 219- 231.

19 - Hambleton, R. (1989). Principles and selected application of item

response theory. In Robert L. Linn, (Ed.), **Educational Measurement** (3rd ed.). New York: American Council on Education, Macmillan Publishing Company.

20 - Hambleton, R. K. & Swaminthan, H. (1985).**Item Response Theory Principle and Application.** Boston: Kluwer: Nijhoff Publishing.

21 - Hambleton, R. K. & Cook, L. (1980). The robustness of latent trait models and effects of test length and sample size on the precision of ability estimates. In D. J. Weiss (Ed.), **Proceeding of the 1979 Computerized Adaptive Testing Conference**, 36-52. Minneapolis, Minnesota: University of Minnesota, Department of Psychology. Psychometric Methods Program, Computerized Adaptive Testing Laboratory.

22 - Hambleton, R. K., & Traub,. R. (1971). Information curves and efficiency of three logistic test models**. British Journal of Mathematical and Statistical Psychology**, 24, 273- 281.

23 - Hambleton, R. K., & Traub, R. (1973). Analysis of empirical data using two logistic latent trait models. **British Journal of Mathematical and Statistical Psychology**, 26, 195- 211.

24 - Hambleton, R. K., Swaminatban, H. & Rogers. H. (1991). **Fundamentals of item response theory.** Thousand oaks. CA: Sage Publication.

25 - Hattie, J. (1985). Methodology review: assessing unidmensionality of tests and items. **Applied Psychological Measurement, 9**(2), 139 - 164.

26 - James, W.; Daniel, B.; Allen, C. & Young-Shun, L. (2002). Recovery of item- parameters in the nominal response model: a comparison of Marginal Maximum Likelihood estimation and Marrow Chain Monte carol Estimation. **Applied Psychological Measurement, 26**(3), 339-343.

27 - Kim, S. (2001). An evaluation of a Markov Chain Monte Carlo Method for the Rasch model. **Applied Psychological Measurement**, 25 (2), 163-176.

28 - Lewy, P, & Nielsen, A. (2002). Comparison of the frequent properties of Bayes and the Maximum Likelihood estimators in an age- structured fish stock assessment model.NRC Research Press web site at http:// cjfas. Nrc. Ca on February. Canada

29 - Linden, W. (1998). Bayesian item selection criteria of adaptive testing. **Psychometrika**, 63(2), 102-118.

30 - Lord, F. M. (1980). **Application of item response theory to practical testing problems**. Hillsadle, New Jersey: Lawrence Erlbaum Associates.

31 - Lord, F. (1986). Maximum likelihood and Bayesian parameter estimation in item response theory. **Journal of Educational Measurement, 23**, 157-162.

32 - Mark, P.; Hafidz, O.; & Michael, C. (2006). A comparison of Winsteps and Bilog- MG3 for vertical scaling with Rasch model. **Educational and Psychological Measurement, 64** (4), 600-616.

33 - Mckinley, R. & Reckase, M. (1981). **Comparison of Bayesian and a Maximum Likelihood tailored testing procedure.** Office of Naval Research, Arlington, VA, Personal and Training Research Programs Office.

34 - Michael, C.; Hafidz, O. & Mark, P. (2006). Vertical scaling with the Rasch model utilizing default and tight convergence settings with WInsteps and Bilog- MG3. **Applied Measurement in Education, 19**(2), 133-149.

35 - Mislevy, R. (1986). Bayes modal estimation in item response models. **Psychometrika, 51**, 177-195.

36 - Mislevy, R., & Bock, R. D. (2003). BILOG (Version 3) [Computer software]. Lincolnwood, IL: Scientific Software International**.**

37 - Reeve, B. (2004). Applications of item response theory (IRT) modeling for building an evaluating questionnaires measuring patient- reported outcomes. Web site: http:// outcomes.cancer.gov/conference/irt/reeve.pdf.

38 - Rondall, P. (2007). Estimating the standard error of the Maximum Likelihood ability estimator in adaptive using the posterior weighted test information. **Educational and Psychological Measurement, 67**(6), 958-975**.**

39 - Rosso, M.& Reckase, M. (1981**).** Comparison of a Maximum Likelihood and Bayesian ability estimation procedure for tailored testing**.** Paper presented at the annual meeting of the national council on measurement in education. Retrieved June 14, 2009, from <http:// www.eric.ed.gov/ERIC> We Portal.

40 - Stocking, M. K. (1990). Specifying optimum examinees for item parameter estimation in item response theory**, Psychometrica, 55**, 461-475

41 - Swaminatban, H. & Gifford, J. (1979**). Estimation of parameters in the three- parameter latent trait model** (Report No. 90). Amherst, Massachusetts: university of Massachusetts, school of education, laboratory of psychometric and evaluation research.

42 - Swaminathan, H., & Gifford, J. A. (1982). Bayesian estimation in the Rasch model. **Journal of Educational Statistics, 7**, 175-191.

43 - Swaminathan, H., & Gifford, J. A. (1985). Bayesian estimation in the two-parameter logistic model. **Psychometrika, 50**, 349-364.

44 - Swaminathan, H., & Gifford, J. A. (1986). Bayesian estimation in the three-parameter logistic model. **Psychometrika, 51**, 589-601.

45 - Trevisan, M. S., Sax, G. & Michael, W. B. (1991). The effect of number of options per item and student ability on test validity and reliability. **Educational and Psychological Measurement, 51**, 829-837.

46 - Wang, S. & Wang, T. (2001). Precision of RE warms weighted likelihood estimation for polytomous model in computerized adaptive testing. **Applied Psychological Measurement, 25**, 4, 317-331.

47 - Wang, S. & Wang, T. (2002). Relative precision of ability estimation in polytomous CAT: A Comparison under the generalized partial credit model and graded response model. **ED** 477926, shudon wang, 19500 Bulverede road, SanAntonion, TX 28259-3701.

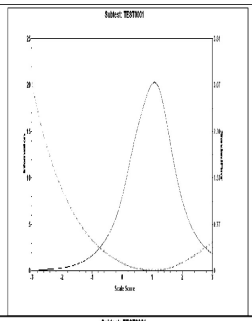
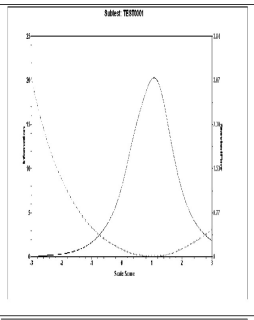
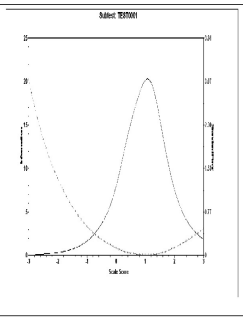
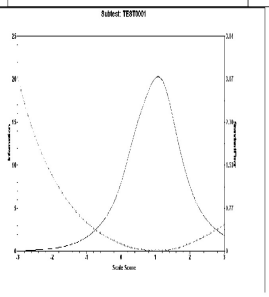
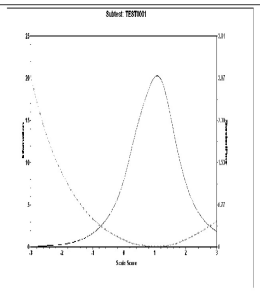
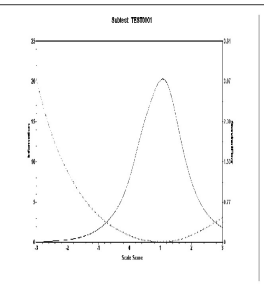
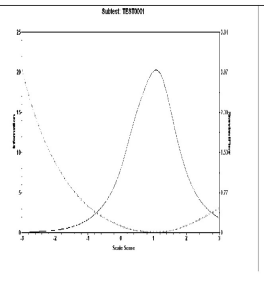
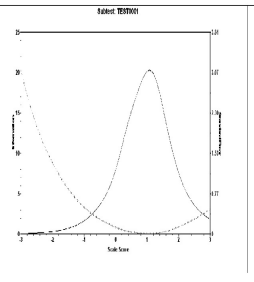
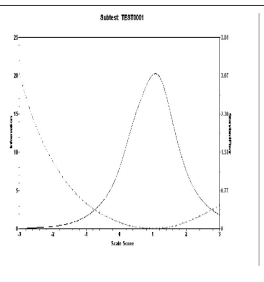
48 - Wang, T. & Vispoel, W. (1998). Properties of ability estimation methods in computerized adaptive testing. **Journal of Educational Measurement, 35**, 109- 135.

49 - Zeng, L. (1997). Implementation of marginal Bayesian estimation with four-parameter beta prior distributions**. Applied Psychological Measurement, 21**, 143-156.

50 - Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. (2003). Bilog Mg3[Computer Software].In M.du Toit (ed), **Irt from SSI: Bilog-MG, Multilog,** Parscale, Testfact. Lincolnwood, IL: Scientific Software International, Inc.

الملحق رقم (١)

منحنيات دالة المعلومات ومنحنيات الأخطاء المعيارية للاختبار تبعا لطرق التقدير وحجم العينة

حجم العينة	طريقة التقدير		
	الأرجحية العظمى	بييز تعظيم الاقتران	بييز توقع الاقتران
1000			
500			
100			

❖ الخط المتصل يمثل منحنى دالة معلومات الاختبار، والخط المتقطع يمثل منحنى الخطأ المعياري للاختبار، ويمثل المحور العمودي الأيسر قيم المعلومات، بينما يمثل المحور العمودي الأيمن قيم الخطأ المعياري.