# اللاتغير في تقديرات معالم فقرات الاختيار من متعدد عندما تطبق في سياقات مختلفة باستخدام نماذج النظرية الحديثة في القياس

**د. حابس سعد الزبون**

كلية التربية - جامعة الطائف

المملكة العربية السعودية

## الملخص

هدفت الدراسة إلى الكشف عن اللاتغير في تقدير معالم فقرات الاختيار من متعدد عندما تطبق في سياقات مختلفة، ولتحقيق أهداف الدراسة تم الاعتماد على اختبار مكون من (١١٣) فقرة من نوع الاختيار من متعدد في مادة الرياضيات بالإضافة إلى (١٥) فقرة مشتركة طبق على عينة تكونت من ٢٥١١ طالب وطالبة. وباستخدام البرامج المناسبة في تحليل البيانات تبين أن معايرة الفقرات ضمن الإطار الكلي للاختبار وعندما حللت لوحدها يؤدي إلى وجود استقرار في تقدير معالم الفقرات وهذه النتيجة متحققة لمعالم الفقرة الثلاثة: التمييز و الصعوبة، والتخمين تبعا لاختلاف النموذج والسياق باستثناء تقدير معلمة التمييز تبعا لاختلاف النموذج فقد وجدت فروق ذات دلالة إحصائية لصالح النموذج الثلاثي مقارنة مع النموذج الثنائي.

كما أظهرت قيم معامل الارتباط بين تقديرات معالم الفقرة: التمييز والصعوبة و التخمين التي تمت معايرتها بصورة مستقلة وضمن الإطار الكلي للاختبار أنها كانت عالية ودالة إحصائياً كما تبين أن قيم معاملات الارتباط كانت هي الأعلى عندما تمت معايرة الفقرات كعينة مستقلة (لوحدها) ويمكن التوصل من ذلك إلى انه يمكن الحصول على دقة أفضل عند معايرة عينة من الفقرات بشكل مستقل مقارنة بمعايرتها ضمن الإطار الكلي للاختبار، وبناء على ذلك توصي الدراسة باستخدام عينه ممثلة من الفقرات بعدد قليل من الفقرات في الاختبارات التكيفية وإجراء المزيد من الدراسات باستخدام دالة المعلومات كما توصي باستخدام النموذج الثلاثي عند معايرة الفقرات.

## مقدمة

تعتبر نظرية الاستجابة للفقرة (IRT) من النظريات السيكومترية المعاصرة التي اتضحت فائدتها في التغلب على الكثير من مشاكل القياس التي عجزت النظرية الكلاسيكية عن مواجهتها، والتي تفترض أنه يمكن التنبؤ بأداء الأفراد أو يمكن تفسير أدائهم في اختبار نفسي أو تربوي معين بناء على خاصية أو خصائص معينة مميزة لهذا الأداء تسمى السمات Traits، وتحاول هذه النظرية تقدير درجات الأفراد في هذه السمات. ونظراً لصعوبة ملاحظة هذه السمات بطريقة مباشرة أو قياسها بصورة مباشرة، فإنه يجب تقديرها أو الاستدلال عليها من خلال استجابات الأفراد الملاحظة على مفردات الاختبار الذى يقيس السمة أو القدرة المطلوبة، وتقوم نظرية استجابة الفقرة على مبدأين أساسين:

١ - أداء المفحوص على الاختبار أو الفقرة يمكن شرحه (عزوه) أو التنبؤ به بالرجوع إلى خصائص المفحوص التي تسمى قدرات أو السمات الكامنة.

٢ - توصف العلاقة بين ما هو ملاحظ (أداء المفحوص على الفقرة) وما هو غير ملاحظ (السمات والقدرات التي تقف وراء الأداء الملاحظ) باستخدام ما يسمى بمنحنى خصائص الفقرة (Item Charastarestic Curve).

ولهذه النظرية ثلاث ميزات رئيسية هي:

١ - تقدير معالم الفقرات مستقل عن مجموعة المفحوصين المأخوذة من مجتمع المفحوصين المصمم لهم الاختبار (Sample Free).

٢ - قدرة المفحوص تقدر بشكل مستقل عن مجموعة الفقرات التي تم اختيارها من مجتمع الفقرات، أي تقديرات القدرة للأفراد متحررة من خصائص الفقرات المستخدمة في تقدير القدرة (Item Free).

٣ - لكل تقدير للقدرة يوجد مؤشر للدقة وهذه الدقة تختلف من فرد إلى آخر (Hambleon and Swaminathan,1985).

وقد يبين لورد (Lord, 1980) أن تقدير صعوبة الفقرات لا يعتمد على طبيعة توزيع عينة الأفراد المستخدمة في ذلك وبناء على ذلك فقد أعطى أهمية كبيرة لخاصية اللاتغير (Invariance) في تقدير معالم الفقرات بتغيير مجموعة الأفراد، وبيّن أن هذه الخاصية هي من أهم الخصائص التي تميز نظرية استجابة الفقرة عن النظرية الكلاسيكية. وقد طورت نماذج رياضية مختلفة وذلك للتعبير عن العلاقة بين القدرة واحتمال الإجابة الصحيحة على الفقرة وتختلف هذه النماذج في صيغتها الرياضية وعدد معالمها كما تختلف من حيث عدد الافتراضات اللازم توافرها في البيانات المستخدمة.

وتقسم نماذج نظرية استجابة الفقرة إلى قسمين:

- النماذج ثنائية التدريج (Dichotomous IRT Models) منها النموذج أحادي المعلمة (One Parameter Model) والنموذج ثنائي المعلمة Two Parameters (Model) والنموذج ثلاثي المعلمة (Three Parameters Model).

- النماذج متعددة التدريج (Polytomous IRT Models) وتميز هذه النماذج بين المفحوصين الذين لديهم جزء من المعرفة والذين ليس لديهم معرفة ومنها نموذج الاستجابة المتدرجة (Graded Response Models GRM)، وهو امتداد للنموذج ثنائي المعلمة، ونموذج التقدير الجزئي (Partial credit model, PCM) وسلم التقدير (Rating Scale Models RSM) وهما امتداد للنموذج أحادي المعلم ونموذج الاستجابة الاسمية (Nominal Response Model NRM) الذي يناسب الاستجابات التي لايمكن تصنيفها في فئات رتبية (Embreston & Rise, 2000).

وليس بالضرورة أن يلائم نموذج رياضي كل البيانات فقد يلائم بعضها ولا يلائم بعضها الآخر لذلك عند استخدام أي نموذج رياضي من الضروري فحص مدى ملائمة النموذج للبيانات، وإذا لاءم النموذج الرياضي البيانات عندها فقط يمكن تقدير معالم الفقرة وقدرة الأفراد (Traub and Wolf, 1981; & Hambleton, 1985).

وللتحقق من مدى مطابقة النموذج للبيانات توجد ثلاث طرق للتحقق من ذلك:

١ - هل بيانات الاختبار تحقق افتراضات النموذج؟

٢ - هل خاصية اللاتغير (Invariance) متحققة لكل من معالم الفقرات والقدرة؟

٣ - هل توفر البيانات معلومات عن مدى قرب ماهو متنبأ به من خلال النموذج وبين ماهو ملاحظ؟ (Hamblton and Swaminathan,1985).

وفي هذه الدراسة تم الاعتماد على النماذج ثنائية التدريج والتي تقوم على عدة افتراضات منها(Hamblton and Swaminathan, 1985; Crocker and Algena, 1986; Embretson, 2000)

١ - افتراض أحادية البعد Unidimentionality:

يشير هذا الافتراض إلى أن هناك قدرة أو سمة واحدة فقط ضرورية لشرح اداء المفحوص على الاختبار وهذا الافتراض لا يمكن تحقيقه في كثير من الأحيان لوجود عوامل معرفية واختبارية وشخصية تتعلق بالمفحوص تؤثر على أدائه على الاختبار مثل القلق والدافعية، وعادةً نعتبر أن هذا الافتراض قد يتحقق إذا وجد هناك عامل مسيطر (Dominant Factor) وحيد يؤثر على أداء المفحوص على الاختبار وهذا العامل المسيطر عادةً يسمى بالقدرة التي يقيسها الاختبار.

وللكشف عن افتراض أحادية البعد يتم ذلك من خلال طريقتين:

أ - معظم الاجراءات التي تتبع للكشف عن أحادية البعد تستخدم إجراء التحليل العاملي بمعنى اخضاع العلامات على الفقرات للتحليل العاملي لمعرفة البنية العاملية لهذه الفقرات لمعرفة هل يوجد عامل رئيسي مسيطر.

ب - التوزيعات المشروطة لعلامات الاختبار عند كل مستوى من مستويات القدرة F(X/è) وتكون لمجموعات جزئية منتهية من المجتمع الكلي وتكون

متطابقة (Identical).

أما إذا كانت هذه التوزيعات متغيرة عبر المجتمعات الجزئية للمجتمع الكلي فهذا يعني أن الاختبار يقيس شيئاً آخر غير القدرة الوحيدة التي يقيسها أصلاً.

٢ - افتراض الاستقلال الموضعي (Local Independence):

يقصد بهذا الافتراض أن استجابات المفحوصين من نفس القدرة للفقرات المختلفة مستقلة عن بعضها بعضاً وبالتالي فإن استجابة المفحوص على فقرة لا تؤثر على أدائه على فقرة أخرى سلباً أو ايجاباً. ويمكن الكشف عن هذا الافتراض بما يلي:

أ - إذا وجد في ترتيب الفقرات تأثير على الأداء فمعنى ذلك أن افتراض الاستقلال الموضعي لا يتحقق.

ب - عدم استيفاء أحادية البعد أي إذا لم تكن فقرات الاختبار تقيس قدرة واحدة فإن افتراض الاستقلال الموضعي لا يتحقق.

٣ - منحنى خصائص الفقرة (Item Characteristic Curve):

هو عبارة عن دالة تربط بين احتمال النجاح في الإجابة على الفقرة والقدرة التي يقيسها الاختبار الذي يحتوي تلك الفقرة، وان هذه الدالة أو العلاقة تتخذ شكل المنحنى اللوغارتمي الذي يفترض أن يكون شكله كحرف S حيث يمثل قدرة المفحوص على المحور السيني واحتمال حصول المفحوص على الإجابة الصحيحة للفقرة التي تمثل المحور الصادي، أو هو معادلة انحدار غير خطي لاحتمال إجابة الفقرة إجابة صحيحة على القدرة أو السمة التي يقيسها الاختبار (Crocker and Algena,1986).

والفرق الرئيسي بين نماذج استجابة الفقرة الشائعة هو في الصيغة الرياضية لـ (ICC) ويستطيع باني الاختبار أن يختار الدالة الرياضية التي يراها مناسبة بعد مطابقة النموذج للبيانات. وإذا كان التوزيع الشرطي لعلامات الفقرة عند مستوى محدد من القدرة متماثل عند المجتمعات المختلفة فهذا يعني

أن (ICC) سوف يكون متماثل عند المجتمعات المختلفة وبالتالي تتحقق خاصة اللاتغير (Invariante)لـ (ICC) بتغير من يستجيب على الاختبار.

٤ - افتراض السرعة Speediness:

ينص هذا الافتراض على أن إجابة المفحوص على الفقرة سواء كانت صحيحة أم خاطئة يجب أن تعتمد على مقدار القدرة التي يقيسها الاختبار وليس على سرعته في الإجابة. فإذا اعتمد الاختبار على السرعة فإن هذا يؤدي إلى انتهاك افتراض أحادية البعد لأنه سوف يكون هناك قدرتان تؤثر على الأداء على الاختبار وهي السرعة والسمة التي يقيسها الاختبار.

## نماذج استجابة الفقرة ثنائية التدريج (Hambelton & swaminathan,1985):

- النموذج اللوجستي أحادي المعلمة (نموذج راش One Parameter Logistic Model) يعد هذا النموذج أبسط نماذج استجابة الفقرة ثنائية التدريج، يفترض هذا النموذج أن جميع الفقرات لها نفس القدرة التمييزية للتمييز بين المفحوصين، أي تكون جميع المنحنيات المميزة لفقرات الاختبار متوازية (متساوية في الميل) وتختلف عن بعضها البعض في نقاط التقائها بالمحور الافقي الذي يمثل متصل القدرة أو السمة الكامنة، كما يفترض عدم لجوء المفحوصين للتخمين العشوائي عند اجابتهم عن فقرات الاختبار، وهذا يعني إن الفقرات تختلف فقط في صعوبتها وتكون معادلة هذا النموذج كما يلي:

$$Pi(\theta) = \exp[Da(\theta-b))/(1+\exp(Da(\theta-b)]$$

حيث $P_i(\theta)$: احتمال اجابة الفقرة (i) إجابة صحيحة من قبل مفحوص له القدرة è تم اختياره عشوائياً من بين المفحوصين الذين لهم نفس القدرة.

D: ثابت مقداره ١,٧.

bi: معلمة الصعوبة للفقرة i.

e: الأساس اللوغاريتمي الطبيعي (٢,٧١٨).

a: تميز الفقرة i وهو قيمة ثابتة تساوي متوسط معاملات تميز الفقرات.

- النموذج اللوجستي ثنائي المعلمة (Two Parameter Logistic Model) اقترحه عالم الإحصاء بيرنبيوم (Birnbum) مع مجموعة من زملائه بجامعة كولومبيا الأمريكية ويفترض هذا النموذج أن الفقرات تختلف في صعوبتها وأيضاً تختلف في درجة التمييز بين المفحوصين وأن التخمين قليل جداً وتكون معادلة هذا النموذج كما يلي:

$$Pi(\theta) = \exp[Da\ (\theta\text{-}b))/(1 + \exp(Da\ (\theta\text{-}b)]$$

- النموذج اللوجستي ثلاثي المعلمة (Three parameter logistic model) يفترض هذا النموذج أن الفقرات تختلف في صعوبتها وأيضاً تختلف في درجة التمييز بين المفحوصين وأن مستويات التخمين لدى المفحوصين تختلف من فقرة إلى أخرى وتكون معادلة هذا النموذج كما يلي:

$$Pi(\theta) = ci + (1 + ci)\exp[Da(\theta\text{-}b))/(1 + \exp(Da(\theta\text{-}b)]$$

Ci: معلمة التخمين للفقرة i ويسمي الخط التقاربي الأدنى (Lower Asymptote) للمنحى المميز للفقرة، ويمثل احتمال الإجابة الصحيحة للأفراد ذوي القدرة المنخفظة.

والفكرة الأساسية لنماذج استجابة الفقرة في أنها تحاول اشتقاق قيم تقديرية للسمات التي تنطوي عليها مجموعة من الاستجابات لمجموعة من الفقرات، وعادة يفترض أن السمة المقاسة هى قدرة معينة أو خاصية من خصائص الفرد الذي يختبر بها، بحيث لا توجد علاقة منتظمة بين مستويات السمة المقاسة لدى أفراد مختلفين واحتمالات الاستجابة الصحيحة لمفردات مختلفة بمعنى آخر (علام، ٢٠٠٠).

وتعتبر خاصية اللاتغير في معالم الفقرات والقدرة حجر الزاوية في نظرية استجابة الفقرة وهذا ما يميزها عن النظرية الكلاسيكية في القياس، حيث تنص هذه الخاصية على أن تقدير معالم الفقرات مستقل عن توزيع مستويات القدرة

للمفحوصين الذين استخدمت إجابتهم في تقديرها وكما أن تقدير معلمة القدرة للمفحوصين لا يعتمد على مجموعة فقرات الاختبار المستخدمة في تقدير تلك القدرة.حيث وصف لورد (Lord, 1985) أن هذه الأهمية هي إحدى اكبر الصفات المهمة في النظرية الحديثة في القياس. وتحقق هذه الخاصية يسمح بإجراء تطبيقات هامة مثل معادلة الاختبارات، بنوك الأسئلة. وقد تم دراسة هذه الخاصية على معالم الفقرات عندما تطبق على مفحصوين مختلفين في القدرة، وحسب علم الباحث لم تدرس هذه الخاصية على الفقرات عندما تطبق في سياقات مختلفة، لذلك جاءت هذه الدراسة لمحاولة دراسة ذلك.

## مشكلة الدراسة

هدفت هذه الدراسة الى الكشف عن اللاتغير في تقدير معالم الفقرات ثنائية الاستجابة عندما تطبق في سياقات مختلفة، فقد درست هذه الخاصية للفقرات في حالة تغيير مجموعة الأفراد الذين يستجيبون على الفقرات. ونظراً لأن الفقرة يمكن أن تطبق في سياقات مختلفة خاصة في الاختبارات المكيّفة فإن هناك حاجة للاطمئنان على أن خصائص الفقرة تبقى مستقرة في هذه السياقات المختلفة، إذ أن الاختبارات قد تطبق ضمن اختبار قصير أحياناً وضمن اختبار مطول أحياناً أخرى، وقد تطبق ضمن اختبارات متجانسة جداً في السمة المقاسة أو ضمن فقرات أقل تجانساً في قياسها للسمة. وهذا يستدعي توفير بيانات عن خصائص الفقرات في هذه السياقات المختلفة. لذا فإن مشكلة الدراسة يمكن تلخيصها في الآتي:

- هل تختلف تقديرات معالم صعوبة الفقرات موضوع الدراسة تبعاً لاختلاف السياق ونموذج نظرية استجابة الفقرة المستخدم والتفاعل بينهما عندما تحلل لوحدها أو ضمن الإطار الكلي للاختبار وما وجهة هذا الاختلاف إن وجد؟
- هل تختلف تقديرات معالم التمييز للفقرات موضوع الدراسة تبعاً لاختلاف السياق ونموذج نظرية استجابة الفقرة المستخدم والتفاعل بينهما عندما تحلل لوحدها أو ضمن الإطار الكلي للاختبار وما وجهة هذا الاختلاف إن وجد ؟

- هل تختلف تقديرات معلمة التخمين للفقرات موضوع الدراسة تبعاً لاختلاف السياق عندما تحلل لوحدها أو ضمن الإطار الكلي للاختبار وما وجهة هذا الاختلاف إن وجد ؟

## أهمية الدراسة

إن لخاصية اللاتغير أهمية كبيرة عند مطوري الاختبارات الذين يستخدمون نماذج نظرية استجابة الفقرة حيث تعتبر هذه الخاصية بمثابة حجر الزاوية لنظرية استجابة الفقرة وأنه بدون هذه الخاصية سوف يتم مواجهة العديد من التعقيدات في الجانبين العملي والنظري، حيث إن هذه الخاصية تعتبر من أهم الصفات المميزة والمستخدمة بشكل واسع في مجالات القياس المختلفة كإحدى الصفات المميزة لنماذج نظرية استجابة الفقرة.

وقد ساهمت هذه الخاصية في حل الكثير من الصعوبات والمشاكل التي تواجه النظرية الكلاسيكية في القياس كالتي تواجه معادلة الاختبارات والاختبارات المكيفة لذا تأتي أهمية هذه الدراسة من دراستها لخاصية اللاتغير في تقدير معالم الفقرات ثنائية الاستجابة عندما تطبق في سياقات مختلفة لما لهذه الخاصية من أهمية كبيرة من حيث تقليص الوقت والجهد في تطبيق الاختبارات المختلفة وخاصة الاختبارات المكونة من عدد كبير من الفقرات.

## متغيرات الدراسة

- المتغيرات المستقلة في هذه الدراسة:

١- السياق الذي تطبق فيه الاختبارات.

٢- النموذج المستخدم.

- المتغيرات التابعة:

معالم الفقرات (الصعوبة، التمييز، التخمين).

## تعريف المصطلحات:

**معالم الفقرة (Item Parameters)**: قيم تحسب باستخدام دوال رياضية معينة لتقدير الخصائص الإحصائية للفقرة و تتضمن معلمة الصعوبة والتمييز والتخمين.

**معلمة الصعوبة (Threshold)** : هي قيمة القدرة التي يكون احتمال إجابة المفحوص عندها للسؤال إجابة صحيحة = $(1+c)/2$.

**معلمة التمييز (Slope)**: نسبة ميل منحنى خصائص الفقرة الذي يقابل النقطة التي تكون فيها علامة القدرة تساوي صعوبة الفقرة.

**معلمة التخمين (Asymptote)**: هي خط المقاربة الادنى Lower Asymptote من منحنى خصائص الفقرة ويمثل احتمالية إجابة المفحوصين ذوي القدرة المتدنية على الفقرة اجابة صحيحة.

**السياق (Context)**: يشير السياق إلى الوضع والشروط التي تطبق فيها فقرات الاختبار، فقد تطبق ضمن اختبار قصير أو طويل، وقد تطبق ضمن ترتيب مختلف للفقرات وقد يكون الأشخاص الذين يطبق عليهم الاختبار مختلفين.

## التصور النظري والأدبيات السابقة:

يوجد العديد من الدراسات التي تحدثت عن خاصية اللاتغير في تقدير معالم الفقرات ومن هذه الدراسات:

فقد قام ين (Yen, 1980) بإجراء دراسة لمعرفة أهمية اثر السياق على خصائص الفقرات عند استخدام النموذج الثنائي أو الثلاثي المعلمة، حيث تم اقتباس الفقرات من اختبار كاليفورنيا للتحصيل والذي يقيس المفاهيم الرياضية والاستعداد القرائي وتم تطبيق الاختبار على الصف الرابع والسادس، وقياس أثر السياق من خلال مقارنة عدم الاستقرار في معالم الفقرات التي تم إيجادها من

خلال نفس السياق حيث تم قياس عدم الاستقرار في معالم الفقرات من خلال تطبيق نفس الفقرات مرتبين على المفحوصين، وتغير السياق كان يتم بتغيير الترتيب للفقرات على الاختبار الثاني. وكانت من نتائج هذه الدراسة أن أثر السياق كان يظهر أكثر أهمية في معايرة الفقرات أكثر من اختيار الفقرات.

وفي دراسة قام بها هولن وآخرون (Hulin et al., 1982) التي هدفت إلى دراسة آثر حجم العينة وطول الاختبار على دقة تقدير قدرة الفرد ومعالم الفقرة باستخدام النموذج اللوغاريتمي ذي المعلمتين وذي الثلاث معالم، تم استخدام معامل الارتباط بين قدرة الفرد الحقيقية والمقدرة وبين معالم الفقرات الحقيقية والمقدرة واستخدام الجذر التربيعي لمتوسط مربعات الانحرافات كمعايير لتقييم دقة التقدير حيث تم استخدام أربعة مستويات لحجم العينة هي (٢٠٠، ٥٠٠، ١٠٠٠، ٢٠٠٠) فردا،وثلاثة مستويات لطول الاختبار هي (٦٠، ٣٠، ١٥) فقرة، وأشارت النتائج فيما يتعلق بالنموذج اللوغاريتمي ذي الثلاثة معالم أن معامل الارتباط بين المعامل الحقيقي لتمييز الفقرة والمعامل المقدر كان (٠,٣٦) عند استخدام اختبار مكون من (١٥) فقرة وحجم عينة (٢٠٠) فرد وكان معامل الارتباط يساوي (٠,٨٤) عند استخدام اختبار مكون من (٦٠) فقرة وحجم عينة مكونة من (٢٠٠٠) فرد كما وجد أن معامل الارتباط بين المعامل الحقيقي لصعوبة الفقرة والمعامل المقدر يساوي (٠,٧٢) عند استخدام الاختبار المكون من (١٥) فقرة وحجم عينة مكونة من (٢٠٠) فرد في حين ازداد هذه المعامل ليصبح (٠,٩٧) عند استخدام الاختبار المكون من (٦٠) فقرة وحجم عينة مكونة من (٢٠٠) فرد، وبشكل عام تشير النتائج إلى أنه بزيادة حجم العينة وطول الاختبار فإن دقة التقدير لقدرة الفرد ولمعالم الفقرة تزداد.

وفي دراسة هامبلتون وكوك (Hambleton and Cook, 1983) بعنوان جودة نموذج استجابة الفقرة وأثر طول الاختبار وحجم العينة على الدقة في تقدير القدرة؛ حيث بيّنا أن تقدير معالم الفقرات يتأثر بعدد المفحوصين المستخدمين في تقدير معالم الفقرة وأيضا بعدد الفقرات المستخدمة في الاختبار.

وفي دراسة لـ ستون ولان (Stone & Lane, 1991) بعنوان استخدام نماذج محددة لاستجابة الفقرة لفحص الاستقرار في تقدير معالم الفقرة مع الزمن، حيث تم تطبيق اختبار للرياضيات مكون من ١٩ فقرة على عينة من المفحوصين لإيجاد معالم الفقرات وبعد فترة من الزمن تم تطبيق الاختبار على المفحوصين وإيجاد معالم الفقرات، ووجدت الدراسة أن هناك درجة عالية من الاستقرار في معالم الفقرات في كلا الحالتين.

وفي دراسة لـ وي وولتر (Way& Walter,1992) بعنوان دراسة استكشافية للخصائص المرتبطة بخاصية اللاتغير في نظرية الاستجابة للفقرة وذلك لمحاولة دراسة الصفات المتعلقة في خاصية اللاتغير لمعالم الفقرات حيث تم الاعتماد على اختبار توفل وذلك لفحص خصائص اللاتغير في معالم الفقرات وقد تم التوصل إلى أنه أحياناً وبشكل قليل يتم عدم تحقيق هذه الخصائص وتم اعزاء ذلك لعدة أسباب منها قلة عدد المفحوصين وأثر موقع الفقرة.

وفي دراسة للتقي (١٩٩٢) بعنوان اللاتغير في تقدير معالم قدرات الأفراد ودرجات صعوبة أسئلة المقال من خلال نموذجي التقدير الجزئي وسلم التقدير كحالتين خاصتين من نماذج راش. وقد صممت هذه الدراسة من أجل فحص مدى تحقق افتراضيين أساسيين من افتراضات نظرية السمات الكامنة في أسئلة المقال وهذان الافتراضان هما:

١ - وجود لا تغير في تقدير قدرات الأفراد من خلال استعمال عينات مختلفة من الأفراد.

٢ - وجود لاتغير في تقدير قدرات الأفراد من خلال التقدم لاختبارات تتضمن أسئلة متفاوتة في درجات الصعوبة.

حيث تم فحص الافتراضين من خلال استعمال كل من نموذج التقدير الجزئي ونموذج سلم التقدير واستعمل في عملية الفحص أداة هي عبارة عن اختبار في الرياضيات وضع بهدف قياس قدرة طلاب الصف الأول الثانوي العلمي على حل المعادلات الجبرية والمثلثية في مجهول واحد و في مجهولين.

وقد وجد أن الافتراضين اللذين هما مجال البحث تحققا عن استعمال نموذج التقدير الجزئي. بينما لم يتحقق هذان الافتراضان عند استعمال نموذج سلم التقدير.

وقد بيّن هامبيلتون وجونس (Hambleton & Jones, 1994) في دراسة بعنوان : الأخطاء في تقدير معالم الفقرات وأثرها على دالة المعلومات للاختبار" حيث تم استخدام طرق المحاكاة (Simulation) وكما تم اختيار حجم عينة المعايرة لتكون في المرة الأولى (٥٠٠) مفحوص وفي المرة الثانية (٢٠٠٠) مفحوص حيث بيّنا وجود عاملين يؤثران على تقدير معالم فقرات الاختبار وبالتالي على دالة معلومات الاختبار، الأمر الذي يؤثر على دقة القياس وهما طول الاختبار و حجم عينة المفحوصين.

وفي دراسة لـ جونج وفلنكا (Jong & Fellyanka,1994) بعنوان حجم العينة وملائمة النموذج للبيانات، حيث حاولت هذه الدراسة التحقق من أثر حجم العينة على اللاتغير في تقدير معالم الفقرات حيث تم الحصول على اختبارات من اختبار ثم تطبيقه سابقاً وتم قسمة مجموعة من المفحوصين إلى عدد من العينات الجزئية وإجراء تحليل مستقل لكل مجموعة لوحدها وأيضاً تم أخذ اختبارات جزئية من الاختبار الرئيسي وذلك لدراسة خاصية اللاتغير في تقدير معالم الفقرات والقدرة بتغيير مجموعة الفقرات المستخدمة، وقد أشارت النتائج إلى وجود استقرار في تقدير معالم الفقرات والقدرة باختلاف مجموعة الفقرات الجزئية المستخدمة وأن معالم الفقرات والقدرة تعتمد على حجم العينة.

وفي دراسة لـ ليم وهون وآخرون (Lim et al., 1997) لبيان كيف يستطيع منهج تحليل الدراسات السابقة أن يقدم معلومات شمولية عن نظرية الاستجابة للفقرة لتوضيح أسئلة بحثية معينة، حيث ركزت الدراسة على تحليل أثر عدد الفقرات وعدد المفحوصين على دقة تقدير المعالم في النموذج اللوجستي الثنائي وتوصلت الدراسة إلى أن أكثر من ٦٣٪ من التغير في المؤشرات المستخدمة لقياس دقة تقديرات المعالم تعزى لهذين العاملين.

وفي دراسة لـ ين (Yen, 1999) بعنوان أثر عدم ملاءمة النموذج على خاصية اللاتغير في تقدير معالم الفقرات حيث تم في هذه الدراسة محاولة معرفة أثر عدم ملائمة النموذج للبيانات على تقدير معالم الفقرات وعلى تقدير قدرات المفحوصين. وذلك باستخدام ثلاث عينات من المفحوصين واختبارين لدراسة أثر عدم ملاءمة النموذج على خاصية اللاتغير في تقدير معالم الفقرات وعلى تقدير المفحوصين، وتوصلت الدراسة إلى أن استخدام النموذج الثلاثي أفضل من استخدام نموذج راش بالنسبة لمطابقة البيانات، وأن حجم عينة المفحوصين المستخدمة يؤثر على خاصية اللاتغير في معالم الفقرات. وأن هناك تغير في تمييز الفقرات عندما يتغير عدد المفحوصين، و اختلاف بسيط في درجة اللاتغير في تقدير معلمة القدرة باختلاف النموذج المستخدم.

وفي دراسة كيلكار وويتمان ولنشت (Kelkar, Wightman & Luencht, 2000) بعنوان تقييم اللاتغير في معالم نظرية استجابة الفقرة لاختبارات الدخول للكليات الطبية حيث تم استخدام نماذج أحادية المعلمة و النماذج ثنائية المعلمة وثلاثية المعلمة ومحاولة دراسة أثر الجنس والعرق واللغة على خاصية اللاتغير في معالم الفقرات وكذلك دراسة خاصية اللاتغير في معلمة القدرة عندما تطبق مجموعات مختلفة من الفقرات. وتوصلت الدراسة إلى أن معالم الفقرات والقدرة ثابتة وتحقق خاصية اللاتغير بالنسبة للجنس والعرق واللغة.

وفي دراسة فيتز باترك (Fitzpatric, 2001) هدفت إلى فحص اثر طول الاختبار وحجم العينة على عينات الاختبارات باستخدام نموذج التقدير الجزئي. وقد توصلت نتائج الدراسة إلى أن دقة تقدير المعالم عندما تم استخدام عينة حجمها (٢٠٠) مفحوص كانت أقل من دقة التقدير عند استخدام عينة حجمها (٥٠٠) مفحوص، وكانت دقة التقدير للمعالم عند استخدام عينة حجمها (٢٠٠) مفحوص وعينة حجمها (٥٠٠) مفحوص اقل من تقديرات المعالم عند استخدام عينة حجمها (١٠٠٠) مفحوص.

كما أجرى بومبلون وريتشي (Pomplun and Ritchie, 2004) دراسة بعنوان

بحث أثر السياق بعشوائية الفقرات في الاختبار الجزئي، حيث تم ترتيب الفقرات عشوائياً ضمن كل اختبار جزئي وتم استخدام الانحدار اللوجستي لمقارنة صعوبة الفقرات حيث توصلت الدراسة أنه لا يوجد أثر لترتيب الفقرات ولا لاستخدام العشوائية في ترتيب الفقرات أثر على معامل صعوبة الفقرات.

وفي دراسة (عبابنة، ٢٠٠٤) بعنوان أثر حجم العينة وطريقة انتقائها وعدد الفقرات وطريقة انتقائها على دقة تقدير معالم الفقرة والقدرة لاختبار قدرة عقلية باستخدام نظرية الاستجابة للفقرة. ولتحقيق أهداف هذه الدراسة تم إعداد اختبار قدرة عقلية مؤلف من أربعة اختبارات فرعية هي اختبار المفردات، اختبار المتشابهات، اختبار المتضادات، واختبار الحساب، حيث بلغ عدد فقرات الاختبار النهائية ٧١ فقرة، طبقت فقرات الاختبار على عينة مكونة من طلاب وطالبات الصف السابع من مدارس التربية والتعليم، وبلغ عدد أفراد عينة الدراسة ١٠٠٠ طالب وطالبة.

تم استخدام البرمجية Biolog 3.11 لتقدير قدرات المفحوص ومعالم فقرات الاختبار والاخطاء المعيارية في التقدير وإحصائيات مطابقة السياقات للنموذج اللوجستي الثلاثي الذي تم اعتمادها لإغراض هذه الدراسة. ومن النتائج التي تم التوصل إليها :

- تزداد الدقة في تقدير معلمة الصعوبة والقدرة عندما يكون مدى القدرة للمفحوص متوافقاً مع مدى صعوبة الفقرات. وكذلك تزداد الدقة في تقدير معلمة التخمين عند استخدام عينة من ذوي القدرة المتدنية في معايرة الفقرات. وتتسم تقديرات معالم الفقرة لعينة مختارة من الفقرات بالاستقرار النسبي عند معايرتها ضمن المقياس ككل أو بصورة مستقلة باستخدام نفس العينة من الأفراد
- تزداد الدقة في تقدير معالم الفقرة بزيادة حجم عينة المفحوصين.

وفي دراسة سادسكي (Sadesky, 2008) بعنوان الكشف عن الشروط المحددة على خاصية اللاتغير لتقدير معالم فقرات نماذج نظرية استجابة

الفقرة. حيث اعتمدت هذه الدراسة على بيانات مولدة و على النموذج ثلاثي المعلمة، وتم استخدام خمسة مستويات من القدرة (١,٥ - ٢، ٠,٠٥ - ١,٠) وإيجاد معالم الفقرات للمجتمع الكلي مقابل إيجاد معالم الفقرات للمجموعات الجزئية عند مستويات القدرة المختلفة وهذه المقارنة تمت عند زيادة الفروق بين مستويات القدرة في المجتمع الكلي والعينات الجزئية، وتغيير حجم العينة. وأشارت الدراسة إلى أن الأخطاء في تقدير معالم الفقرات للعينات تزداد عندما يوجد فرق في القدرة بين العينة والمجتمع وأن الأخطاء تقل عندما نزيد حجم العينات.

مما سبق يتبين أن الدراسة الحالية تتميز عن الدراسات السابقة بما يلي:

١ - الدراسات السابقة كانت تركز على اللاتغير في تقدير قدرات المفحوص ومعالم الفقرات باستخدام اختبارات جزئية مختلفة تطبق على نفس المفحوصين عند مستويات مختلفة من القدرة، وهذه الدراسة بحثت اللاتغير في تقدير معالم الفقرات لفقرات محددة عندما تحلل لوحدها في سياقات مختلفة وعندما تحلل ضمن الاختبار الكلي.

٢ - استخدمت هذه الدراسة ثلاث نماذج لاستجابة الفقرة في حين أن الدراسات السابقة الأخرى كانت تركز على النموذج الثلاثي وأيضا سوف تبحث الدراسة أثر النموذج واثر السياق وأثرهما معاً.

٣ - الدراسات السابقة كانت تحلل للاختبار الكلي وتأخذ مجموعات جزئية من الاختبار الكلي وتحللها وبالتالي كأنها أعادت تحليل فقرات الاختبار مره أخرى.

٤ - الدراسات السابقة كانت تأخذ مجموعات مختلفة من الفقرات وتحللها عند مستويات قدرة مختلفة وتقارن بينها وبين معالم الفقرات للاختبار الكلي عند نفس مستويات القدرة، بينما هذه الدراسة تركز على أثر السياق حيث إن هذه الدراسة أخذت نفس الفقرات (الفقرات ذات

الاهتمام) وتطبقها في ظروف مختلفة (على مفحوصين مختلفين وضمن سياقات مختلفة) وبعد ذلك يتم تطبيق الاختبار الكلي وعلى نفس المفحوصين الذين طبقت عليهم السياقات السابقة ونحلل جميع فقرات الاختبار الكلي، وبعدها نقارن معالم الفقرات ذات الاهتمام عندما حللت لوحدها في سياقات مختلفة وعندما حللت ضمن الاختبار الكلي.

وفي دراسة سكولز (Schulz, 2010) بعنوان اختبار اللاتغير في تقدير مؤشرات معالم الاستبيان باستخدام التحليل العاملي التوكيدي والنظرية الحديثة في القياس، حيث تم استخدام الطريقتين السابقتين لدراسة خاصية اللاتغير في معالم الفقرات بالاعتماد على البيانات التي تم الحصول عليها من الاختبار الدولي بيزا (PISA). وقد أظهرت نتائج الدراسة أن هناك مستوى عالي من الاستقرار في تقدير معالم الفقرات عند تقديرها بالاعتماد على نماذج النظرية الحديثة في القياس.

وفي دراسة ميرز وويلر ويلتر (Meyers, Miller & Way, 2010) التي كانت بعنوان موقع الفقرة والتتغير في صعوبة الفقرة بالاعتماد على تصميم المعادلة باستخدام النظرية الحديثة في القياس، وتمت دراسة معالم الفقرات ومهددات اللاتغير في معالم الفقرات عندما تظهر في مواقع مختلفة من الاختبار، بالاعتماد على النظرية الحديثة في القياس حيث تم الاعتماد على طريقة المحاكاة في تحليل البيانات بالاعتماد على نموذج راش. وأظهرت النتائج أن لتغيير مواقع الفقرات أثر على صعوبة الفقرة بالاعتماد على نموذج راش.

## مجتمع الدراسة

تكون مجتمع الدراسة من جميع طلبة الصف الثامن الأساسي من الذكور والإناث المنتظمين في المدارس الحكومية في المملكة الأردنية الهاشمية.

## عينة الدراسة

تكونت عينة الدراسة (٢٥١١) من طلبة الصف الثامن الأساسي في مديرتي عمان الاولى والثانية والبادية الوسطى وعين الباشا منهم (١١٨٥) طالباً وطالبة

طبق عليهم النموذجان الأول والثاني و (١٣٢٦) طالباً وطالبة طبق عليهم النموذجان الثالث والرابع، حيث تم اختيار (١٤) مدرسة من المدارس التي تضم الصف الثامن الأساسي في هذه المديريات، بالاعتماد على الطريقة العنقودية العشوائية.

## أداة الدراسة

لغرض تحقيق أهداف الدراسة اعتمد الباحث على اختبار معد من قبل وزارة التربية والتعليم مكون من (١١٣) فقرة من نوع الاختيار من متعدد في مادة الرياضيات بالإضافة إلى (١٥) فقرة مشتركة للصف الثامن الأساسي والذي تم تقسيمه إلى أربعة نماذج مكونة من (٥٠،٥١، ٣٦،٣٦) فقرة على التوالي بالإضافة إلى (١٥) فقرة مشتركة لكل نموذج وبترتيب مختلف لمواقع هذه الفقرات المشتركة وقد مرت عملية بناء الأداة بعدد من الخطوات يمكن تلخيصها على النحو الآتي:

١ - الرجوع إلى النواتج أو النتاجات التعلمية الخاصة بمبحث الرياضيات للصف الثامن الأساسي.

٢ - تشكيل فرق عمل لوصف مستويات الإتقان لكل نتاج تعلمي من النتاجات الجوهرية.

٣ - وضع أوصاف لما يجب على الطالب معرفته أو القيام به.

٤ - وضع فقرات اختبارية (أسئلة) لقياس كل مستوى من النتاجات التعلمية.

٥ - تم حساب الثبات باستخدام معادلة كرونباخ ألفا للاتساق الداخلي لنماذج الاختبار الأربعة وكانت على التوالي (٠,٨٤، ٠,٨٣، ٠,٨٤، ٠,٧٩).

## اجراءات الدراسة

للإجابة على أسئلة الدراسة قام الباحث بعمل الإجراءات التالية :

بعد القيام بالإجراءات الرسمية من موافقات وزيارات للمدارس المختارة والالتقاء بالمديرين والمديرات والمعلمين والمعلمات والطلبة لضمان التعاون في تنفيذ عملية تطبيق الاختبار بنماذجه الأربعة على الطلبة، تم تحديد مواعيد

تطبيق نماذج الاختبار خلال الفصلين الأول والثاني وتحديد جدول للمراقبة على الاختبار. وطبق النموذج الأول على (٥٦٢) طالباً وطالبة والنموذج الثاني على (٦٢٣) طالباً وطالبة في نهاية الفصل الأول، وطبق النموذج الثالث على (٦٢٢) طالباً وطالبة والنموذج الرابع على (٦٨٣) طالباً وطالبة في نهاية الفصل الدراسي الثاني من العام الدراسي ٢٠٠٨/٢٠٠٩. وقد تم حذف استجابات (٢١) مفحوصاً على النموذج الثالث لعدم الجدية في الاستجابة. وتمت عملية تصحيح فقرات النماذج الأربعة، وإدخال إجابات الطلبة في الحاسوب بالاعتماد على برمجية (SPSS) لأغراض التحليل.

## المعالجة الإحصائية

تم الاعتماد على برنامج Bilog-mg3 لتحليل استجابات الأفراد، حيث تضمنت عملية التحليل إيجاد معالم الفقرات بالاعتماد على النظرية الحديثة في القياس حسب النموذج الأحادي (راش) والثنائي والثلاثي،حيث تم إيجاد معلمة الصعوبة لفقرات نماذج الاختبار الأربعة وفق النموذج الأحادي والثنائي والثلاثي بعد تطبيقها على عينة من المفحوصين، كما تم إيجاد معلمة التمييز لفقرات النماذج الأربعة وفق النموذجين الثنائي والثلاثي بعد تطبيقها على عينة من المفحوصين مختلفة عن العينة الأولى، وأخيراً تم إيجاد معلمة التخمين لفقرات النماذج الأربعة وفق النموذجين الثلاثي بعد تطبيقها على عينة من المفحوصين مختلفة عن العينة الأولى والثانية. كما تم تحليل معالم الفقرات المشتركة لوحدها دون ارتباطها لفقرات الاختبار كاملة.

## النتائج ومناقشتها

**السؤال الأول: هل تختلف تقديرات معالم صعوبة الفقرات موضوع الدراسة تبعاً لاختلاف السياق ونموذج نظرية استجابة الفقرة المستخدم والتفاعل بينهما عندما تحلل لوحدها أو ضمن الإطار الكلي للاختبار وما وجهة هذا الاختلاف إن وجد ؟**

للإجابة عن هذا السؤال تم تقدير معلمة الصعوبة لجميع فقرات الاختبار الكلي وتقديرها للفقرات ذات الاهتمام في الحالتين عندما حللت لوحدها وعندما حللت ضمن الاختبار الكلي، تم استخدام تصميم القياسات المتكررة ثنائي العوامل، حيث يشير العامل الأول إلى السياق الذي تطبق فيه الفقرات والعامل الثاني إلى النموذج المستخدم، ويكون المتغير التابع هو تقدير صعوبة الفقرة، ويبين الجدول رقم (١) تحليل التباين الثنائي للمتوسطات الحسابية لمعلمة الصعوبة بالاعتماد على النماذج الثلاث.

**جدول رقم (١)**

**تحليل التباين الثنائي للمتوسطات الحسابية لمعالم الصعوبة عندما حللت لوحدها وضمن الإطار الكلي للاختبار**

| | مصدر التباين | مجموع المربعات | درجات الحرية | متوسط المربعات | قيمة F | الدلالة الإحصائية (sig) |
|---|---|---|---|---|---|---|
| عندما حللت ضمن السياق الكلي للاختبار | السياق | ١٠,٦٥ | ٣ | ٣,٥٥ | ٠,٨٠ | ٠,٥٠٠ |
| | النموذج | ٣١,٣٠ | ٢ | ١٥,٦٥ | ٣,٥١ | ٠,٣ |
| | التفاعل | ٦,٦٥ | ٦ | ١,١١ | ٠,٢٥ | ٠,٩٦ |
| | الخطأ | ٧٤٨,٩٥ | ١٦٨ | ٤,٤٦ | ٠,٣١ | |
| | المجموع | ٧٩٧,٥٥ | ١٧٩ | ٣١,٩٣ | ٣,٩٩ | |
| عندما حللت لوحدها | السياق | ٩٥,٧٦ | ٣ | ٤١٦,٨٣ | ٠,١٩ | ٠,٨٢ |
| | النموذج | ٨٣٣,٦٦ | ٢ | ١٩,٨٣ | | ٠,٢ |
| | التفاعل | ١١٨,٩٩ | ٦ | ١٠٤,٤٥ | | ٠,٩٨ |
| | الخطأ | ١٧٥٤٦,٨٠٠ | ١٦٨ | | | |
| | المجموع | ١٨٥٩٥,٢٤ | ١٧٩ | | | |

يتبين من الجدول رقم (١) أنه لا توجد فروق ذات دلاله إحصائية في معلمة الصعوبة عند مستوى دلالة ($\alpha$ = ٠,٠٥) تبعا لاختلاف السياق والنموذج والتفاعل بينهما عندما حللت الفقرات موضوع الاهتمام ضمن السياق الكلي

للاختبار وعندما حللت لوحدها وهذا يبين أن تقدير معلمة الصعوبة مستقل عن مجموعة المفحوصين المصمم لهم الاختبار والنموذج المستخدم في التحليل أي أن معالم صعوبة الفقرات تتمتع بخاصية اللاتغير (Invariance).

كما تم ايجاد معاملات الارتباط بين معاملات الصعوبة عند استخدام السياقات والنماذج المختلفة والجدول رقم (٢) يبين ذلك.

**جدول رقم (٢)**

**معاملات الارتباط بين معاملات الصعوبة تبعا لاختلاف النموذج السياق**

| | | عندما حللت لوحدها | | | عندما حللت ضمن الإطار الكلي للاختبار | | | | |
|---|---|---|---|---|---|---|---|---|---|
| تبعاً لاختلاف النموذج | | الأول | الثاني | الثالث | | الأول | الثاني | الثالث | |
| | الأول | | ٠,٩٢❖ | ٠,٨٤❖ | | | ٠,٧٣❖ | ٠,٦٤❖ | |
| | الثاني | | | ٠,٩٣ | | | | ٠,٢٧❖ | |
| تبعاً لاختلاف السياق | | الأول | الثاني | الثالث | الرابع | الأول | الثاني | الثالث | الرابع |
| | الأول | | ٠,٩٢❖ | ٠,٩٧❖ | ٠,٩٨❖ | | ٠,٥٣❖ | ٠,٤٣ | ٠,٥٢❖ |
| | الثاني | | | ٠,٨٩❖ | ٠,٨٩❖ | | | ٠,٣٢ | ٠,١٩ |
| | الثالث | | | | ٠,٩٩❖ | | | | ٠,٦٣❖ |

❖: معامل الارتباط ذو دلالة احصائية عند مستوى دلالة ٠,٠١

يتبين من الجدول رقم (٢) أن معظم معاملات الارتباط بين معاملات الصعوبة كانت عالية وذات دلالة إحصائية عند مستوى دلالة ($\alpha$ = ٠,٠١) وهذا يؤكد على خاصية اللاتغير في تقدير معلمة الصعوبة عند استخدام السياقات والنماذج المختلفة.

**السؤال الثاني: هل تختلف تقديرات معالم التمييز للفقرات موضوع الدراسة تبعاً لاختلاف السياق ونموذج نظرية استجابة الفقرة المستخدم والتفاعل بينهما عندما تحلل لوحدها أو ضمن الإطار الكلي للاختبار وما وجهة هذا الاختلاف إن وجد ؟**

للإجابة عن هذا السؤال تم تقدير معلمة التمييز لجميع فقرات الاختبار

الكلي وتقديرها للفقرات ذات الاهتمام في الحالتين عندما حللت لوحدها وعندما حللت ضمن الاختبار الكلي، وعليه تم استخدام تصميم القياسات المتكررة ثنائي العوامل، حيث يشير العامل الأول إلى السياق الذي تطبق فيه الفقرات والعامل الثاني إلى النموذج المستخدم، ويكون المتغير التابع هو تقدير تمييز الفقرة، ويبين الجدول رقم (٣) تحليل التباين الثنائي للمتوسطات الحسابية لمعلمة التمييز بالاعتماد على النموذجين ثنائي المعلمة والنموذج ثلاثي المعلمة.

**جدول رقم (٣)**

**تحليل التباين الثنائي للمتوسطات الحسابية لمعالم التمييز عندما حللت لوحدها وضمن الإطار الكلي للاختبار**

| | مصدر التباين | مجموع المربعات | درجات الحرية | متوسط المربعات | قيمة F | الدلالة الإحصائية (sig) |
|---|---|---|---|---|---|---|
| عندما حللت ضمن السياق الكلي للاختبار | السياق | ٠,٢٨٤ | ٣ | ٠,٠٩٥ | ٠,٩٨ | ٠,٤١ |
| | النموذج | ٦,٢٩ | ١ | ٦,٢٩ | ٦٥,١٨ | ٠,٠٠ |
| | التفاعل | ٠,١٣٨ | ٣ | ٠,٠٤٦ | ٠,٤٨ | ٠,٧٠ |
| | الخطأ | ١٠,٨١ | ١١٢ | ٠,٠٩٧ | | |
| | المجموع | ١٧,٥٣ | ١١٩ | | | |
| عندما حللت لوحدها | السياق | ٠,٩٨ | ٣ | ٠,٣٣ | ٢,٠٦ | ٠,١١ |
| | النموذج | ٢,٥١ | ١ | ٢,٥١ | ١٥,٨٢ | ٠,٠٠ |
| | التفاعل | ١,٣٣ | ٣ | ٠,٤٤ | ٢,٨٠ | ٠,٤٣ |
| | الخطأ | ١٧,٧٥ | ١١٢ | ٠,١٦ | | |
| | المجموع | ٢٢,٥٧ | ١١٩ | | | |

يتبين من الجدول رقم (٣) أنه لا توجد فروق ذات دلالة إحصائية في معلمة التمييز عند مستوى دلالة ($\alpha$ = ٠,٠٥) تبعا لاختلاف السياق والتفاعل بينهما عندما حللت الفقرات موضوع الاهتمام ضمن السياق الكلي للاختبار

وعندما حللت لوحدها وهذا يبين أن تقدير معلمة التمييز مستقل عن مجموعة المفحوصين المصمم لهم الاختبار والنموذج المستخدم في التحليل أي أن معالم تمييز الفقرات تتمتع بخاصية اللاتغير (Invariance)، كما يتبين وجود فروق ذات دلالة إحصائية في معلمة التمييز عند مستوى دلالة ($\alpha$ = ٠,٠٥) تبعا لاختلاف النموذج عندما حللت الفقرات موضوع الاهتمام لوحدها وعندما حللت ضمن الإطار الكلي للاختبار. وتم الاعتماد على اختبار توكي للمقارنات البعدية لبيان دلالة الفروق وتبين أن الفروق وفي الحالتين كانت لصالح النموذج الثلاثي مقارنة مع النموذج الثنائي.

كما تم إيجاد معاملات الارتباط بين معاملات التمييز عند استخدام السياقات والنماذج المختلفة والجدول رقم (٤) يبين ذلك.

**جدول رقم (٤)**

**معاملات الارتباط بين معاملات التمييز تبعا لاختلاف النموذج السياق**

| | | عندما حللت لوحدها | | | | عندما حللت ضمن الإطار الكلي للاختبار | | | |
|---|---|---|---|---|---|---|---|---|---|
| تبعا لاختلاف النموذج | | | الثاني | الثالث | | | الثاني | الثالث | |
| | الثاني | | | ٠,٥٧❖ | | | | ٠,٤٣❖ | |
| تبعا لاختلاف السياق | | الأول | الثاني | الثالث | الرابع | الأول | الثاني | الثالث | الرابع |
| | الأول | | ٠,١٥ | ٠,٤٩❖ | ٠,٥٢❖ | | ٠,٣٨❖ | ٠,٦١❖ | ٠,٦٣ |
| | الثاني | | | ٠,٣٠ | ٠,٣١ | | | ٠,٤٦ | ٠,٢٨ |
| | الثالث | | | | ٠,٥٥❖ | | | | ٠,٨١❖ |

❖: معامل الارتباط ذو دلالة احصائية عند مستوى دلالة ٠,٠١

يتبين من الجدول رقم (٤) أن معظم معاملات الارتباط بين معاملات التمييز كانت عالية وذات دلالة إحصائية عند مستوى دلالة ($\alpha$ = ٠,٠١) وهذا يؤكد على خاصية اللاتغير في تقدير معلمة التمييز عند استخدام السياقات والنماذج المختلفة.

**السؤال الثالث:هل تختلف تقديرات معلمة التخمين للفقرات موضوع الدراسة تبعاً لاختلاف السياق عندما تحلل لوحدها أو ضمن الإطار الكلي للاختبار وما وجهة هذا الاختلاف إن وجد ؟**

للإجابة عن هذا السؤال تم تقدير معلمة التخمين لجميع فقرات الاختبار الكلي وتقديرها للفقرات ذات الاهتمام في الحالتين عندما حللت لوحدها وعندما حللت ضمن الاختبار الكلي وتم الاعتماد على تحليل التباين الأحادي والجدول رقم (٥) يوضح نتائج التحليل.

**جدول رقم (٥)**

**تحليل التباين الأحادي للمتوسطات الحسابية لمعالم التخمين عندما حللت لوحدها وضمن الإطار الكلي للاختبار**

| | مصدر التباين | مجموع المربعات | درجات الحرية | متوسط المربعات | قيمة F | الدلالة الإحصائية (sig) |
|---|---|---|---|---|---|---|
| عندما حللت ضمن السباق الكلي للاختبار | بين المجموعات | ٠,٠٣٨ | ٣ | ٠,٠١٣ | ٠,٩٩ | ٠,٤٠ |
| | داخل المجموعات | ٠,٧٢ | ٥٦ | ٠,٠١٣ | | |
| | الكلي | ٠,٧٦ | ٥٩ | | | |
| عندما حللت لوحدها | بين المجموعات | ٠,٢١ | ٣ | ٠,٠١ | ٠,٤٥ | ٠,٧١ |
| | داخل المجموعات | ٠,٨٢ | ٥٦ | ٠,٠٢ | | |
| | الكلي | ٠,٨٤ | ٥٩ | | | |

يتبين من الجدول رقم (٥) أنه لا توجد فروق ذات دلالة إحصائية في معلمة التخمين عند مستوى دلالة ($\alpha$ = ٠,٠٥) تبعاً لاختلاف السياق عندما حللت الفقرات موضوع الاهتمام ضمن السباق الكلي للاختبار، وعندما حللت الفقرات موضوع الاهتمام لوحدها.

كما تم ايجاد معاملات الارتباط بين معاملات التخمين عند استخدام السياقات المختلفة والجدول رقم (٦) يبين ذلك.

**جدول رقم (٦)**

**معاملات الارتباط بين معاملات التخمين تبعا لاختلاف السياق**

| | | عندما حللت لوحدها | | | | عندما حللت ضمن الإطار الكلي للاختبار | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | الأول | الثاني | الثالث | الرابع | الأول | الثاني | الثالث | الرابع |
| تبعا لاختلاف السياق | الأول | | ٠,٥٥❖ | ٠,٦٩❖ | ٠,٥٥❖ | | ٠,٢٩ | ٠,٦٤❖ | ٠,٣٢ |
| | الثاني | | | ٠,٦١❖ | ٠,٩٥❖ | | | ٠,٣٥ | ٠,٣٥ |
| | الثالث | | | | ٠,٦٣❖ | | | | ٠,٧١❖ |

❖: معامل الارتباط ذو دلالة عند مستوى دلالة ٠,٠١

يتبين من الجدول رقم (٦) أن معظم معاملات الارتباط بين معاملات التخمين كانت عالية وذات دلالة إحصائية عند مستوى دلالة ($\alpha$ = ٠,٠١) وهذا يؤكد على خاصية اللاتغير في تقدير معلمة التخمين عند استخدام السياقات المختلفة.

مما سبق يتبين أن معايرة الفقرات ضمن الإطار الكلي للاختبار وعندما حللت لوحدها يؤدي إلى وجود استقرار في تقدير معالم الفقرات وهذه النتيجة متحققة لمعالم الفقرة الثلاثة: التمييز والصعوبة، والتخمين تبعا لاختلاف النموذج والسياق باستثناء تقدير معلمة التمييز تبعاً لاختلاف النموذج فقد وجدت هناك فروق ذات دلالة إحصائية لصالح النموذج الثلاثي مقارنة مع النموذج الثنائي.

كما أظهرت قيم معامل الارتباط بين تقديرات معالم الفقرة: التمييز والصعوبة و التخمين التي تمت معايرتها بصورة مستقلة وضمن الإطار الكلي للاختبار أنها كانت عالية ودالة إحصائياً كما تبين أن قيم معاملات الارتباط كانت هي الأعلى عندما تمت معايرة الفقرات كعينة مستقلة (لوحدها) ويمكن

التوصل من ذلك إلى أنه يمكن الحصول على دقة أفضل عند معايرة عينة من الفقرات بشكل مستقل مقارنة بمعايرتها ضمن الإطار الكلي للاختبار.

وهذه النتيجة تتفق مع ما أشار إليه جونج وفلنكا (Jong & Fellyanka, 1994) بأن هناك استقرار في تقدير معالم الفقرات والقدرة باختلاف مجموعة الفقرات الجزئية المستخدمة وأن معالم الفقرات والقدرة تعتمد على حجم العينة، كما تتفق مع ما أشار إليه ين (Yen, 1999) بأن استخدام النموذج الثلاثي أفضل من استخدام نموذج راش بالنسبة لمطابقة البيانات، وأن حجم عينة المفحوصين المستخدمة يؤثر على خاصية اللاتغير في معالم الفقرات.وان هناك تغير في تمييز الفقرات عندما يتغير عدد المفحوصين، واختلاف بسيط في درجة اللاتغير في تقدير معلمة القدرة باختلاف النموذج المستخدم.

وتتفق أيضا مع ما أشار اليه عبابنة (٢٠٠٤) والذي بيّن أن تقديرات معالم الفقرة لعينة مختارة من الفقرات بالاستقرار النسبي عند معايرتها ضمن المقياس ككل أو بصورة مستقلة باستخدام نفس العينة من الأفراد.

ولكنها لا تتفق مع ما أشار إليه هامبلتون وكوك (Hambleton and Cook, 1983) حيث بيّنا أن تقدير معالم الفقرات يتأثر بعدد المفحوصين المستخدمين في تقدير معالم الفقرة وأيضاُ بعدد الفقرات المستخدمة في الاختبار.

**التوصيات:**

١ - الاعتماد على النموذج الثلاثي عند إعداد بنوك الأسئلة.

٢ - استخدام عينة ممثلة من الفقرات بعدد قليل من الفقرات في الاختبارات التكيفية.

٣ - استخدام النموذج الثلاثي عند معايرة الفقرات.

٤ - إجراء المزيد من الدراسات باستخدام دالة المعلومات.

# The Invariance of Multiple-choice Item Parameters Estimates when Applied in Different Contexts Using Modern Test Theory Models

**Dr. Habes S. AL-Zaboon**
College of Education - ALTaef University
K.S.A

## Abstract

**The study aims to study the Invariance of multiple-choice item parameters estimates when applied in different contexts. To achieve the goals of the study, the researcher used a test consisting of (113) items multiple-choice type in mathematics in addition to (15) common items applied on a sample of 2511 students.**

**Using the suitable software to analyze data, the results of the analysis showed that the items calibration within all items of the test, when analyzed alone, leads to invariance in the items parameters estimate and this result achieved for the flowing items parameters: Discrimination and difficulty, and guessing depending on the different model and contexts, except estimating a discrimination parameter depending on the different model, has been found. There were significant differences in favor of the Three-parameter model compared with two-parameter model.**

**As shown by the values the correlation coefficient between the estimates of items parameters: Discrimination and the difficulty and guessing that has been calibrated independently and within the all items of the test they were high and statistically significant as shown by the values correlation coefficients were high when it was calibrated item sample independently (alone).**

**This value can be reached from that we get better accuracy when calibrating a sample of items independently compared within calibrating within the all items of the test.**

**Based on the results of the study, it was recommended to use a representative sample of items with a few items in the tests of adaptive and further studies using a information function also recommends using the Three-parameter Model, when we calibrate the items and when we design and development the item banking.**

# المراجع


١ - التقي، أحمد (١٩٩٢). اللاتغير في تقدير معالم قدرات الأفراد ودرجات صعوبة أسئلة المقال من خلال نموذجي التقدير الجزئي وسلم التقدير كحالتين خاصتين من نماذج راش. رسالة دكتوراه غير منشورة. الجامعة الأردنية، الأردن.

٢ - صلاح الدين، علام (٢٠٠٠). **القياس والتقويم التربوي والنفسي - أساسياته وتطبيقاته وتوجهاته المعاصرة**. القاهرة، دار الفكر العربى.

٣ - عبابنة، 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، الأردن.

4 - Crocker, L. & Algina, J. (1986). **Introduction to classical and modern test theory**. New York: CBS College Publishing.

5 - Embretson, S. & Reiase, S. (2000). **Item Response Theory for Psychologists**. New Jersey: Lawrence Erlbaum Associates, Inc..

6 - De Jong, John H. A. L.; & Stoyanova, Fellyanka (1994). Theory Building: Sample Size and Data - Model Fit. Paper presented at the Annual Language Testing Research Colloquium (March). ERIC No: Ed 451001.

7 - Fitzpatrick, Ann R. (2001). The effects of test length and sample size on the reliability constructed - response items. **Applied Measurement in Education**, 14(1), 31-57.

8 - Greg, S. (2008). Exploring boundary condition of the invariance property of item parameter estimates of item response models. Report - research. ERIC No: Ed 479923.

9 - Hambleton, R., & Swaminathan, H. (1985). **Item Response Theory :Principles and applications**. Boston MA: Kluwer-Nyjhoff.

10 - Hamleton, R.H. & Jones, R. W. (1994). Item Parameter estimation errors and their Influence on test information function. **Applied Measurement in education**. 7 (3), 171 - 186.

11 - Hambleton, R. K. & Cook. L. L. (1983). Robustness of item response models and effects of test length and sample size in the precision of ability estimates, In D. J.. Weiss (ed.), **New horizons in Testing**. Pp.31-49, New York.

12 - Hulin, G.L., Lissak, R.I. & Drasgow,F. (1982). Recovery of two and three - Parameter Logistic item characteristic curves: A Monte Carlo study. **Applied Psychological Measurements**, 6, 249 - 260.

13 - Jone, H. A & Fellyanka. (1994). Sample size and data model fit. Paper presented at annual language testing research colloquium. ERIC No: Ed 451001.

14 - Kelkar V.; Wightman, L. F. C.; & Luencht, R. M. (2000). Evaluation of IRT Parameter Invariance Property for the MCAT, the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.

15 - Meyers, J.L.; Miller, J. E; & Way, W. D. (2010). Item Position and Item Difficulty Change in an IRT-Based Common Item Equating Design. **APPLIED** Measurement in Education, 22: 38-60.

16 - Lim, T.; Hoon, B.; & Shardwell, Michael (1997). Effects of number of items and examines on parameter estimation in item response theory. A research synthesis. Paper presented at the annual meeting of American educational Research Association.

17 - Lord, F. (1980). **Application of Item Response Theory to Practical Testing Problems**. NJ: Larence Erlbum Associates.

18 - Pomplun, Mark; Ritchie, Timothy (2004). An Invistigation of context effect for item randomization test. **Journal of Education Computing Research**, 30(3), 243-254.

19 - Schulz, W. (2010). Testing Parameter Invariance for Questionnaire Indices using Confirmatory Factor Analysis and Item Response Theory. Paper prepared for the Annual Meetings of the American Educational Research. Association in San Francisco, 7-11 April 2005.

20 - Stone,C.A. & Lane, S. (1991). Use of Restricted Item Response Theory Models for Examining the Stability of Item Parameter Estimates Over Time. **Applied Measurement in Eeducation**, 4(2), 123-141.

21 - Traub, R. E. & Wolfe, R.G. (1981). Latent trait theories and the assessment of educational achievement. In D.C Berliner (ED). **Review of Research in education**: vol.9. Washington, DC., American Educational Research Association.

22 - Way, Walter D. (1992). An exploratory study of characteristics Related to IRT parameter Invariance a foreign language Toefl Technical Report. ERIC No: Ed 390890.

23 - Yen, W.M. (1980). The extent, cause and Importance of context effects on Item parameters for two latent trait models. **Journal of Educational Measurement**, 17(4), 297-311.

24 - Yin, P. & Xitao, F. (1999). Assessing the effect of model-data misfit on the invariance property of IRT parameter estimates. Paper presented at the annual meeting of the American educational research association. (Montreal, Quebec, April 19 -23). ERIC No: Ed 430050.