

# مقارنة بين النظرية الكلاسيكية والنموذج ثلاثي المعلم في انتقاء فقرات اختبار تحصيلي للغة الإنجليزية

رنا ثاني ضامن المومني  
أستاذ مشارك في القياس والتقويم  
جامعة القصيم - السعودية

رحاب بنت صالح الصيخان  
ماجستير في القياس والتقويم  
جامعة القصيم - السعودية  
rehab-alsaikhan@hotmail.com

قبول البحث: 2021/4/3

مراجعة البحث: 2021 /2/20

استلام البحث: 2021 /2/3

DOI: <https://doi.org/10.31559/EPS2021.10.1.8>



file is licensed under a [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/)

## مقارنة بين النظرية الكلاسيكية والنموذج ثلاثي المعلم في انتقاء فقرات اختبار تحصيلي لغة الإنجليزية

رحاب بنت صالح الصيخان

ماجستير في القياس والتقويم- جامعة القصيم- السعودية  
rehab-alsaikhan@hotmail.com

رنا ثاني ضامن المومني

أستاذ مشارك في القياس والتقويم- جامعة القصيم- السعودية

استلام البحث: 2021/2/3 مراجعة البحث: 2021/2/20 قبول البحث: 2021/4/3 DOI: <https://doi.org/10.31559/EPS2021.10.1.8>

### الملخص:

هدفت هذه الدراسة إلى المقارنة بين النظرية الكلاسيكية والنموذج ثلاثي المعلم لنظرية استجابة الفقرة في انتقاء فقرات اختبار تحصيلي لمادة اللغة الإنجليزية. ولتحقيق أهداف هذه الدراسة تم بناء اختبار تحصيلي لمادة اللغة الإنجليزية للصف الأول ثانوي مكون بصورته النهائية من (42) فقرة من نوع الاختيار من متعدد لأربعة بدائل، وشملت عينة الدراسة على (1076) طالبة تم اختيارهم بالطريقة العشوائية العنقودية متعددة المراحل من (11) مدرسة للبنات.

وقد تم استخدام البرامج الإحصائية (4) (SPSS , Cocron, AMOS , IRT-PRO) لتحليل بيانات الدراسة، وبعد التحقق من افتراضات نظرية استجابة الفقرة فقد طبقت (36) فقرة للنموذج ثلاثي المعلم، وهو ما نسبته (85.71%) من فقرات الاختبار، وطابقت (38) فقرة لنظرية الاختبار الكلاسيكية، وهو ما نسبته (90.45%) من فقرات الاختبار.

وفيما يتعلق بالخصائص السيكومترية للاختبار (الصدق والثبات)، فقد بلغ قيمة معامل الثبات للنظرية الكلاسيكية بإستخدام معامل الثبات كودر-ريتشاردسون (KR-20) (0.89)، بينما بلغ معامل الثبات التجريبي للنموذج ثلاثي المعلم (0.78)، وتم تقدير معامل الصدق المحكي لكلا النظريتين من خلال حساب معامل ارتباط بيرسون بين نتائج الطالبات وتحصيلهن في مادة اللغة الإنجليزية، فكانت قيمة معامل الصدق المحكي في نظرية استجابة الفقرة (0.59)، بينما بلغت قيمة معامل الصدق المحكي للنموذج ثلاثي المعلم بعد حذف الفقرات غير المطابقة (0.60). وكشفت نتائج هذه الدراسة عن وجود فروق ذات دلالة إحصائية بين معاملي الثبات بين النظرية الكلاسيكية والنموذج ثلاثي المعلم لصالح النظرية الكلاسيكية، وفروق ذات دلالة إحصائية بين معاملي الصدق المحكي بين النظرية الكلاسيكية والنموذج ثلاثي المعلم لصالح النموذج ثلاثي المعلم.

كما أشارت هذه الدراسة إلى المميزات التي تتمتع بها نظرية استجابة الفقرة أثناء تحليل الفقرات؛ إذ توفر معلومات أكثر عن أداء الفقرات عند كل مستوى من مستويات القدرة، مما يساهم في استخدامها في بناء وتطوير مختلف الإختبارات في المجالات التعليمية.

الكلمات المفتاحية: نظرية الاختبار الكلاسيكية؛ نظرية استجابة الفقرة؛ النموذج اللوغاريتي ثلاثي المعلم؛ انتقاء الفقرات؛ الخصائص السيكومترية.

## 1. المقدمة:

تعددت أدوات القياس والتقويم وأساليب استخدامها في مختلف المجالات التربوية والنفسية والمهنية، واكتسبت أهميتها كعنصر أساسي في تلك المجالات بشكل عام وفي المجال التربوي بشكل خاص، نظراً لإسهامها في توجيه البرامج التربوية وعملياتها واتخاذ قرارات بشأن أفرادها. وقد شهد مجال القياس والتقويم خلال العقود الماضية تطوراً هائلاً في أسسه النظرية وأساليب التحليل المختلفة، وظهور نظريات حديثة تساعد على تجاوز الكثير من المشكلات المتعلقة ببناء الاختبارات والمقاييس، يواكبه تطور وانتشار العديد من البرامج الإحصائية المعنية بهذا المجال، مما جعل عملية القياس أكثر مرونة وملاءمة لطبيعة الظواهر السلوكية، وهو ما ساهم في تقدم علم القياس التربوي والنفسية. (علام، 2011)

وتُعد الاختبارات بمختلف أشكالها من أبرز أدوات القياس المستخدمة في العملية التربوية، فزاد الاهتمام بتصميم وبناء الاختبارات وطريقة اختيار فقراتها، وتحري الدقة في خصائصها السيكمومترية، وتبني الآليات والاستراتيجيات الحديثة في البناء والتحليل، الأمر الذي يزيد من درجة الوثوقية في نتائجها، ومحاولة فهم وتقييم هذه النتائج، فالاختبار الجيد هو الذي يوضِّح العلاقة بين خصائص الفقرات وقدرة الطالب اثناء الإجابة عليها. (Hambelton & Jones, 1993)

وعند تتبع التطور في مجال القياس النفسي والتربوي، نجد أن نظرية الاختبار الكلاسيكية [Classical Test Theory (CTT)] من أولى النظريات التي تم تبنيها، وهي نظرية تمتاز ببساطتها وسهولة تطبيقها، فقد شاع استخدامها في بناء الاختبارات منذ ظهورها في هذا المجال، بإعتمادها في تقديرها للدرجة الحقيقية على استخدام الدرجة الملاحظة التي يحصل عليها المفحوص في الاختبار؛ فوفقاً لهذه النظرية يمكن التعبير عن الدرجة الحقيقية للمفحوص والتي يمكن التنبؤ بها من أدائه على الاختبار من خلال المعادلة:

$$X=T+E$$

حيث أن:

$X$  = الدرجة الملاحظة للمفحوص على الاختبار.

$T$  = الدرجة الحقيقية للمفحوص.

$E$  = خطأ القياس (Courville, 2004).

تبعاً لهذه المعادلة فإن الدرجة الملاحظة للمفحوص هي حاصل جمع خطي لدرجة المفحوص الحقيقية والخطأ العشوائي، وبما أن الدرجة الحقيقية وخطأ القياس العشوائي مكونات غير ملاحظة؛ فاعتبرت هذه النظرية أن الدرجة الحقيقية هي معدل الدرجات التي يحصل عليها المفحوص على أدائه لعدة اختبارات متكافئة، وأن أخطاء القياس هي أخطاء عشوائية، كما تفترض عدم ارتباط الدرجة الحقيقية بخطأ القياس، وأن متوسط أخطاء القياس يساوي صفراً، وكذلك عدم ارتباط أخطاء القياس فيما بينها (Alharbi, 2003).

كما تهتم نظرية الاختبار الكلاسيكية بإحصائيات الاختبار بشكل عام (كالصدق والثبات)، وتتيح حساب كل من خصائص الفقرة كالصعوبة والتميز بسهولة (Fan, 1998)، إلا أن هذه الخصائص تعتمد في تقديرها على عينة المفحوصين والعكس صحيح، وهو ما يجعل استخراج الدرجة الحقيقية أمراً في غاية الصعوبة، بالإضافة إلى صعوبة تعميم نتائج الاختبار. (Hambelton & Jones, 1993)

وقد عانت النظرية الكلاسيكية من بعض نقاط القصور نظرياً وعملياً؛ فقد تبنت في الجانب النظري افتراضات سهلة نسبياً ولا يمكن التحقق منها في بعض الأحيان، وركزت بشكل أساسي على الدرجة الكلية للاختبار للوصول إلى الدرجة الحقيقية والتي تمثل قدرة المفحوص، كما عبّرت عن الدرجة الحقيقية بأنها متوسط الدرجة التي يحصل عليها المفحوص من أدائه لاختبار معين عدة مرات، والذي بدوره سيؤدي إلى مشكلات عملية تتضمن تأثير خصائص الاختبار على تحديد هذه القدرة، وصعوبة تكرار نفس الاختبار وتحت نفس الظروف لتحديد الدرجة الحقيقية (Hambelton & Van der Linden, 1982). بالإضافة إلى صعوبة تحقق تكافؤ الاختبارات وتساوي أخطاء القياس على أرض الواقع (Spela & Gregor, 2008)، ولعل المشكلة الأساسية للنظرية الكلاسيكية تتلخص بعدم استقلالية إحصائياتها لكل من المفحوص والفقرات؛ فإحصائيات المفحوص المتمثلة بقدرته والتي تعكسها درجته الظاهرة غير مستقلة عن إحصائيات عينة الفقرات التي يشتمل عليها الاختبار، كما أن إحصائيات الفقرات كالصعوبة والتميز تعتمد على عينة المفحوصين الذين تطبق عليهم تلك الفقرات، مما أدى إلى قصور النظرية الكلاسيكية في بعض تطبيقات القياس كمعادلة الاختبارات والاختبارات التكيفية المحوسبة. (Abedalaziz & Leng, 2013)

فظهرت تبعاً لذلك نظرية استجابة الفقرة [Item Response Theory (IRT)] كمنعى جديد يعمل على سد الثغرات التي عانت منها النظرية الكلاسيكية، والوصول إلى درجة عالية من الدقة والاستقلالية الأزمن للقياس، فحازت على إهتمام كبير من العلماء في مجال تقدير القدرة واختبار الفقرات وتحديد خصائصها؛ فهي نظرية تحاول الكشف عن قدرات المفحوصين وسماتهم الكامنة من خلال استجاباتهم على فقرات الاختبار، وذلك بإيجاد العلاقة بين هذه الاستجابات وخصائص الفقرات على متصل القدرة (ability scale)، بحيث تستقل خصائص القدرة للمفحوصين عن خصائص عينة الفقرات والعكس، وهي بذلك تحقق خاصية الاستقلالية لكل من إحصائيات المفحوص وإحصائيات الفقرة، وهو جانب مهم من جوانب القياس قد تفوقت به نظرية استجابة الفقرة على النظرية الكلاسيكية. (Hambelton & Jones, 1993)

تقوم نظرية استجابة الفقرة على عدة افتراضات أساسية تبعاً لما أشار إليه (Allen & Yen, 1979). وهي كالتالي:

- أحادية البعد (Unidimensionality): ويقصد بذلك أن جميع الفقرات تقيس قدرة أو سمة واحدة.
- الاستقلال الموضوعي (Local Independence): ويقصد به أن إجابة المفحوص على أي من فقرات المقياس مستقلة إحصائياً عن إجابته على بقية الفقرات.

- منحني خصائص الفقرة (Item Characteristic Curve): والذي يُفترض أن يكون شكله لوجستياً.
- التحرر من السرعة (Speededness): أي أن عامل الوقت لا يجب أن يؤثر على استجابات المفحوصين، وأن عدم استجاباتهم يعود لمستوى قدرتهم وليس لضيق الوقت.

وقد تميزت نظرية استجابة الفقرة عند بناء الاختبار واختيار الفقرات بوضعها لمعايير ساهمت في تقدم بناء الاختبارات بشكل واضح؛ وذلك من خلال إيجاد دوال رياضية تربط بين كلاً من خصائص المفحوصين والفقرات للوصول إلى احتمالية الاستجابة الصحيحة للمفحوص بالسمة أو الخاصية التي يقيسها الاختبار، فهي تعتمد على منحني خصائص الفقرة بشكل رئيسي (Item Characteristic Curve) والتي تفترض هذه النظرية الشكل اللوجستي له (Spela & Gregor, 2008). بالإضافة إلى ارتكازها على دالة معلومات الفقرة (Item Information Function)، ودالة معلومات الاختبار (Test Information Function) (المومني، 2011). مما جعلها تتميز بخصائص تفسيرية لاستجابات المفحوص على فقراتها تتجاوز هذه التفسيرات إيجاد مجموع درجات المفحوص على الفقرات.

ويندرج تحت هذه النظرية ثلاثة نماذج لوجستية رئيسية للإستجابات الثنائية (1, 0)، والتي ساهمت في حل عدة مشاكل مثل: تكافؤ الإختبارات، وبناء اختبارات توائم مستوى قدرة المفحوص، وإيجاد أساس إحصائي يُعتمد عليه في بناء الاختبارات، وتشمل هذه النماذج: النموذج أحادي المعلم (One-Parameter Model) والمعروف بنموذج راش (Rasch Model) والذي يفترض اختلاف الفقرات في صعوبتها، والنموذج ثنائي المعلم (Two-Parameters Model) والذي تختلف فيه فقرات الاختبار في كل من صعوبتها وتميزها، والنموذج ثلاثي المعلم (Three-Parameters Model) والذي سوف تركز عليه هذه الدراسة، نظراً لأنه يُتيح اختلاف الفقرات في كل من معلم الصعوبة (Difficulty Parameter)، ومعلم التمييز (Discrimination Parameter)، ومعلم التخمين (Pseudo-guessing Parameter)، والذي تظهر أهميته عند استخدام فقرات اختبار من نوع الإختيار من متعدد، (Adedoyin & Adedoyin, 2013)

ونظراً لاختلاف كلا النظريتين في افتراضاتها الأساسية، وطريقة بنائها وتحليلها لفقرات الإختيار ومعالجتها للنتائج فمن المتوقع أن يكون هناك اختلافات بين النظريتين بناءً على إحصائيات كل من المفحوصين والفقرات. (Fan, 1998)

ففي دراسة أجراها كل من جولر ويانيك وتيكر (Güler, Uyanik & Taker, 2013) تمت مقارنة معالم الفقرات بين النظرية الكلاسيكية ونظرية استجابة الفقرة بنماذجها الثلاث (الأحادي، والثنائي، والثلاثي)، وذلك باستخدام اختبار القبول في المرحلة الثانوية والمكون من (25) فقرة من نوع الإختيار من متعدد تم تطبيقه على عينة عشوائية مكونة من (1250) طالب. وقد أظهرت نتائج الدراسة أن أفضل مطابقة للبيانات كانت للنموذج ثلاثي المعلم، كما أن معالم الفقرات لهذا النموذج حققت أدنى ارتباط بينها وبين المعالم المناظرة لها باستخدام النظرية الكلاسيكية، واستخلصت الدراسة تشابهاً في استخدام النظرية الكلاسيكية للنموذجين الأحادي والثنائي لنظرية استجابة الفقرة، واختلاف دال إحصائياً بين النظرية الكلاسيكية والنموذج ثلاثي المعلم وخاصة في حال ارتفاع احتمالية التخمين.

وقد أجرى كل من عبدالعزيز ولينغ (Abedalaziz & Leng, 2013) دراسة قارنت بين كل من صعوبة وتميز الفقرات لمقياس القدرة الرياضية والمكون من (74) فقرة باستخدام نظرية الاختبار الكلاسيكية والنماذج الثلاثة لنظرية استجابة الفقرة (الأحادي، والثنائي، والثلاثي). وقد تم تطبيق أداة الدراسة على عينة مكونة من (602) طالب من طلاب الصف العاشر، وقد أظهرت النتائج أن النموذج ثلاثي المعلم كان الأفضل في مطابقة البيانات، كما أظهرت النتائج بأن المؤشرات الناتجة من النموذج ثلاثي المعلم كانت الأكثر مقارنة لمؤشرات النظرية الكلاسيكية، وقد أوصت الدراسة بإمكانية استخدام النظرية الكلاسيكية ونماذج نظرية استجابة الفقرة بشكل منفصل أو مشترك لوصف طبيعة خصائص الفقرات.

وقام كيانجتي (Kyungtae, 2015) بدراسة لاختبار مدى قوة وضعف النماذج السيكومترية للتحليل ممثلةً بالنظرية الكلاسيكية ونظرية استجابة الفقرة في تحليل فقرات اختبار القراءة والفهم والمكون من (22) فقرة على عينة مكونة من (10897) طالب. وقد أظهرت النتائج أن النموذج ثلاثي المعلم حقق أفضل مطابقة لبيانات هذه الاختبار مقارنة بالنماذج الأخرى، كما تفوق النموذج ثلاثي المعلم على النظرية الكلاسيكية بتوفير معلومات أكثر عن خصائص الفقرات (التمييز، والصعوبة، والتخمين).

وركزت دراسة النعيمي (2015) على مقارنة معالم الفقرات والمفحوصين واستقلالية كل منهما عن الآخر باستخدام النظرية الكلاسيكية ونماذج نظرية استجابة الفقرة، مستعيناً ببيانات الإختيار الوطني والمكون من (90) فقرة من نوع الإختيار من متعدد لأربعة بدائل طبقت عينة من (1000) طالب وطالبة من طلبة الصف الثامن، وقد أظهرت النتائج وجود علاقة قوية وطرديّة بين توزيع العلامات الخام في النظرية الكلاسيكية وبين توزيع القدرات في نظرية استجابة الفقرة، وكانت أعلى قيمه لمعامل الارتباط بين النظرية الكلاسيكية والنموذج أحادي المعلم، كما أظهرت النتائج وجود علاقة تامة بين معاملات الصعوبة في النظرية الكلاسيكية والنموذج أحادي المعلم، بينما كانت قوية وطرديّة باستخدام النموذج الثنائي والثلاثي. كما أظهرت

النتائج ارتباط قوي إلى حد ما بـقيم تقدير معلم التمييز بين النظرية الكلاسيكية والنموذج ثنائي المعلم والذي تراوح بين (0.87 - 0.75)، بينما تمتعت العلاقة بين معاملات التمييز في كل من النظرية الكلاسيكية والنموذج ثلاثي المعلم بضعف واضح حيث تراوحت بين (0.44 - 0.11). وقارنت دراسة اوبيجو وافلابي (Awopeju & Afolabi, 2016) بين مؤشرات الصعوبة والتمييز للنظرية الكلاسيكية ونظرية استجابة الفقرة بنماذجها الثلاث، وعلاقتها بقدرة المفحوصين في اختبار في الرياضيات، وقد تكونت عينة الدراسة من مجموعتين مختلفتين في الحجم (100, 1000) والذين قد خضعوا لاختبار الرياضيات المكون من (60) فقرة من نوع الاختيار من متعدد. وقد أظهرت نتائج الدراسة ارتباطاً قوياً في تقديرات مؤشرات الصعوبة بين النظرية الكلاسيكية ونظرية استجابة الفقرة وخاصة مع النموذج أحادي المعلم بالنسبة لمعلم الصعوبة، كما أظهرت النتائج ارتباطاً قوياً بين مؤشرات التمييز وخاصة مع النموذج ثنائي المعلم، وخلصت الدراسة إلى تشابه النظريتين في تقديرها للخصائص الإحصائية والسيكومترية للاختبارات.

وهدف دراسة أبو فودة (2016) إلى الكشف عن التوافق بين الخصائص السيكومترية للاختبار وفقراته بين نظرية الاختبار الكلاسيكية والنموذج ثنائي المعلم في مطابقة اختبار محكي المرجع، من خلال اختبار في مادة الرياضيات مكون من (30) فقرة من نوع الاختيار من متعدد لأربعة بدائل تم تطبيقه على عينة مكونة من (140) طالب، وقد أظهرت النتائج مطابقة (29) فقرة للنظرية الكلاسيكية، و(28) فقرة للنموذج ثنائي المعلم، كما أظهرت النتائج وجود فروق ذات دلالة إحصائية بين معاملي الصدق لصالح النظرية الكلاسيكية، ووجود فروق ذات دلالة إحصائية بين معاملي الثبات لصالح النموذج ثنائي المعلم.

وقارن كلاً من البيجي، واونا، وابانوبي (Eleje, Onah & Abanobi, 2018) بين النظرية الكلاسيكية والنموذج ثلاثي المعلم لنظرية استجابة الفقرة في نتائج تحليل فقرات اختبار مهارات التحليل الكمي الاقتصادي في السنة الثالثة، وعند مقارنة نتائج التحليل في كلا النظريتين أظهرت النتائج اختلاف في انتقاء الفقرات تبعاً لمعالم الفقرات (الصعوبة والتمييز)؛ حيث تم استبعاد فقرتين تبعاً للمعايير المقبولة لمعلم التمييز ضمن نظرية استجابة الفقرة، و(11) فقرة تبعاً لمعايير النظرية الكلاسيكية لمؤشر التمييز، كما أنه تم استبعاد (8) فقرات لم تتوافق مع قيم الصعوبة المقبولة تبعاً لنظرية استجابة الفقرة و (15) فقرة تبعاً للنظرية الكلاسيكية، وقد أظهرت النتائج أن معامل الثبات في النظرية الكلاسيكية قد بلغ (0.66) بينما بلغ في النموذج ثلاثي المعلم (0.86)، وقد أوصت هذه الدراسة باستخدام النموذج ثلاثي المعلم ضمن نظرية استجابة الفقرة كنموذج أكثر فاعلية في الكشف عن الخصائص السيكومترية.

### 1.1. مشكلة الدراسة:

جاءت التوجهات المعاصرة في علم القياس النفسي والتربوي وتقنية المعلومات بما تحمله من تطورات لتشير إلى وجوب إعادة النظر في نظم وأساليب القياس، والحاجة إلى الإهتمام ببناء الاختبارات وتصميم فقراتها، وذلك استناداً إلى نظريات معرفية؛ إذ تقوم الاختبارات التحصيلية في قياس بُنى افتراضية وسمات كامنة- كقدرات المفحوصين وتحصيلهم- والتي لا يمكن قياسها أو ملاحظتها بشكل مباشر، مما يستلزم الإعتماد على نماذج لترجمة أداء المفحوصين على هذه الاختبارات بشكل دقيق، والإستعانة بأساليب إحصائية يمكن الوثوق في نتائجها، ومعرفة كيف لهذه القدرة أو المهارة أن تعكس استجابة المفحوص، مما يساعد بشكلٍ إيجابي في التحليل والتشخيص والعلاج، والذي بدوره سيساهم في التقدم التربوي، وقياس وتطور النمو الإنساني بما يصل إلى الكفاءة المرجوة. (علام، 2011)

وبالرغم من شيوع استخدام النظرية الكلاسيكية في بناء الاختبارات فإنها تعاني من نقاط ضعف عديدة والتي يمكن تلخيصها بعدم استقلالية خصائص الإختبار عن خصائص العينة؛ فلا يمكن وضع خصائص المفحوص والفقرة على متصل واحد، الأمر الذي يحول دون إمكانية تعميم نتائج هذا الإختبار ومقارنة المفحوصين الذين تعرضوا لنماذج مختلفة للاختبار (Wiberg, 2004)، وتركيزها على الدرجة الكلية للاختبار بإعتبارها تمثل القدرة الحقيقية للطلاب وهو ما قد يكون مضللاً لكلٍ من الطالب والمعلم (عيد، 2004). كما أنها لا توفر معلومات عن كيفية أداء المتحنيين عند مستويات قدرة مختلفة للسمة المقاسة، ولا توفر معلومات تفصيلية عن مقدار الخطأ بإختلاف مستويات القدرة (Crocker & Algina, 1986). إضافة إلى عجزها عن توفير قواعد للتنبؤ بإستجابة المفحوصين على الفقرات المختلفة (Lord, 1980).

وفي سبيل تطوير الاختبارات التحصيلية بما يواكب التقدم في هذا المجال، وتجاوز عيوب ومشكلات النظرية الكلاسيكية، ظهرت أهمية تطبيق نظرية استجابة الفقرة كنظرية حديثة تعد بالوصول إلى نموذج اختبار تحصيلي أكثر دقة وموضوعية (عودة، 2014).

ونظراً لكون فقرات الإختبار من متعدد من أكثر أشكال الفقرات شيوعاً في بناء الاختبارات التحصيلية، واعتماد معظم واضعي هذه الاختبارات بشكل أساسي على نظرية الاختبار الكلاسيكية وإجراءاتها في بناء وتطوير هذه الإختبارات، ونظراً لتضارب الدراسات السابقة حول النتائج التي خلصت إليها من كلتا النظريتين؛ فقد جاءت فكرة هذه الدراسة لإلقاء مزيداً من الضوء حول استخدام نظرية استجابة الفقرة في بناء اختبار تحصيلي لمادة اللغة الإنجليزية، ومعرفة أهم الخصائص السيكومترية التي تحققها، ومدى تفوقها على النظرية الكلاسيكية -إن وُجد- وذلك في مجال انتقاء فقرات الاختبار، والكشف عن الخصائص السيكومترية لكلٍ من الفقرات و للاختبار ككل، لتتحدد مشكلة الدراسة بالإجابة عن الأسئلة التالية:

- ما مدى التوافق بين النظريتين الكلاسيكية ونظرية استجابة الفقرة في انتقاء فقرات اختبار اللغة الإنجليزية المعد لهذه الدراسة؟

- ما الخصائص السيكومترية (الصدق والثبات) لاختبار تحصيلي في اللغة الإنجليزية المستخدم في هذه الدراسة وفق كل من نظرية الاختبار الكلاسيكية والنموذج ثلاثي المعلم؟
- هل توجد فروق ذات دلالة إحصائية لمعاملات الثبات المحسوبة وفق كل من نظرية الاختبار الكلاسيكية والنموذج ثلاثي المعلم؟
- هل توجد فروق ذات دلالة إحصائية لمعاملات الصدق المحسوبة وفق كل من نظرية الاختبار الكلاسيكية والنموذج ثلاثي المعلم؟

### 2.1. أهداف الدراسة:

- الكشف عن الفروق بين كل من النظرية الكلاسيكية والنموذج ثلاثي المعلم في مطابقة فقرات اختبار تحصيلي في اللغة الإنجليزية للصف الأول ثانوي.
- التعرف على الخصائص السيكومترية لاختبار تحصيلي في مادة اللغة الإنجليزية على عينة من طلاب المرحلة الثانوية باستخدام النظرية الكلاسيكية.
- التعرف على الخصائص السيكومترية لاختبار تحصيلي في مادة اللغة الإنجليزية على عينة من طلاب المرحلة الثانوية باستخدام النموذج ثلاثي المعلم.
- توفير صورة موضوعية لاختبار تحصيلي يتمتع بخصائص الاختبار الجيد تبعاً للنظريتين.

### 3.1. أهمية الدراسة:

- تستمد هذه الدراسة أهميتها من خلال:
- الأهمية النظرية: تساهم في تدعيم جوانب التوافق والاختلاف بين النظريتين الكلاسيكية ونظرية استجابة الفقرة متمثلة بالنموذج ثلاثي المعلم، وإيضاح جوانب القوة والقصور في كل منهما.
- الأهمية العملية: ستوفر هذه الدراسة نموذج اختبار لمادة اللغة الإنجليزية باستخدام النموذج الثلاثي المعلم، والذي سيمكن بخصائص سيكومترية جيدة، كما أنها ستزود المعلمين بطرق إحصائية دقيقة لتحليل فقرات الاختبار وطرق جيدة لانتقاء الفقرات وتطوير أدوات القياس، وبالتالي الوصول إلى دقة أكبر في تقدير قدرات المفحوصين.

### 4.1. مصطلحات الدراسة:

- نظرية الاختبار الكلاسيكية (Classical Test Theory): هي إحدى نظريات القياس التي تهدف لتقدير ما يعرف بالدرجة الحقيقية للمفحوص (T) والخطأ العشوائي للقياس (E)، والتي تنص على أن الدرجات الملاحظة (X) عند تطبيق المقاييس المختلفة تمثل مكون يعكس القدرة الحقيقية للطلاب، ويعبر عنه بالدرجة الحقيقية (T) ومكون الخطأ (E) (Lord, 1980).
- نظرية استجابة الفقرة (Item Response Theory): وتسمى بنظرية السمات الكامنة (Latent Trait Theory)، كما تُسمى بالنظرية الحديثة (Modern Test Theory)، والتي تفترض وجود قدرات أو سمات كامنة خلف أداء المفحوصين على فقرات الاختبار، ويمكن وصف وتمثيل العلاقة بين هذه السمة وأداء المفحوص من خلال نماذج رياضية توفرها هذه النظرية. (Harris, 1989)
- النموذج ثلاثي المعلم (Three Parameter Logistic Model): هو أحد نماذج نظرية الاستجابة الفقرة ثنائية التدرج، وتبعاً لهذا النموذج تتباين الفقرات في معالمها الثلاث: معلم الصعوبة، معلم التمييز، ومعلم التخمين للفقرة. (بني عطا والشريفين، 2012)
- الخصائص السيكومترية للاختبار (Test Psychometric Properties): يُقصد بها ثبات الاختبار (Test Reliability) وصدق الاختبار (Test Validity) (علام، 2011).
- انتقاء الفقرات (Item Selection): هي مجموعة من الطرق والأساليب الإحصائية المستخدمة في انتقاء الفقرات (هياجنة، 2011)، ومن أكثرها شهرة الإعتماد على مؤشرات الصعوبة والتمييز في نظرية الاختبار الكلاسيكية، وعلى كمية المعلومات التي تقدمها الفقرات إضافة إلى معالم الفقرات عند انتقاء الفقرات في نظرية استجابة الفقرة (Mead & Mead, 2010). وفي هذه الدراسة تم الإعتماد على المعايير التي وضعها كلاً من (Lord, 1953) و (Ebel, 1965) في انتقاء فقرات النموذج الكلاسيكي، و (Warm, 1978) و (جمحاوي، 2000) في انتقاء فقرات النموذج ثلاثي المعلم.

### 5.1. حدود الدراسة:

- الحد الموضوعي: تقتصر هذه الدراسة على دراسة النموذج ثلاثي المعلم ونظرية الاختبار الكلاسيكية لانتقاء فقرات اختبار تحصيلي في اللغة الإنجليزية.
- الحد الزمني: تم تطبيق هذه الدراسة في الفصل الثاني من العام الدراسي (1439-1438).

- الحد المكاني: تم تطبيق أداة الدراسة في مدينة عنيزة فقط.
- الحد البشري: تكونت عينة الدراسة من طالبات الصف الأول ثانوي فقط.

## 2. منهجية البحث وإجراءاته:

### 1.1. منهج الدراسة:

سيتم استخدام المنهج الوصفي التحليلي نظراً لملاءمته لإجراءات الدراسة.

### 2.2. مجتمع الدراسة:

تكون مجتمع الدراسة من طالبات الصف الأول ثانوي في مدارس محافظة عنيزة الحكومية للعام الدراسي (1438-1439 هـ).

### 3.2. عينة الدراسة:

#### أولاً: العينة الاستطلاعية:

حيث تم تطبيق الاختبار بصورته الأولية والمكونة من (52) فقرة من قبل الباحثة على عينة استطلاعية حجمها (100) طالبة، وقد ساهمت نتائج هذا التطبيق في تحسين الاختبار وحساب الزمن المناسب للإستجابة على الاختبار.

#### ثانياً: العينة الأساسية:

تكونت عينة الدراسة الأساسية من (1076) طالبة من طالبات الصف الأول ثانوي المنتظمات في مدارس محافظة عنيزة الحكومية للعام الدراسي (1438-1439 هـ). وتم اختيار العينة بواسطة الطريقة العشوائية العنقودية متعددة المراحل، حيث كانت المدرسة والشعبة هي وحدات الإختيار، ويظهر جدول (1) توزيع أفراد مجتمع وعينة الدراسة.

جدول (1): توزيع أفراد مجتمع وعينة الدراسة

المدرسة	المجتمع	العينة	النسبة المئوية للعينة
الثانوية الثالثة	200	160	14.9%
الثانوية السادسة	106	101	9.4%
الثانوية الحادية عشر	119	107	9.9%
الثانوية الثانية	110	98	9.1%
الثانوية السابعة	156	156	14.5%
الثانوية الخامسة	224	197	18.3%
ثانوية أم المؤمنين	72	57	5.3%
الثانوية الأولى	77	74	6.9%
ثانوية التحفيظ الأولى	70	69	6.4%
الثانوية الثامنة	49	29	2.7%
الثانوية الرابعة	62	28	2.6%
المجموع	1245	1076	100%

### 4.2. أداة الدراسة:

تكونت أداة الدراسة من اختبار تحصيلي لمادة اللغة الإنجليزية للصف الأول الثانوي من إعداد الباحثة، والذي هدف الى قياس تحصيل الطالبات في الوحدة الأولى والثانية من المقرر الدراسي لهذه المرحلة (Traveller 2).

### 5.2. تصحيح الاختبار:

قامت الباحثة بتصحيح الاختبار يدوياً ثم إدخال البيانات إلكترونياً؛ حيث أعطيت القيمة (1) للإجابة الصحيحة، والقيمة (0) للإجابة الخاطئة.

### 6.2. الخصائص السيكومترية للاختبار:

#### أولاً: صدق الاختبار:

- صدق المحتوى: من خلال عرض فقرات الاختبار والأهداف التي تمثلها على (14) محكماً من ذوي الاختصاص (معلمات اللغة الإنجليزية للصف الأول ثانوي، ومشرفات تربويات مادة اللغة الإنجليزية، وأكاديميين بدرجة الدكتوراة والمجستير في تخصص اللغة الإنجليزية ومناهج وطرق تدريس اللغة الإنجليزية)، وبناءً على نتائج التحكيم تم حذف فقرات كانت نسبة الاتفاق عليها أقل من (80%).
- صدق البناء: والذي يعتمد على التحديد العملي للعوامل التي تؤثر في الأداء (أبو علام، 2011). وقد تم التحقق من هذا الصدق من خلال التحليل العاملي الاستكشافي (Exploratory factor Analysis) والتحليل الموازي (Parallel Analysis)؛ حيث يتم حساب الجذور الكامنة من مصفوفة

البيانات الأصلية ومقارنتها بقيم الجذور الكامنة العشوائية للبيانات الموازية، واعتبار العامل على أنه جوهري إذا زاد مقدار الجذور الكامنة له في البيانات الأصلية عن نظيرتها في العشوائية (المومني، 2017).

ثانياً: ثبات الاختبار:

- ثبات التجانس الداخلي (الإتساق الداخلي): تم التحقق من ثبات الإتساق الداخلي باستخدام معادلة كيودر ريتشاردسون (KR-20) حيث بلغت قيمة معامل الثبات (0.924) مما يُشير إلى تمتع الإختبار بثبات جيد. كما تم استخراج معامل ارتباط ألفا في حال حذف الفقرة (Alpha if Item Deleted) للتحقق من ثبات تجانس فقرات الإختبار كل على حده، والذي أشار إلى أن جميع الفقرات كانت مناسبة.

### 7.2. الأساليب الإحصائية المستخدمة في الدراسة:

- لتحقيق أهداف الدراسة والاجابة على الأسئلة، تم إجراء المعالجات الإحصائية اللازمة لذلك وهي كما يلي:
- استخراج الخصائص السيكومترية للفقرات (الصعوبة والتمييز) تبعاً لنظرية الاختبار الكلاسيكية وذلك باستخدام برنامجي (Excel, SPSS-19).
- التحقق من مطابقة البيانات لإفترض أحادية البعد للنموذج ثلاثي المعلم بواسطة التحليل العاملي الاستكشافي باستخدام برنامج (SPSS-19)، والتحليل العاملي التوكيدي باستخدام برنامج (AMOS-24).
- استخراج مؤشر مطابقة الفقرات (S-X2) ومؤشر المطابقة العامة (M2) للنموذج البارامتري ثلاثي المعلم باستخدام البرنامج الإحصائي (IRT-PRO4).
- التحقق من الاستقلال الموضوعي من خلال الاختبار الإحصائي مربع كاي للاستقلال الموضوعي ( $LD \chi^2$ ).
- استخراج الخصائص السيكومترية للفقرات تبعاً للنموذج ثلاثي المعلم (الصعوبة، والتمييز، والتخمين).
- استخراج الخصائص السيكومترية للاختبار (الصدق والثبات) تبعاً لنظرية الاختبار الكلاسيكية والنموذج ثلاثي المعلم من خلال استخراج معامل كيودر ريتشاردسون (KR-20) للثبات في النظرية الكلاسيكية ومعامل الثبات التجريبي في النموذج ثلاثي المعلم، واستخراج معامل الصدق المحكي لكلا النظريتين.
- حساب نسبة الاتفاق بين النظرية الكلاسيكية والنموذج ثلاثي المعلم في انتقاء فقرات الاختبار من خلال معادلة كوبر لحساب نسبة الاتفاق (Cooper, 1974):

$$100 \times \frac{\text{عدد مرات الاتفاق}}{\text{عدد مرات الاختلاف} + \text{عدد مرات الاتفاق}}$$

- حساب دلالة الفرق بين معاملي الثبات الناتجين من النظرية التقليدية والنموذج ثلاثي المعلم بواسطة إحصائي (X2)، وذلك باستخدام الحزمة الإحصائية كوكرون (Cocron)، وهي حزمة إحصائية جزئية ضمن الحزمة الإحصائية (R- Package).
- الكشف عن دلالة الفرق بين معاملي الصدق الناتجين من النظرية التقليدية والنموذج ثلاثي المعلم باستخدام اختبار (T) لحساب دلالة الفرق بين معاملي صدق للعينات المترابطة.

### 3. النتائج وتفسيرها:

#### 1.3. النتائج المتعلقة بالسؤال الأول: "ما مدى مطابقة فقرات اختبار تحصيلي في اللغة الإنجليزية لنظرية الاختبار الكلاسيكية؟"

للإجابة عن هذا السؤال فقد تم استخدام البرنامج الإحصائي (SPSS-19) لحساب خصائص الفقرات (الصعوبة والتمييز) لاختبار اللغة الإنجليزية المستخدم في هذه الدراسة وفق نظرية الاختبار الكلاسيكية، حيث يمثل معامل الصعوبة عدد الإجابات الصحيحة إلى العدد الكلي للإستجابات على كل فقرة من فقرات الاختبار، كما تم حساب معامل التمييز من خلال إيجاد معامل الإرتباط المصحح (Point Biseral) لكل فقرة بالإختبار بعد حذف قيمة الفقرة (Corrected Item - Total Correlation)، وجدول (2) يبين قيم معاملات الصعوبة والتمييز لفقرات الاختبار.

جدول (2): معاملات الصعوبة والتمييز لفقرات الاختبار التحصيلي وفق النظرية الكلاسيكية

رقم الفقرة	الصعوبة	التمييز	رقم الفقرة	الصعوبة	التمييز	رقم الفقرة	الصعوبة	التمييز
1	0.36	0.44	15	0.47	0.42	29	0.41	0.35
2	0.29	0.43	16	0.47	0.47	30	0.44	0.43
3	0.41	0.42	17	0.39	0.31	31	0.30	0.41
4	0.47	0.39	18	0.33	0.36	32	0.39	0.38



0.35	0.32	33	0.58	0.39	19	0.51	0.48	5
0.43	0.31	34	0.41	0.52	20	0.31	0.43	6
0.49	0.28	35	0.41	0.38	21	0.48	0.55	7
0.40	0.50	36	0.34	0.35	22	0.49	0.63	8
0.15	0.24	37	0.32	0.47	23	0.52	0.58	9
0.52	0.40	38	0.42	0.63	24	0.46	0.41	10
0.32	0.33	39	0.34	0.42	25	0.49	0.50	11
0.42	0.41	40	0.39	0.35	26	0.37	0.47	12
0.29	0.46	41	0.35	0.27	27	0.47	0.42	13
0.29	0.36	42	0.43	0.50	28	0.29	0.24	14

يبين الجدول (2) أن قيم معاملات الصعوبة لفقرات الاختبار قد تراوحت بين (0.24 - 0.63)، كما تم حساب المتوسط الحسابي لمعاملات صعوبة الفقرات والذي بلغ (0.41) بإنحراف معياري (0.09). كما يبين الجدول (2) أن قيم معاملات تمييز الفقرات قد تراوحت بين (0.15 - 0.58) كما تم حساب المتوسط الحسابي لمعاملات التمييز والذي بلغ (0.40)، بإنحراف معياري (0.08). ويوضح الجدول (3) معايير للحكم على هذه الفقرات وانتقاءها تبعاً لمعاملات الصعوبة والتمييز لها بالرجوع إلى (Lord, 1953 ; Ebel, 1965).

الفقرات المحذوفة	المدى المقبول	معلم الفقرة
8, 14, 24, 37	$0.25 \leq p_i \leq 0.6$	الصعوبة
37	$d_i \geq 0.2$	التمييز

وبالاعتماد على هذه المعايير تم استبعاد الفقرات (8, 24) لسهولةها وتجاوز معامل صعوبتهما عن (0.6)، وكذلك تم حذف الفقرتين (14, 37) لإرتفاع صعوبتهما وانخفاض معامل الصعوبة لهما عن (0.25)، كما أن تمييز الفقرة (37) كان أقل من الحد الأدنى المقبول للتمييز؛ وعلى ذلك تم استبعادها.

وبناءً على ماسبق فقد طبقت (38) فقرة من فقرات الإختبار التحصيلي في اللغة الإنجليزية المعد في هذه الدراسة بصورته النهائية للنظرية الكلاسيكية، وهو يشكل ما نسبته (90.48%) من فقرات الاختبار. وهي نسبة تُعد مرتفعة وتؤكد على أن افتراضات النظرية الكلاسيكية والتي تمتاز بالبساطة وأنها تطابق معظم البيانات (Harris, 1989).

### 2.3. النتائج المتعلقة بالسؤال الثاني: "ما مدى مطابقة فقرات اختبار تحصيلي في اللغة الإنجليزية للنموذج ثلاثي المعلم؟"

للإجابة عن هذا السؤال تم استخدام التحليلات الخاصة بنظرية استجابة الفقرة، وحيث أن هذه النظرية تتطلب تحقق عدداً من الافتراضات الأساسية كمبررات مسبقة لإستخدامها؛ فقد تم بدايةً التحقق من هذه الافتراضات، وذلك كمايلي:

أ- التحقق من الافتراضات الأساسية لنظرية استجابة الفقرة:

#### • أحادية البعد (Unidimensionality):

أولاً: التحليل العاملي الإستكشافي (Exploratory Factor Analysis):

بواسطة برنامج (SPSS-19) تم استخدام التحليل العاملي الإستكشافي للتحقق من افتراض أحادية البعد والذي يُعد أحد أهم طرق التحقق من أحادية البعد تبعاً لما أشار له (Hattie, 1985). وقد قامت الباحثة أولاً من التحقق من مدى ملاءمة بيانات عينة الدراسة لإجراءات التحليل العاملي باستخدام اختبار بارتليت (Bartlett's Test) ومؤشر كايزر-ماير اولكن (KMO) [Meyer Olken] [Keiser-]، وذلك كما هو مبين في جدول (4):

الاختبار المستخدم	قيمته	الدلالة
Bartlett's Test	8619.273	0.01
KMO	.941	تدل على جودة المعاينة.

يُظهر جدول (4) تمتع اختبار بارتلبيت (Bartlett's Test) بدلالة إحصائية، كما يظهر أن قيمة مؤشر كايزر-ماير اولكن (KMO) قد بلغت (0.941). وهي قيمة مرتفعة وتزيد عن الحد الأدنى المقبول والبالغ (0.5)، مما يُفيد ان البيانات ملائمة وبشكل كبير لأغراض اجراء التحليل العاملي (الشافعي، 2014).

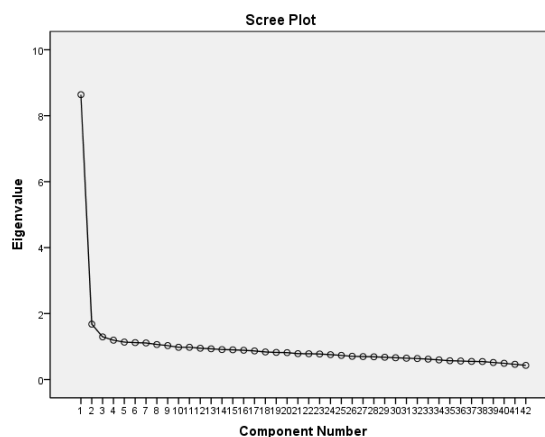
بعد التحقق من ملائمة البيانات، تم إجراء التحليل العاملي الاستكشافي بطريقة المكونات الرئيسية (Principal Components Analysis) لإستجابات الطالبات على الاختبار التحصيلي والمكون من (42) فقرة من نوع الإختيار من متعدد لأربعة بدائل، حيث بلغ عدد الطالبات المستجيبات (1076) طالبة، وقد افرزت النتائج وجود (9) عوامل بلغت قيمة الجذر الكامن لكل واحد منها أكثر من واحد، وفسرت بمجموعها قدرأ من التباين بلغ (43.46%)، وببين جدول (5) قيم الجذور الكامنة (Eigenvalues) لهذه العوامل ونسبة التباين المفسر من قبل كل منها (Explained variance)، إضافة إلى نسبة التباين التراكمي (Cumulative variance).

جدول (5): قيم الجذور الكامنة ونسبة التباين المفسر لكل عامل ونسبة التباين التراكمي

Components	Initial Eigenvalues			Components	Initial Eigenvalues		
	Total	% of Variance	Cumulative %		Total	% of Variance	Cumulative %
1	8.637	20.565	20.565	6	1.118	2.663	35.849
2	1.680	3.999	24.565	7	1.109	2.641	38.490
3	1.294	3.080	27.645	8	1.061	2.525	41.015
4	1.192	2.839	30.483	9	1.029	2.451	43.466
5	1.135	2.703	33.186				

يتضح من الجدول (5) أن اختبار اللغة الإنجليزية التحصيلي المستخدم في هذه الدراسة يحقق افتراض أحادية البعد وذلك تبعاً لما يلي:

- بلغت قيمة الجذر الكامن (Eigen value) للعامل الأول (8.63)، كما فسّر نسبة من التباين الكلي بلغت (20.56%)، وهي نسبة تزيد عن (20%)، كما أن نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني بلغت (5.14) وهي قيمة أكبر من (4)، الأمر الذي يؤكد أن الاختبار يقيس عاملاً واحداً (Reckase, 1979).
- كما تم تمثيل الرسم البياني للجذور الكامنة باستخدام اختبار (Scree Plot) وذلك كما هو موضح في شكل (1):



شكل (1): التمثيل البياني لقيم الجذور الكامنة للعوامل المكونة للاختبار

يظهر الرسم البياني للجذور الكامنة (Scree Plot) في شكل (1) تفوق العامل الأول على بقية العوامل وتقارب بقية العوامل الأخرى، الأمر الذي يعد دليلاً ثالثاً على أحادية البعد للاختبار المستخدم في هذه الدراسة.

ثانياً: التحليل العاملي التوكيدي (Confirmatory Factor Analysis):

وذلك من خلال استخدام برنامج (AMOS-24) كدليل أكثر دقة للحكم على بُنية المقياس العاملية، حيث تم حساب مؤشرات المطابقة [Goodness of Fit (GOF)] لنموذج العامل العام (أحادي البعد)، وذلك كما هو موضح في جدول (6).

جدول (6): مؤشرات جودة المطابقة لنموذج التحليل العاملي التوكيدي

المؤشر	المعيار	القيمة
مؤشر جودة المطابقة [ Goodness of Fit Index (GFI) ]	$(GFI) \geq 0.80$	0.944
مؤشر المطابقة المقارنة [ Comparative Fit Index (CFI) ]	$(CFI) \geq 0.80$	0.938
مؤشر توكير-لوييس [ Tucker - Lewis Index (TLI) ]	$(TLI) \geq 0.80$	0.933
مؤشر المطابقة التدريجي [ Incremental Fit Index (IFI) ]	$(IFI) \geq 0.80$	0.94
الجذر التربيعي لمتوسط مربعات أخطاء التقدير (RMSEA)	$(RMSEA) \leq 0.06$	0.024

- ويُظهر جدول (6) أن مؤشرات جودة مطابقة البيانات للنموذج أحادي البُعد وقعت جميعها في المدى المقبول (الشافعي، 2014). إضافة إلى تمتعها بقيمة مرتفعة وهو ما يُشير إلى مطابقة جيدة للنموذج؛ الأمر الذي يُعد دليلاً رابعاً وأكثر الطرق دقة للتحقق من أحادية البُعد.
- **الاستقلال الموضوعي (Local Independence):** تم التحقق من خلال إحصائي فحص الإستقلال الموضوعي ( $\chi^2_{LD}$ )، والذي يقوم على مقارنة التكرارات المتوقعة والملاحظة لكل زوج من الفقرات، حيث يشير ارتفاع قيمته إلى عدم وجود استقلالية موضعية (Local dependence)، فيجب أن لا تزيد قيمة هذا الإحصائي عن ( $LD\chi^2 > 10$ ) (Cai, Thissen & du Toit, 2011)؛ وعلى ذلك تم حذف الفقرات التي لم تحقق هذا الشرط وذات الخصائص الأضعف إحصائياً من كل زوج والتي أظهرت انتهاكاً لهذا الشرط، وهما (17,35).
  - **التحرر من الوقت (Nonspeeded):** تم احتساب الزمن الذي أتمت به (80%) من الطالبات الاستجابة على فقرات الاختبار، وذلك تبعاً لما أشار إليه كلاً من هامبلتون وسواميناثان (Hambleton & Swaminathan, 1985) إلى أنه يمكن التحقق هذا الافتراض عند استجابة (80%) من المفحوصين على فقرات الاختبار، أو عند الإجابة على ما نسبته (75%) من الفقرات.
  - **مطابقة الفقرات للنموذج ثلاثي المعلم:** تم تحليل البيانات باستخدام برنامج (IRT-PRO4) للتحقق من مطابقة البيانات للنموذج ثلاثي المعلم بواسطة مؤشر المطابقة للفقرات، وهو إحصائي يقيس مدى تحقق شرطي التزايد المطرد والشكل اللوجستي لفقرات الاختبار، وقد تم ذلك باستخدام كلاً من:
    1. مؤشر مطابقة الفقرة ( $S-X^2$ ): حيث تُشير القيم الدالة إحصائياً إلى عدم مطابقة الفقرات للنموذج (Meseganaw, 2014). ويظهر جدول (7) قيم مؤشر المطابقة ( $S-X^2$ ) لفقرات اختبار اللغة الإنجليزية التحصيلي.

جدول (7): إحصائيات ( $S-X^2$ ) لمطابقة الفقرات للنموذج ثلاثي المعلم

Item	S-X <sup>2</sup>	d.f.	Probability	Item	S-X <sup>2</sup>	d.f.	Probability
1	23.59	35	0.9290	22	46.04	34	0.0812
2	25.10	33	0.8366	23	36.46	34	0.3540
3	30.05	31	0.5158	24	15.98	27	0.9535
4	30.66	33	0.5851	25	27.22	32	0.7083
5	22.09	28	0.7779	26	23.92	32	0.8478
6	30.48	35	0.6868	27	28.29	35	0.7825
7	35.53	27	0.1256	28	24.17	32	0.8385
8	15.62	19	0.6833	29	45.05	34	0.0972
9	32.67	20	0.0365	30	23.37	31	0.8358
10	22.81	29	0.7858	31	27.15	32	0.7115
11	35.39	28	0.1584	32	22.33	35	0.9524
12	20.38	33	0.9580	33	23.66	33	0.8844
13	27.59	31	0.6432	34	30.70	32	0.5333
14	19.35	36	0.9894	35	32.43	31	0.3978
15	23.92	32	0.8477	36	31.70	32	0.4832
16	29.28	29	0.4520	37	28.56	36	0.8070
17	19.95	32	0.9523	38	26.83	30	0.6332
18	24.12	35	0.9172	39	22.74	35	0.9453
19	31.82	26	0.1986	40	31.35	32	0.5007
20	32.67	30	0.3363	41	32.99	35	0.5666
21	30.35	32	0.5511	42	31.78	35	0.6252

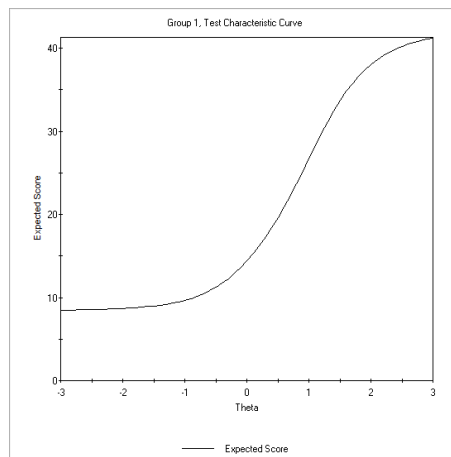
- ويظهر من جدول (7) عدم مطابقة الفقرة (9) للنموذج ثلاثي المعلم وفقاً للمؤشر ( $S-X^2$ )، حيث كانت القيمة الاحتمالية لهذه الفقرة أقل من (0.05)، وبالتالي تُحذف هذه الفقرة تبعاً لنظرية استجابة الفقرة.
2. **المؤشر الإحصائي (M2) للمطابقة العامة للنموذج:** ويقوم هذا المؤشر بفحص مطابقة البيانات بشكل عام للنموذج من خلال اختبار توزيع إحصائيات مربع كاي، حيث تعبر القيم المتدنية له عن مطابقة أفضل للنموذج، ويتم الحكم عليه تبعاً لمحك متوسط مربعات أخطاء التقدير

والذي يجب أن تقل قيمته عن ( $RMSEA < 0.06$ ), مما يُشير إلى جودة مطابقة الاختبار للنموذج، وهو مؤشر جيد يُعالج حساسية إحصائي ( $X^2$ ) لحجم العينة وعدد المعالم (Brown & Cudeck, 1993; Meseganaw, 2014). وبين جدول (8) قيمة مؤشر ( $M_2$ ) للمطابقة العامة للنموذج.

جدول (8): نتائج المؤشر الإحصائي ( $M_2$ ) للمطابقة العامة للنموذج		
$M_2$	Prob.	RMSEA
1019.25	0.0001	0.02

يظهر من جدول (8) أن قيمة متوسط مربعات أخطاء التقدير بلغت ( $RMSEA = 0.02$ ) وهي قيمة تقل عن ( $0.06$ ) مما يشير إلى مطابقة البيانات للنموذج العام لنظرية استجابة الفقرة.

3. منحى خصائص الاختبار (**Test Characteristic Curve**): يقوم منحى خصائص الاختبار على وصف العلاقة بين كل من القدرة واحتمالية الإستجابة الصحيحة ممثلةً بالدرجة المتوقعة للمفحوصين (الوسيدي، 2017). كما أنه يمثل المنحى البياني لمطابقة الاختبار ككل للنموذج، ويوضح شكل (2) تمثيل منحى خصائص الاختبار بيانياً.



شكل (2): منحى خصائص الاختبار (**Test Characteristic Curve**) لاختبار اللغة الإنجليزية التحصيلي

يبين شكل (2) أن منحى خصائص الاختبار قد تمتع بتوزيع يأخذ شكل ( $S$ )، كما أنه متزايد بإطراد مما يؤكد ويُعد دليلاً آخر على مطابقة النموذج العام للبيانات والذي تم الحكم عليه سابقاً من خلال الإحصائي ( $M_2$ ). كما يُظهر الشكل تمتع هذا المنحى بإنحدار جيد (ميل المنحى)، وذلك لتمتع الاختبار بقيمة معلم تمييز بلغت ( $4.22$ )، إضافة إلى أن القاطع مع محور الصادات زاد عن ( $0$ )، وذلك لوجود نسبة من التخمين بلغت ( $0.29$ ).

4. خصائص الفقرات السيكومترية في النموذج ثلاثي المعلم: تم حساب خصائص الفقرات تبعاً للنموذج الثلاثي المعلم باستخدام البرنامج الإحصائي ( $IRT-PRO4$ )، وذلك كما هو موضح في جدول (9) والذي يبين قيم معالم كل فقرة والأخطاء المعيارية لتقدير كل معلم.

جدول (9): قيم معالم الفقرات [التمييز ( $ai$ )، والصعوبة ( $bi$ )، والتخمين ( $ci$ )، والأخطاء المعيارية في تقديرها ( $s.e.$ ).

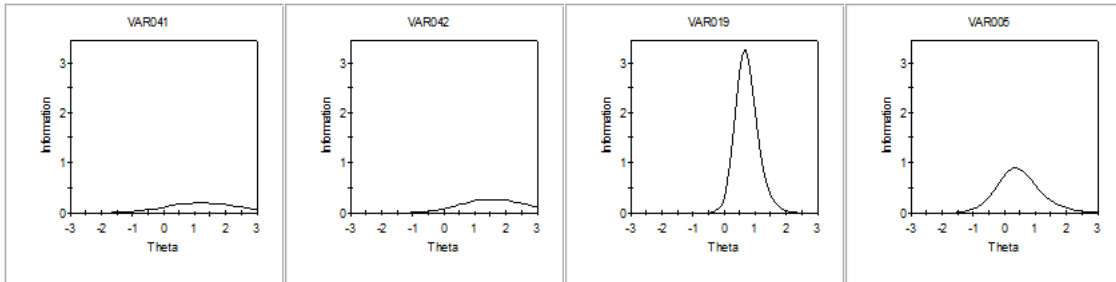
Item	$Ai$	$s.e.$	$bi$	$s.e.$	$ci$	$s.e.$	Item	$Ai$	$s.e.$	$bi$	$s.e.$	$ci$	$s.e.$
1	1.65	0.37	0.90	0.14	0.13	0.04	24	1.65	0.22	-0.24	0.17	0.15	0.07
2	2.66	0.62	1.19	0.08	0.16	0.02	25	2.50	0.82	1.16	0.09	0.29	0.02
3	2.92	0.70	0.94	0.10	0.25	0.02	26	2.82	0.80	1.15	0.08	0.22	0.02
4	1.38	0.30	0.63	0.22	0.20	0.06	27	2.08	0.49	1.49	0.12	0.16	0.02
5	2.12	0.31	0.28	0.14	0.12	0.04	28	1.61	0.25	0.36	0.14	0.17	0.07
6	1.30	0.31	1.06	0.18	0.22	0.05	29	1.96	0.48	1.12	0.10	0.25	0.03
7	2.12	0.30	0.10	0.13	0.16	0.05	30	2.50	0.53	0.80	0.08	0.24	0.03
8	3.50	0.67	-0.10	0.16	0.22	0.05	31	3.52	0.84	1.24	0.07	0.19	0.02
10	3.62	0.90	0.87	0.08	0.24	0.02	32	1.15	0.22	0.84	0.17	0.11	0.06
11	2.42	0.40	0.34	0.14	0.19	0.04	33	3.33	0.91	1.32	0.08	0.22	0.02
12	1.34	0.25	0.60	0.17	0.18	0.06	34	3.54	0.91	1.17	0.07	0.18	0.02
13	2.39	0.47	0.70	0.12	0.19	0.03	36	1.55	0.28	0.44	0.17	0.20	0.08
14	1.74	0.54	1.79	0.15	0.14	0.02	37	3.39	2.40	1.97	0.11	0.20	0.02
15	1.75	0.30	0.50	0.13	0.18	0.05	38	2.18	0.30	0.58	0.08	0.12	0.04
16	3.09	0.67	0.63	0.11	0.26	0.03	39	1.93	0.47	1.44	0.10	0.22	0.03
18	1.62	0.41	1.27	0.11	0.18	0.03	40	2.45	0.44	0.91	0.08	0.23	0.03
19	4.22	0.79	0.61	0.09	0.15	0.02	41	1.09	0.34	0.94	0.31	0.22	0.11
20	1.98	0.44	0.54	0.14	0.27	0.04	42	1.23	0.35	1.33	0.17	0.19	0.07
21	2.47	0.57	0.99	0.08	0.21	0.03	Mean	2.24		0.88		0.19	
22	1.69	0.43	1.32	0.11	0.20	0.03	S.d.	0.82		0.46		0.04	
23	1.12	0.25	0.81	0.21	0.21	0.07							

يظهر جدول (9) مايلي:

- معلم الصعوبة [Threshold parameters (bi)]: قد تراوحت القيم بين (-1.97 - 0.24) بمتوسط عام للصعوبة بلغ (0.88)، وانحراف معياري (0.46).
  - معالم التمييز [Slope parameters (ai)]: قد تراوحت القيم بين (1.09 - 4.22) بمتوسط بلغ (2.24)، وانحراف معياري (0.82)، وحيث انه من النادر الحصول على قيم تمييز تتجاوز المدى الشائع للتمييز وهو (2.5)، إلا أن ارتفاع قيمتها تعكس جودة التمييز وان كان في مدى ضيق للقدرة (Warm, 1978).
  - معالم التخمين [Pseudo-guessing parameters (ci)]: قد تراوحت القيم بين (0.11 - 0.29) بمتوسط حسابي بلغ (0.19)، وانحراف معياري (0.04).
- وتبعاً لجميع ماقد سبق فقد قامت الباحثة بانتقاء فقرات الاختبار تبعاً للمدى المقبول لمعالم الفقرات في النموذج ثلاثي المعلم، بالرجوع الى ما أشار اليه كل من جمحوي (2000) و وارم (Warm, 1978)، والمشار اليها في الجدول (10).

جدول (10): معايير إنتقاء فقرات النموذج ثلاثي المعلم		
معالم الفقرات	المدى المقبول	الفقرات المحذوفة
الصعوبة (bi)	$-2.5 \geq bi \geq 2.5$	لا يوجد
التمييز (ai)	$ai \geq 0.8$	لا يوجد
التخمين (ci)	$ci \leq 0.25$	16, 20, 25

- وبالأخذ بالمعايير المقترحة في جدول (10) فقد تم استبعاد الفقرات (16, 20, 25)؛ لتجاوز قيم التخمين لها الحد الأعلى المقبول في فقرات الاختيار من متعدد لأربعة بدائل، ليصبح الاختبار في صورته النهائية مكون من (36) فقرة، أي مانسبته (85.71%) من الفقرات.
- دالة معلومات الفقرة [Item Information Function (  $I_i(\theta)$  ]: تم استخراج دوال المعلومات لكل فقرة من فقرات الاختبار التحصيلي والمكون بصيغته النهائية من (36) فقرة، والتي تعكس كمية المعلومات التي تساهم بها كل فقرة في الاختبار وهو ما يتناسب عكسياً مع قيمة الخطأ المعياري للقياس، وهي تعتمد على صعوبة وتمييز الفقرة؛ فالصعوبة تتحكم بموقع الفقرة على متصل القدرة، وكلما زادت الصعوبة انسحبت الدالة نحو اليمين على متصل القدرة، الأمر الذي يُفيد بأنها تقدم معلومات أكبر للطلبة ذوي القدرة المرتفعة، بينما تتحكم قيمة التمييز بارتفاع المنحنى بالنسبة للفقرة ومدى القدرة التي تشملها، وكلما زاد ارتفاع المنحنى ارتفعت كمية المعلومات، والذي يعكس بدوره دقة الفقرة في القياس وانخفاض قيمة الخطأ المعياري في القياس (Meseganaw, 2014). ويظهر شكل (3) دوال معلومات لبعض فقرات الاختبار.

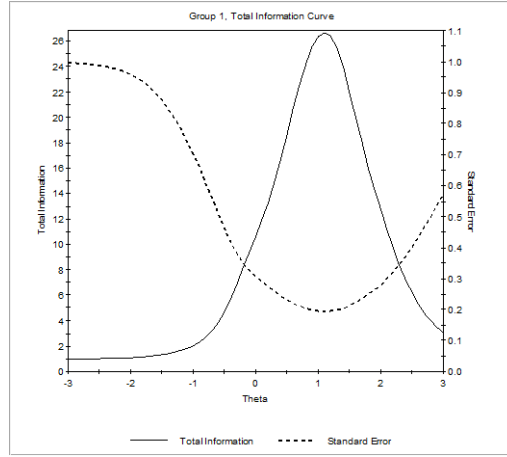


شكل (3): دوال معلومات بعض الفقرات (Item Information Functions)

ويظهر من خلال شكل (3) تباين الفقرات في مدى مساهمتها في تقديم معلومات عن الاختبار و المفحوصين والذي يعكسه ارتفاع منحنياتها على متصل القدرة؛ حيث كانت الفقرة (19) أعلى الفقرات ارتفاعاً في منحناها، فقد بلغ حجم مساهمتها بالمعلومات (3.03) عند مستوى القدرة (0.08)، أي أنها توفر معلومات أكبر في المستويات المتوسطة- المرتفعة نوعاً ما للقدرة، وعند الرجوع لإحصاءات هذه الفقرة نجد أنها تمتعت بمعلم تمييز مرتفع وإن صعوبتها كانت متوسطة نوعاً ما، إضافة إلى انخفاض معلم التخمين بها، الأمر الذي يقودنا إلى الإستنتاج إلى أن ارتفاع تمييز الفقرة إضافةً إلى مقدار صعوبة متوسط واحتمال تخمين متدني يقدم معلومات بالغة عن الفقرات والمفحوصين على حد سواء. وقد اتسع مدى القدرة لبعض الفقرات لتوفر معلومات تشمل المستويات المتوسطة كالفقرة (5)، بينما كانت الفقرة (41) هي الأدنى في الارتفاع وبكمية المعلومات، وهو ما يرتبط بالقدرة التمييزية المنخفضة لهذه الفقرة، ويجدر الذكر أن معظم الفقرات التي قدمت كمية من المعلومات المتدنية قد امتلكت معاملات تخمين مرتفعة عند مراجعتها

كالفقرات (41,42)، الأمر الذي يؤكد تأثير وجود التخمين على خصائص فقرات الإختبار والمعلومات التي يقدمها، وبالتالي معامل ثبات الإختبار، وهو ما يزيد من مقدار الخطأ في القياس ويجعل النتائج أقل دقة وموثوقية.

• دالة معلومات الإختبار (Test Information Function): تعكس دالة معلومات الإختبار مدى جودة الإختبار ككل في قياس السمة المراد قياسها، وهي تتناسب عكسياً مع مقدار الخطأ في القياس (Momani, 2018)، وقد تم تمثيل دالة معلومات الإختبار التحصيلي في اللغة الإنجليزية المستخدم في هذه الدراسة، وذلك كما هو موضح في شكل (4).



شكل (4): دالة معلومات الإختبار التحصيلي في اللغة الإنجليزية (Test Information Function)

حيث يُظهر شكل (4) أن الخطأ المعياري والذي يمثله الخط المتقطع يتناقص تدريجياً وبشكل عكسي مع معلومات الإختبار كلما انتقلنا من المستويات الدنيا للقدرة إلى المستويات العليا؛ حيث بلغ الخطأ المعياري أقل قيمة وهي (0.19) عند مستوى قدرة (1.2) وهي القيمة المناظرة لأعلى قيمة لدالة معلومات الإختبار (26.4)، كما عادت قيمة الخطأ بالتزايد وكمية المعلومات بالتناقص لمدى القدرة [3 - 1.2]، إلا أن الشكل يشير إلى أن الإختبار يقدم معلومات عند هذا المستوى أكبر مما يقدمه عند مستويات القدرة المتدنية [0.5 - -3] وبمقدار خطأ أقل، مما يُشير عموماً إلى أن الإختبار المستخدم في هذه الدراسة يميل لقياس القدرة المتوسطة - العليا، كما يقدم معلومات أكبر للمفحوصين الذين يتمتعون بمقدار متوسط - مرتفع من السمة.

وحيث يتعلق جودة الإختبار ضمن هذه النظرية بكمية المعلومات التي يقدمها؛ فقد تم حساب المعلومات التي تساهم بها كل فقرة والإختبار بشكل كامل بعد حذف الفقرات الضعيفة، فقد بلغت أكبر كمية معلومات للإختبار على متصل القدرة (26.40) عند مستوى قدرة (1.2)، ثم (24.16) عند مستوى (0.8)، وأقل كمية معلومات كانت (1.01) عند مستوى (-2.8)؛ أي أن الإختبار يقيس المستويات العليا من القدرة.

3.3. النتائج المتعلقة بالسؤال الثالث: "ما مدى التوافق بين النظرية الكلاسيكية ونظرية استجابة الفقرة في انتقاء فقرات اختبار اللغة الإنجليزية المُعد لهذه الدراسة؟"

للإجابة عن هذا السؤال قامت الباحثة بإجراء مقارنة بين المؤشرات الإحصائية لإنتقاء الفقرات للإختبار التحصيلي المُعد لهذه الدراسة وفقاً للنظرية الكلاسيكية ومؤشرات النموذج ثلاثي المعلم لنظرية استجابة الفقرة، ويبين جدول (10) عدد الفقرات ونسبتها حسب مطابقتها لكلا النظريتين:

جدول (10): عدد الفقرات ونسبتها حسب مطابقتها وفقاً للنظريتين			
حالة الفقرة	عدد الفقرات	الفقرات المحذوفة	النسبة المئوية للفقرات
غير مطابقة في النظريتين	0	0	0 %
غير مطابقة في النظرية الكلاسيكية ومطابقة للنموذج ثلاثي المعلم	4	8, 14, 24, 37	9.52%
غير مطابقة للنموذج ثلاثي المعلم ومطابقة في النظرية الكلاسيكية	6	9, 16, 17, 20, 25	14.3%
مطابقة في النظريتين	32	-	76.2%
المجموع الكلي	42		100%

يبين الجدول (10) مايلي:

- عدم وجود اتفاق بين المؤشرات الإحصائية لإنتقاء فقرات الاختبار بين كل من النظرية الكلاسيكية والنموذج ثلاثي المعلم عند استبعاد الفقرات؛ فقد تم استبعاد الفقرات (8,14,24,37) تبعاً للنتائج التي أفرزتها نظرية الاختبار الكلاسيكية، والتي تُشكل نسبة (9.52%) من الفقرات، بينما تم استبعاد الفقرات (9,16,17,20,25,35) في النموذج ثلاثي المعلم والتي تشكل نسبة (14.28%) من الفقرات.
- طابقت (38) فقرة من فقرات الاختبار أسس ومعايير انتقاء الفقرات تبعاً للنظرية الكلاسيكية، وهو ما يشكل نسبة (90.48%) من فقرات الاختبار.
- طابقت (36) فقرة من فقرات الاختبار أسس ومعايير انتقاء الفقرات للنموذج ثلاثي المعلم لنظرية استجابة الفقرة، والتي شكلت مانسبته (85.71%) من فقرات الاختبار الأصلي.
- عند الجمع بين أسس انتقاء الفقرات لكل من النظرية الكلاسيكية والنموذج ثلاثي المعلم فسيكون الإختبار بصورته النهائية مكون من (32) فقرة، والذي يشكل ما نسبته (76.2%) من فقرات الاختبار الأصلي، وقد تم حساب نسبة الإتفاق بين نتائج المؤشرات لإنتقاء الفقرات بشكل عام في كل من نظرية الكلاسيكية في الاختبار والنموذج ثلاثي المعلم من خلال معادلة كوبر (Cooper, 1974):

$$76.2\% = 100 \times \frac{32}{4+6+32} \quad \text{عدد مرات الاتفاق،} \quad 100 \times \frac{\text{عدد مرات الاتفاق}}{\text{عدد مرات الاختلاف} + \text{عدد مرات الاتفاق}}$$

تُشير النتائج المُستخرجة من المعادلة إلى وجود نسبة اتفاق بين النظرية الكلاسيكية والنموذج ثلاثي المعلم عند انتقاء فقرات الاختبار التحصيلي المُعد لهذه الدراسة تصل إلى (76.2%)، ويمكن تفسير الإختلاف بينهما إلى اختلاف طريقة انتقاء الفقرات بين النظريتين؛ فالنظرية الكلاسيكية تسعى إلى إيجاد الفقرات الصالحة للقياس بدلالة معالمها والإفادة من ذلك في خصائص الاختبار ومُعتمدة في ذلك على خصائص عينة المفحوصين، بينما تسعى نظرية استجابة الفقرة إلى تدريج الفقرات (Item Calibration) عند تقدير سمات المفحوصين مع ضمان استقلالية تقدير هذه المعالم عن بعضهما (أبو فودة، 2016). وبشكل عام تمتع كلا الاختبارين بدرجة متوسطة إلى مرتفعة من الصعوبة والتمييز، كما اتفقتا على أصعب الفقرات وأسهلها وأكثرها تمييزاً.

وقد اتفقت هذه الدراسة مع دراسة (Abedalaziz & Leng, 2013) بتشابه خصائص الفقرات في النموذج الكلاسيكي مع النموذج ثلاثي المعلم بشكل خاص، ومع دراسة كلاً من (Fan, 1998)، وجمحاوي (2000)، وحجازي والخطيب (2014) في وجود تشابه بين معالم الفقرات بين النظرية الكلاسيكية ونظرية استجابة الفقرة بنماذجها الثلاث بشكل عام؛ حيث كانت قيم معاملات الارتباط بين معالم الفقرات في هذه الدراسات مرتفعة. بينما اختلفت هذه الدراسة بنسبة التوافق بين النظرية الكلاسيكية ونظرية استجابة الفقرة عند انتقاء الفقرات مع دراسة كلاً من سالم (2011) والتي كانت نسبة الاتفاق بين النظريتين (92%)، وأبو فودة (2016) التي كانت نسبة الاتفاق بينهما (96.6%) وحجازي والخطيب (2014) بنسبة اتفاق (97.56%). وقد يُعزى اختلاف النتائج عن الدراسات السابقة إلى اختلاف النماذج المستخدمة عند مقارنة النظريتين، فقد اعتمدت بعض الدراسات سابقة الذكر على النموذج ثنائي المعلم عند مقارنتها، والذي يقتصر على معلمي الصعوبة والتمييز عند انتقاء فقرات الإختبار، إضافة إلى اختلاف المعايير التي تم من خلالها انتقاء فقرات الإختبار في كلا النظريتين، وكذلك قد يعود اختلاف نسب التوافق مع الدراسات السابقة إلى اختلافها في عدد فقرات الاختبارات ونوعها، وأحجام عيناتها وطبيعتها.

4.3. النتائج المتعلقة بالسؤال الرابع: "ما الخصائص السيكومترية (الصدق والثبات) للاختبار التحصيلي في اللغة الإنجليزية المستخدم في هذه الدراسة وفق كل من نظرية الاختبار الكلاسيكية والنموذج ثلاثي المعلم؟"

لقد تم حساب خصائص الاختبار السيكومترية تبعاً لكل من نظرية الاختبار الكلاسيكية ونظرية استجابة الفقرة ممثلةً بالنموذج ثلاثي المعلم، وذلك كما يلي:

أولاً: الثبات:

تم استخدام برنامج (SPSS-19) لتقدير معامل الثبات للاختبار ضمن نظرية الاختبار الكلاسيكية باستخدام معامل الثبات كودر-ريتشاردسون (KR-20)، واستخدام برنامج (IRT-PRO4) لتقدير معامل الثبات التجريبي للنموذج ثلاثي المعلم بعد حذف الفقرات غير المطابقة، ويظهر من خلال جدول (11) قيم معاملات الثبات للاختبار للنظريتين بعد حذف الفقرات غير المطابقة.

جدول (11): قيم معاملات الثبات لكل من النظرية الكلاسيكية والنموذج ثلاثي المعلم بعد حذف الفقرات غير المطابقة

النظرية	معامل الثبات		عدد الفقرات	عدد الطالبات
	قبل الحذف	بعد الحذف		
النظرية الكلاسيكية	0.92	0.89	38	1076
النموذج ثلاثي المعلم	0.79	0.78	36	

يُبين جدول (11) أن قيمة معامل ثبات الاختبار في النظرية الكلاسيكية بعد حذف الفقرات قد انخفضت عن قيمة الثبات للنموذج الأصلي وهو (0.92) لتصبح (0.89)، كما انخفضت قيمة معامل الثبات التجريبي في النموذج ثلاثي المعلم والتي قد بلغت (0.79) لتصبح بعد حذف الفقرات غير المطابقة لتصبح (0.78)، ويعود انخفاض الثبات بشكل عام في كلا النظريتين إلى انخفاض عدد الفقرات والذي له تأثير مباشر على معاملات الثبات؛ حيث يتأثر معامل الثبات بعدد الفقرات وطول الاختبار (علام، 2011). بالإضافة إلى ذلك فقد يعود انخفاض الثبات في النموذج ثلاثي المعلم عن النظرية الكلاسيكية إلى تأثر معامل الثبات التجريبي للنموذج ثلاثي المعلم بالفقرات المستبعدة والتي قد يكون لها تحيز بكمية المعلومات، مما يؤدي إلى انخفاض تباين القدرة وارتفاع تباين الخطأ؛ وهو ما يبرر انخفاض الثبات في النموذج ثلاثي المعلم.

ثانياً: الصدق:

تم التحقق من صدق الاختبار المرتبط بمعك باستخدام تحصيل الطالبات في مادة اللغة الإنجليزية كمحك؛ حيث تم حساب معاملات الصدق المحكي لصورتي الاختبار الناتجتين من استخدام كل من النظرية الكلاسيكية والنموذج ثلاثي المعلم، فتم حساب معامل ارتباط بيرسون بين درجات الطالبات على الاختبار الذي أعدته الباحثة ومعدل درجات الطالبات على مادة اللغة الإنجليزية خلال الفصل الدراسي، والذي يمثل مقياس آخر لنفس السمة المقاسة، ويجدر الذكر إلى انخفاض عدد الطالبات إلى (1000) طالبة لعدم تمكن الباحثة من الحصول على معدل درجاتهن من المدرسة، وبين جدول (12) قيم معاملات الصدق المحكي الناتجين من استخدام كل من النظرية الكلاسيكية والنموذج ثلاثي المعلم قبل وبعد حذف الفقرات الغير مطابقة.

جدول (12): قيم معاملات الصدق المحكي الناتجين من كل من النظرية الكلاسيكية والنموذج ثلاثي المعلم

النظرية	معامل الصدق		عدد الفقرات	عدد الطالبات
	قبل الحذف	بعد الحذف		
النظرية الكلاسيكية	0.60	0.59	38	1000
النموذج ثلاثي المعلم	0.60	0.60	36	

يبين جدول (12) أن قيمة معامل الصدق المحكي للاختبار في النظرية الكلاسيكية قد انخفضت لتصبح (0.59)، بينما لم تتغير قيمة معامل الصدق للنموذج ثلاثي المعلم (0.60)، وهي إجمالاً قيم جيدة وكافية لتعكس نسبة التباين المشترك بين كل من الاختبار التحصيلي والمحك تبعاً لما أشار له (أبو علام، 2011)، حيث اعتبر أن الحصول على معاملات مرتفعة للصدق المحكي وتزيد عن (0.6 - 0.5) أمراً صعباً لعدم اعتبار المحك اختباراً مكافئاً للاختبار المراد قياس صدقه. كما يبين جدول (12) أن قيمة معامل الصدق المحكي في النموذج ثلاثي المعلم أعلى بنسبة بسيطة عن النظرية الكلاسيكية، وقد يعود ذلك إلى انخفاض الخطأ المعياري في التقدير بعد حذف الفقرات وانخفاض التباين بين درجات الاختبار والمحك.

5.3. النتائج المتعلقة بالسؤال الخامس: "هل توجد فروق ذات دلالة إحصائية لمعاملات الثبات المحسوبة وفق كل من نظرية الاختبار الكلاسيكية والنموذج ثلاثي المعلم؟"

تم الكشف عن دلالة الفروق بين معاملي الثبات المحسوب باستخدام إجراءات كل من النظرية الكلاسيكية والنموذج ثلاثي المعلم في نظرية استجابة الفقرة باستخدام الحزمة الإحصائية كوكرون (Cocron). وهي حزمة إحصائية جزئية ضمن الحزمة الإحصائية (R- Package)، والتي تستخدم إحصائي (X2) للكشف عن الفروق في المعاملات بالإعتماد على ماخلص له فلدت (Feldet, 1980) ضمن هذا السياق. وبين جدول (13) نتائج اختبار الفروق بين معاملات الثبات المستخرجة تبعاً لنظرية الاختبار الكلاسيكية مرة والنموذج ثلاثي المعلم مرة أخرى.

جدول (13): نتائج دلالة الفروق بين معاملي الثبات في النظريتين

المعالجة وفق:	معامل الثبات	عدد الفقرات	عدد الطالبات	(X <sup>2</sup> )	درجات الحرية	الدلالة الإحصائية
نظرية الاختبار الكلاسيكية	0.89	38	1076	119.6915	1	0.000
النموذج ثلاثي المعلم	0.78	36				

ويُشير جدول (13) إلى أن قيمة (X<sup>2</sup> = 119.69)، وهي قيمة دالة إحصائياً عند مستوى (0.05)، والذي يعني وجود فروق ذات دلالة إحصائية في تقدير معاملي الثبات بين كل من نظرية الاختبار الكلاسيكية والنموذج ثلاثي المعلم لصالح نظرية الاختبار الكلاسيكية، ويمكن إرجاع هذا الاختلاف في قيم معاملات الثبات بين النظريتين نظراً لاختلاف طريقة تقديرهما واختلاف عدد الفقرات في كل اختبار؛ فبالإضافة إلى انخفاض عدد الفقرات في



اختبار النموذج ثلاثي المعلم مقارنة باختبار النظرية الكلاسيكية، فقد تم تقدير معامل الثبات كيبورر- ريتشاردسون (KR-20) في نظرية الاختبار الكلاسيكية، والذي يعكس ثبات الاختبار بشكل عام ولجميع مستويات القدرة مع افتراض تساوي قيمة الخطأ المعياري، بينما يعتمد تقدير معامل الثبات التجريبي للنموذج ثلاثي المعلم على تقدير الثبات والخطأ المعياري لكل مستوى من مستويات القدرة؛ وهو ما يبرر انخفاض الثبات وارتفاع الخطأ المعياري عند مستويات القدرة المتدنية في النموذج ثلاثي المعلم.

وقد اتفقت هذه النتيجة مع دراسة جمحاوي (2000) التي أظهرت فروقاً دالة إحصائياً بين معاملي الثبات لصالح النظرية الكلاسيكية والتي قاربت لهذه الدراسة في حجم العينة الذي بلغ (1061) وعدد فقرات الاختبار الذي بلغ (39) فقرة وكذلك في المرحلة العمرية التي تم تطبيق الدراسة عليها، بينما اختلفت مع دراسة أبو فودة (2016) والتي اشارت إلى تفوق النموذج ثنائي المعلم لنظرية استجابة الفقرة بخاصية الثبات، والتي قد اختلفت مع هذه الدراسة في النموذج المستخدم، وحجم العينة والتي شملت (140) طالب في الصف السادس، وتطبيق اختبار في مادة الرياضيات مكون من (30) فقرة.

6.3. النتائج المتعلقة بالسؤال السادس: "هل توجد فروق ذات دلالة إحصائية لمعاملات الصدق المحسوبة وفق كل من نظرية الاختبار الكلاسيكية والنموذج ثلاثي المعلم؟"

يهدف التعرف على دلالة الفروق بين معاملات الصدق المحسوبة وفقاً للنظرية الكلاسيكية والنموذج ثلاثي المعلم لنظرية استجابة الفقرة فقد تم استخدام اختبار (t-test) لمعاملات الارتباط للعينات المترابطة بدرجات حرية (n-3)، وذلك من خلال استخدام المعادلة المقترحة من قبل كلاً من جلاس وهوكنز لمقارنة معاملي الصدق (Glass & Hokins, 1984). وهي كما يلي:

$$t = (r_{xy} - r_{xz}) \sqrt{\frac{(n-3)(1+r_{yz})}{2(1-r_{xy}^2 - r_{xz}^2 - r_{yz}^2 + 2r_{xy}r_{xz}r_{yz})}}$$

حيث أن:

n: عدد أفراد العينة.

$r_{xy}$ : معامل الصدق للنموذج ثلاثي المعلم.

$r_{xz}$ : معامل الصدق للنظرية الكلاسيكية.

$r_{yz}$ : معامل ارتباط الدرجة الكلية المحسوبة تبعاً لنظرية الكلاسيكية مرة ولنموذج ثلاثي المعلم مرة أخرى.

ويبين جدول (14) نتائج اختبار (t-test) لدلالة الفروق في معاملات الصدق المحكي لكل من نظرية الاختبار الكلاسيكية ونظرية استجابة الفقرة.

جدول (14): نتائج دلالة الفروق بين معاملي الصدق المحكي في النظرية الكلاسيكية ونظرية استجابة الفقرة							
المعالجة وفق:	معامل الصدق	عدد الفقرات	عدد الطالبات	عدد المحسوبة	(T) درجات الحرية	(T) الجدولية	الدلالة الإحصائية
النظرية الكلاسيكية	0.59	38	1000	2.33	997	1.96	دالة إحصائياً
النموذج ثلاثي المعلم	0.60	36					

يلاحظ من جدول (14) وجود فروق في تقدير معاملي الصدق باستخدام كل من النظرية الكلاسيكية والنموذج ثلاثي المعلم ولصالح النموذج ثلاثي المعلم؛ حيث بلغت قيمة (T) المحسوبة (2.33) عند درجة الحرية (997)، وهي قيمة أعلى من قيمة (T) الجدولية (1.96) عند مستوى الدلالة (0.05). ويمكن تفسير ذلك لارتفاع قيم معاملات الارتباط بين الفقرات في النموذج ثلاثي المعلم وتجانسها وتأكيدها على وجود سمة واحدة تقيسها؛ وهو ما يؤكد قوة افتراضات نظرية استجابة الفقرة، مقارنة بافتراضات نظرية الاختبار الكلاسيكية والتي يمكن اعتبارها سهلة التحقيق لمعظم البيانات وسهلة الإنهاء (Harris, 1989)، كما وأنه بالرغم من انخفاض معامل الثبات للنموذج ثلاثي المعلم مقارنةً بنظرية الاختبار الكلاسيكية لإنخفاض عدد الفقرات وتأثر معامل الثبات بكمية المعلومات التي يقدمها إلا أن ارتفاع قيمة صدق الاختبار للنموذج ثلاثي المعلم يعكس الموثوقية بالنتائج المستخلصة منه؛ حيث أن الصدق اشمل من الثبات وهي خاصية متقدمة على الثبات. (عودة، 2014)

وقد اتفقت هذه النتيجة مع دراسة كلاً من سالم (2011) ودراسة أبو فودة (2016) والتي أظهرت فروقاً ذات دلالة إحصائية بين معاملات الصدق المحكي المحسوب وفقاً لنظرية الاختبار الكلاسيكية والنموذج ثنائي المعلم ولصالح النموذج ثنائي المعلم، بينما اختلفت هذه النتيجة مع دراسة حجازي والخطيب (2014) التي لم تظهر فروقاً دالة إحصائياً بين معاملات الصدق المحكي لكل من نظرية الاختبار الكلاسيكية والتي بلغ معامل الصدق فيها (0.73) والنموذج ثنائي المعلم الذي بلغ معامل الصدق فيه (0.65)

## 7.3. التوصيات ومقترحات الدراسة:

- إجراء المقارنة بين الخصائص السيكومترية للنظرية الكلاسيكية والنموذج ثلاثي المعلم على اختبارات بتوزيعات مختلفة، للكشف عن سلوك النظريتين في ظل اختلاف التوزيع الإجمالي.
- إجراء بحوث مقارنة بين النظرية الكلاسيكية ونماذج نظرية استجابة الفقرة باستخدام اختبارات تحصيلية لمواد دراسية أخرى، وعلى مراحل دراسية مختلفة.
- توجيه المعلمين والمعلمات إلى استخدام نماذج نظرية استجابة الفقرة عند تحليل الاختبارات التحصيلية للمواد الدراسية المختلفة بهدف تحسين جودة الإختبار المختلفة.

## المراجع:

## أولاً: المراجع العربية:

1. حجازي، تغريد والخطيب، عبدالله (2014). التوافق بين النظرية الكلاسيكية والنموذج ثنائي المعلمة في مطابقة فقرات اختبار محكي المرجع في أحكام التلاوة والتجويد. مجلة جامعة النجاح للأبحاث (العلوم الإنسانية): 28 (10): 2241-2269.
2. جمحاوي (2000). مقارنة خصائص الفقرات وفق النظرية التقليدية ونظرية استجابة الفقرة في مقياس للقدرة الرياضية. رسالة ماجستير غير منشورة. جامعة اليرموك، الأردن.
3. سالم، حسني (2011). مدى التوافق بين النموذج ثنائي المعلمة والنظرية الكلاسيكية في بناء اختبار تحصيلي في مبحث العلوم العامة للصف السادس الأساسي. رسالة ماجستير غير منشورة. جامعة اليرموك، الأردن.
4. الشافعي، محمد (2014). الإحصاء التقليدي والمتقدم في البحوث العلمية والإنسانية: أسس نظرية وتطبيقية باستخدام برامج-SPSS-LISREL. AMOS الرياض: مكتبة الرشد.
5. الشمراني، محمد (2014). استخدام نظرية استجابة الفقرة والنظرية التقليدية في تقدير الخصائص الإحصائية السيكومترية لاختبار مهارات التفكير والتعلم لطلبة السنة التحضيرية الجامعية. مجلة كلية التربية: 751 (2): 717-802.
6. بني عطا، زايد والشريفين، نضال (2012). أثر اختلاف شكل توزيع القدرة على معالم الفقرة ودالة المعلومات للاختبار. المجلة الأردنية في العلوم التربوية: 8 (2): 151-166.
7. علام، صلاح الدين (2011). القياس والتقويم التربوي والنفسي -أساسياته وتطبيقاته وتوجهاته المعاصرة. القاهرة: دار الفكر العربي.
8. أبو علام، محمود (2011). مناهج البحث في العلوم النفسية والتربوية. القاهرة: دار النشر.
9. عودة، أحمد (2014). القياس والتقويم في العملية التدريسية. إربد، الأردن: دار الأمل.
10. عيد، غادة (2004). الدرجة الحقيقية المقدرة باستخدام نظرية السمات الكامنة والنظرية الكلاسيكية. "دراسة سيكومترية". مجلة جامعة أم القرى للعلوم التربوية والاجتماعية والإنسانية: 16 (2): 228-287.
11. أبو فودة، باسل (2016). التوافق بين النظرية التقليدية في القياس ونظرية استجابة الفقرة في مطابقة فقرات محكي المرجع في وحدة الهندسة التحليلية. دراسات عربية في التربية وعلم النفس (ASEP): 73: 189-215.
12. المومني، رنا (2011). مقارنة مطابقة الفقرات بين نموذج موكن اللابارامتري والنموذج ثنائي المعالم البارامتري. رسالة دكتوراه غير منشورة، كلية التربية، جامعة اليرموك، إربد، الأردن.
13. المومني، رنا (2017). التكامل بين التحليل العاملي الاستكشافي والتوكيدي كطريقتين من البنية العاملية لمقياس مكثري للكفاءات المتعددة (الصورة السعودية). مجلة العلوم التربوية والنفسية: 18 (4): 503-542.
14. النعيمي، عز الدين (2015). معالم الفقرات والأفراد وخاصة اللاتغاير في الإختبارات الوطنية لضبط جودة التعليم في الأردن مقارنة بين النظرية الكلاسيكية والنظرية الحديثة في القياس. مجلة اتحاد الجامعات العربية للتربية وعلم النفس، سوريا، 13 (1): 135-136.
15. هياجنة، علي (2011). أثر طريقة انتقاء الفقرات في دالة معلومات الاختبار. رسالة دكتوراه غير منشورة، كلية التربية، جامعة اليرموك، الأردن.
16. الوسيدي، أسامة (2017). مقارنة خصائص المفردات وفق النظرية التقليدية ونظرية الاستجابة للمفردة للقياس باستخدام المحاكاة. رسالة ماجستير غير منشورة، كلية التربية، جامعة أم القرى، المدينة المنورة، السعودية.

ثانياً: المراجع الأجنبية:

1. Abedalaziz, A. & Leng, C. H. (2013). The Relationship between CTT and IRT Approaches in Analyzing Item Characteristics. *The Malaysian Online Journal of Educational Science*, 1(1): 64-70.
2. Adedoyin, O. & Adedoyin, J. (2013). Assessing the Comparability between Classical Test Theory and Item Responses Theories (IRT) Models in Estimating Test Item Parameters. *Herlad Journal of Education and General Studies*, 2 (3): 107-114.
3. Adegoke, B. A. (2013). Comparison of Item Statistics of Physics Achievement Test using Classical Test and Item Response Theory Frameworks. *Journal of Education and Practice*, 4(22), 87-95.
4. Alharbi, K. A. (2003). *An Empirical Investigation of Comparability and Invariance of Classical Test Theory and Item Response Theory Person and Item Parameter Estimates*. Unpublished doctoral dissertation, Ohio University.
5. Allen, M. & Yen W. (1979). *Introduction to Measurement Theory*. Monterey, California: Brooks\ Cole Publishing Company.
6. Awopeju, O. A. & Afolabi, E. R. I. (2016). Comparative Analysis of Classical Test Theory and Item Response Theory Based Item Parameter Estimates of Senior School Certificate Mathematics Examination. *European Scientific Journal*, 12(28). <https://doi.org/10.19044/esj.2016.v12n28p263>
7. Browne, M. & Cudeck, R. (1993). *Alternative ways of assessing model fit*. In Bollen, K.A. & Long, J.S. [Eds.] *Testing structural equation models*. Newbury Park, CA: Sage, 136-162.
8. Cai, L., Thissen, D., & du Toit, S. H. C. (2011). *IRTPRO for Windows [Computer software]*. Lincolnwood, IL: Scientific Software International.
9. Cooper, J. (1974). *Measuring and Analysis of Behavioural Techniques*. Columbus, Ohio: Charles E. Morit.
10. Courville, T. (2004). *An Empirical Comparison of Item Response Theory and Classical Test Theory Item/Person Estimate*. Unpublished doctoral dissertation, A&M University, Texas.
11. Crocker, L. & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. New York: Holt, Rinehart & Winston.
12. Ebel, R. L. (1965). *Measuring Educational Achievement*. Englewood Cliffs, N. J: Prentice-Hall.
13. Eleje, L. I., Onah, F. E. & Abanobi, C. C. (2018). Comparative study of classical test theory and item response theory using diagnostic quantitative economics skill test item analysis results. *European Journal of Educational and Social Sciences*, 3 (1): 71 - 89.
14. Fan, X. (1998). Item Response Theory and Classical Test Theory: An Empirical Comparison of their Item/Person Statistics. *Educational and Psychological Measurement*, 58 (3): 357-385. <https://doi.org/10.1177/0013164498058003001>
15. Glass, G. V., & Hopkins, K. D. (1984). *Statistical methods in education and psychology*. Upper Saddle River, NJ: Prentice-Hall.
16. Güler, N., Uyanık, G. K. & Teker, G. T. (2013). Comparison of Classical Test Theory and Item Response Theory in Terms of Item Parameters. *European Journal of Research on Education*, 2(1): 1-6.
17. Hambleton, R. K., & Jones, R. W. (1993). Comparison of Classical Test Theory and Item Response Theory and their Applications to Test Development. *Educational Measurement: Issues and Practice*, 12(3): 38-47
18. Hambleton, R. K. & Swaminathan, H. (1985). *Item response theory: Principles and Applications*. Boston, MA: Kluwer Academic Publishers.
19. Hambleton, R.K., Swaminthan, H. & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage Publications.
20. Hambleton, R. K. & Van del Linden, W. J. (1982). Advances in item response theory and applications: an introduction. *Applied Psychological Measurement*, 6: 373-378.
21. Harris, D. (1989). Comparison of 1-, 2-, and 3-parameter IRT models. *Educational Measurement: Issues and Practice*, 8, 35-41.
22. Kyungtae, K. (2015). *Model Comparisons among Testlet Responses Theories (IRT) on a Reading Comprehension Test*. Unpublished doctoral dissertation, Middle Tennessee State University.
23. Lord, F.M. (1953). A Theory of Test Scores. *Psychometric Monographs*, No. 7.
24. Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Erlbaum.
25. Mead, A. D., & Meade, A. W. (2010). *Item selection using CTT and IRT with unrepresentative samples*. Paper presented at the 25<sup>th</sup> Annual Meeting of the Society for Industrial and Organizational Psychology, Atlanta, GA.

26. Meseganaw, N. (2014). Master thesis Department of Statistics Masteruppsats, Statistiska institutionen Applying Item Response Theory Models to the Survey of Adult Skills (PIAAC) Geremew.
27. Momani, Rana TH. (2018). Using Item Response Theory to Evaluate Self-Directed Learning Readiness Scale. *Journal of Educational and Developmental Psychology*, 8(1): 14. <https://doi.org/10.5539/jedp.v8n1p14>
28. Reckase. D. (1979). Unifactor Latent Trait Models Applied to Multifactor Tests: Results and Implications. *Journal of Educational Statistics*, 4(3): 207-230. <https://doi.org/10.2307/1164671>
29. Spela, P. & Gregor, S. (2008). An empirical comparison of Item Response Theory and Classical Test Theory. *Horizons of Psychology*, 17(3): 5-24.
30. Spence, R, Owens, M., and Goodyer, I. (2012). Item response theory and validity of the NEO-FII in adolescents. *Pers Individ Dif Journal*. 53(6-4): 801-807, <https://doi.org/10.1016/j.paid.2012.06.002>
31. Warm, T. (1978). *A primer of Item Response Theory*. Oklahoma City: U.S. Coast Guard Institute.
32. Wiberg, M. (2004). *Classical test theory vs. item response theory: An evaluation of the theory test in the Swedish driving-license test* (No. 50). Umea: Kluwer Academic Publications.

## Comparison between classical test theory and the three parameters logistic model for item selection of English language achievement test

Rehab bint Saleh Al-Saikhan

Master's degree in Measurement and Evaluation, Qassim University, KSA  
rehab-alsaikhan@hotmail.com

Rana Thani Damen Al-Momani

Associate Professor of Measurement and Evaluation, Qassim University, KSA

Received : 3/2/2021 Revised : 20/3/2021 Accepted : 3/4/2021 DOI : <https://doi.org/10.31559/EPS2021.10.1.8>

**Abstract:** This study aimed to examine the differences between Classical Test Theory (CTT) and the three-parameter logistic model in terms of selecting items for an English achievement test. To achieve the aim of the study, an English achievement test consists of (42) multiple-choice with four alternatives items was developed and administered on a sample consisted of (1076) female students randomly selected from (11) female secondary schools.

IRT-PRO4, SPSS-19, AMOS- 24 & Cocron were used for statistical analysis. The results revealed that (38) items fit the Classical Test Theory with percentage of (90.45%) while (36) items fit the three-parameter logistic model with percentage of (85.71 %), and (36) items out of (42) fit both theories

Reliability coefficient (KR-20) for the classical test theory was (0.86) while the empirical reliability coefficient value for the three-parameter model was (0.78). The criterion-related validity coefficient was (0.59) according to the classical theory, and (0.60) according to the three-parameter model. The results indicated that there are statistically significant differences between the two reliability coefficients resulting from the two theories in the favor of the Classical Test Theory. Results also revealed statistically significant differences between the two criterion-related validity coefficients resulting from the two theories in the favor of the three-parameter logistic model.

Overall, IRT provides richer information about measurement precision. The findings support the use of IRT for test construction and development in the educational fields.

**Keywords:** *Classical Test Theory; Item Response Theory; Three Parameter Logistic Model; Item selection.*

### References:

1. 'lam, Slah Aldyn (2011). Alqyas Waltqwym Altrbwy Walnfsy –Asasyath Wttbyqath Wtwjhath Alm'asrh. Alqahrh: Dar Alfkr Al'rby.
2. Abw 'lam, Mhmwd (2011). Mnahh Albhth Fy Al'lwm Alnfsyh Waltrbwyh. Alqahrh: Dar Alnshr.
3. Bny 'ta, Zayd Walshryfyn, Ndal (2012). Athr Akhtlaf Shkl Twzy' Alqdrh 'la M'alm Alfqrh Wdahl Alm'lwmat Llakhtbar. Almjhl Alardnyh Fy Al'lwm Altrbwyh: 8 (2): 151-166.
4. 'wdh, Ahmd (2014). Alqyas Waltqwym Fy Al'mlyh Altdrysyh. Erbd, Alardn: Dar Alaml.
5. 'yd, Ghadh (2004). Aldrjh Alhqyqyh Almqrh Bastkhdam Nzryt Alsmat Alkamnh Walnzryh Alkasykyh."Drash Sykwmttryh". Mjlt Jam't Am Alqra Ll'lwm Altrbwyh Walajtma'yhwalensanyh: 16(2): 228-287.
6. Abw Fwdh, Basl (2016). Altwafq Byn Alnzryh Altqlydyh Fy Alqyas Wnzryt Astjabh Alfqrh Fy Mtabqh Fy Mtabqh Fqrat Mhky Almrj' Fy Whdt Alhndsh Althlylyh. Drasat 'rbyh Fy Altrbyh W'Im Alnfs (Asep): 73: 189-215.
7. Hjazy, Tghryd Walkhty, 'bdalh (2014). Altwafq Byn Alnzryh Alkasykyh Walnmwdj Thna'y Alm'lmh Fy Mtabqh Fqrat Akhtbar Mhky Almrj' Fy Ahkam Altlawh Waltjwyd. Mjlt Jam't Alnjah Llabhath (Al'lwm Alensanyh): 28 (10): 2241-2269.

8. Hyajnh, 'ly (2011). Athr Tryqt Antqa' Alfqrar Fy Dalh M'lwmat Alakhtbar. Rsalt Dktwrah Ghyr Mnshwrh, Klyt Altrbyh, Jam't Alyrmwk, Alardn.
9. Jmhawy (2000). Mqarnt Khsha's Alfqrar Wfq Alnzryh Altqlydyh Wnzryt Astjabh Alfqrh Fy Mqyas Llqdrh Alryadyh. Rsalt Majstyr Ghyr Mnshwrh. Jam't Alyrmwk, Alardn.
10. Almwmny, Rna (2011). Mqarnt Mtabqh Alfqrar Byn Nmwdj Mwkn Allabaramtry Walnmwdj Thna'y Alm'alm Albaramtry. Rsalt Dktwrah Ghyr Mnshwrh, Klyt Altrbyh, Jam't Alyrmwk, Arbd, Alardn.
11. Almwmny, Rna (2017). Altkaml Byn Althlyl Al'amly Alastkshafy Waltwkydy Ktryqtyn Mn Albnyh Al'amlyh Lmqyas Mknzy Lldka'at Almt'ddh (Alswrh Als'wdyh). Mjlt Al'lwm Altrbwyh Walnfsyh: 18 (4): 503-542.
12. Salm, Hsny (2011). Mda Altwafq Byn Alnmwdj Thna'y Alm'lmh Walnzryh Alklasykyh Fy Bna' Akhtbar Thsyly Fy Mbhth Al'lwm Al'amh Llsf Alsads Alasasy. Rsalt Majstyr Ghyr Mnshwrh. Jam't Alyrmwk, Alardn.
13. Alshafy, Mhmd (2014). Alehsa' Altqlydy Walmtqdm Fy Albhwth Al'lmyh Walensanyh: Ass Nzryt Wttbyqyh Bastkhdam Bramj. Spss-Lisrel-Amos Alryad: Mktbt Alrshd.
14. Alwsydy, Asamh (2017). Mqarnt Khsha's Almfrdat Wfq Alnzryh Altqlydyh Wnzryt Alastjabh Llmfrdh Llqyas Bastkhdam Almhakah. Rsalt Majstyr Ghyr Mnshwrh, Klyt Altrbyh, Jam't Am Alqra, Almdynh Almnwrh, Als'wdyh.