أثر طريقة اختيار المموهات في فقرات الاختيار من متعدد في خصائصها السيكومترية وفق النظرية الكلاسيكية والنظرية الحديثة في القياس

نايل عودة الكعابنه[1] و أحمد سليمان عودة[2]

(1) كلية المجتمع بحريملا، جامعة شقراء، شقراء، المملكة العربية السعودية
(2) قسم علم النفس التربوي، كلية التربية، جامعة اليرموك، إربد، الأردن

الملخص

هدفت الدراسة إلى بيان أثر طريقة اختيار المموهات في فقرات الاختيار من متعدد على القيم المقدرة لمؤشرات معالم الفقرة (الصعوبة، والتمييز، والتخمين)، ومؤشر معلمة القدرة، وفق النظرية الكلاسيكية والنظرية الحديثة في القياس.

ولتحقيق أهداف الدراسة تم إعداد اختبار تحصيلي لطلبة الصف التاسع الأساسي في مادة الرياضيات، يشتمل على (41) فقرة، وقد تم بناء مموهات تلك الفقرات في ثلاثة نماذج وفق ثلاث إستراتيجيات مختلفة. واستخدم برنامج (SPSS) وبرنامج (BILO-MG) لتقدير المؤشرات الإحصائية اللازمة.

وقد تم التوصل إلى أنه لا يوجد أثر لطريقة اختيار المموهات لفقرات الاختيار من متعدد على مواقع الأفراد على تدريج متصل القدرة، لكن يوجد فروق ذات دلالة إحصائية بين متوسطات الأخطاء المعيارية لتقديرات معلمة القدرة، كذلك يوجد فروق ذات دلالة إحصائية بين قدرات المفحوصين المقدرة بالعلامة الخام تعزى للطريقة المتبعة في بناء مموهات فقرات الاختبار، كما يوجد فروق ذات دلالة إحصائية بين تقديرات معلمة الصعوبة، وتقديرات معلمة التمييز والأخطاء المعيارية في تقديرها، وبين الأخطاء المعيارية لتقديرات معلمة التخمين، بينما لا توجد فروق ذات دلالة إحصائية بين تقديرات معلمة التخمين، ولا بين الأخطاء المعيارية لتقديرات معلمة الصعوبة. أما النتائج وفق النظرية الكلاسيكية فأشارت إلى وجود فروق ذات دلالة إحصائية بين قيم معاملات الصعوبة وقيم معاملات التمييز في ضوء اختلاف طريقة اختيار المموهات في فقرات اختبار الاختيار من متعدد.

الكلمات المفتاحية: إستراتيجية اختيار المموهات، نظرية القياس الحديثة، نظرية القياس الكلاسيكية.

المقدمة

يُلاحظ المتتبع لحركة القياس والتقويم في المؤسسات التعليمية بشكل عام، والاختبارات بشكل خاص تزايد الاهتمام باختبارات الاختيار من متعدد، لما لها من فوائد في جميع المجالات التعليمية والمهنية، وقد أشاد بها كثير من الباحثين؛ منهم Gay (1980)، الذي بين أن اختبارات الاختيار من متعدد تستطيع أن تقيس نواتج التعلم في جميع المستويات العقلية، ويمكن أن تغطي مجالاً واسعًا من السلوكيات، وبالتالي تكون أكثر صدقًا وثباتًا من بقية الاختبارات.

بينما أشار عودة (2004) إلى أن اختبارات الاختيار من متعدد هي أكثر الاختبارات شيوعًا وانتشارًا، خاصة في اختبارات التحصيل المقننة، فهي قادرة على قياس نواتج تعلم في المستويات العليا من المجال المعرفي.

ويجب الأخذ بعين الاعتبار أن هناك عدة متغيرات أساسية تؤثر في الخصائص السيكومترية للاختبار والخصائص السيكومترية للفقرة (الصعوبة، والتمييز، والتخمين) عند تطوير اختبار من نوع الاختيار من متعدد. أول هذه المتغيرات هو عدد المموهات؛ حيث أكدت معظم الدراسات مثل دراسة Straton and Catts (1980) ودراسة Budescu and Nevo (1985) أن اختلاف عدد المموهات يؤثر في الخصائص السيكومترية للاختبار ولفقراته، وفي هذا الصدد يرى Ebel and Frisbie (1986) أن اختلاف عدد المموهات في اختبارات الاختيار من متعدد يؤدي إلى اختلاف عدد المقارنات للإجابة المعطاة؛ حيث يؤثر في صعوبة الفقرة ومعامل تمييزها؛ حيث إن انخفاض القدرة على تمييز الفقرة يجعل اختيارها مرتبطًا بالصدفة، وهذا لا يقيس المقدرة الحقيقية للطلبة. ومدى صعوبة الفقرة يؤثر أيضًا في ثبات الاختبار، لأنه عندما تكون الفقرات متوسطة الصعوبة يكون الثبات أعلى منه في حال الفقرات السهلة جدًّا أو الصعبة جدًّا.

أما المتغير الثاني المؤثر في الخصائص السيكومترية للاختبار والفقرة فهو شكل المموهات التقليدية، والتي تتلخص في وجود إجابة واحدة صحيحة وعدد من المموهات، أو مركبة؛ كأن تحتوي على

مموهات على شكل «لا شيء مما ذكر» أو «جميع ما ذكر صحيح» أو الجمع بين المموهات مثل «أ+ب»، وهذا ما يلاحظ في كثير من الاختبارات التحصيلية التي يعدها المعلمون (Haladyna *et al*., 2002).

ومن ناحية أخرى، يتم استخدام المموهات المركبة لفقرات الاختبار، إما لتوفير الوقت والجهد في بناء مموهات جيدة، أو لعدم توفر الخبرة لدى مطوري الاختبارات بالمحتوى الدراسي أو بالمفحوصين، أو لعدم وجود إستراتيجية معينة يتم من خلالها بناء مموهات مناسبة لفقرات الاختبارات (عودة، 2004م).

وأشارت بعض الدراسات مثل دراسة Crehan and Haladna (1991) إلى أن الفقرات التي تحوي مموهات مركبة تؤدي إلى زيادة الصعوبة والتمييز لتلك الفقرات.

أما المتغير الثالث المؤثر في الخصائص السيكومترية للاختبار والفقرة فهو الطريقة التي يتم من خلالها اختيار المموهات لفقرات اختبار الاختيار من متعدد، فمن التوصيات المطروحة في الدراسات السابقة للقياس والتقويم في المؤسسات التعليمية أن تكون المموهات متجانسة مع بعضها البعض، مما يزيد من قدرة المموهات على جذب انتباه المفحوصين؛ حيث يتوقع أن تزداد صعوبة فقرات الاختبار بزيادة تجانس المموهات، وإذا كانت مموهات الفقرة ضعيفة وغير جذابة يتوقع أن يؤدي ذلك إلى انخفاض صعوبة الفقرة، ويتوقع أن ينعكس ذلك على القدرة على تمييزها؛ حيث إن الفقرات التي معامل صعوبتها قريب من الصفر أو الواحد تكون قدرتها التمييزية ضعيفة (Murphy and Hofer, 1994). وتأثير جاذبية المموهات يمثل المتغير الرابع المؤثر في الخصائص السيكومترية للاختبار وخصائص الطلاب، وطبيعة المادة أيضًا من العوامل المؤثرة في الخصائص السيكومترية للاختبار.

ونجد أن أي إجراء لاختيار المموهات قد يكون له دور في تحديد خصائص الفقرات والاختبار، وبالتوازي مع هذه المتغيرات فإن التوصيات التي يمكن اتباعها من قبل المختصين والمطبقين للاختبارات في المدارس والجامعات والمؤسسات التعليمية الأخرى يتوقع أن تؤثر في خصائص الاختبار كأداة قياس، مما يعني أن مدى الاهتمام بتطوير اختبار من نوع الاختيار من متعدد يعكس مدى الاهتمام بتطبيق القياس العلمي في المجال التربوي والنفسي.

النظرية الحديثة في القياس

تعرف النظرية الحديثة في القياس باسم نظرية السمات الكامنة (Latent Trait Theory)، كما تسمى بنظرية استجابة الفقرة (Item Response Theory)، لاهتمامها بالربط بين استجابة المفحوص لفقرة اختبار ذات خصائص معينة وقدرته؛ أي تفترض أنه يمكن التنبؤ بأداء المفحوص، أو يمكن تفسير أدائه في اختبار نفسي أو عقلي، في ضوء خاصية أو مجموعة خصائص مميزة لهذا الأداء تسمى السمات، ونظرًا لصعوبة ملاحظة هذه السمات بصورة مباشرة، فإنه يجب تقديرها أو الاستدلال عليها من خلال أداء المفحوص على مجموعة من فقرات الاختبار، ولهذا السبب يطلق عليها السمات الكامنة (Hamblton and Swaminathan, 1985) .

وأشارت الأدبيات ذات الصلة بالنظرية الحديثة في القياس إلى أنها تقدم الكثير من الحلول لمشاكل بناء الاختبارات وتطويرها، كما تتمتع بخصائص تجعلها أكثر تلبية لحاجات الباحثين ومطوري الاختبارات، بسبب أن النظرية الكلاسيكية تعاني من جوانب قصور في تلبية حاجات الباحثين ومطوري الاختبارات؛ منها (Hamblton and Swaminathan, 1985):

1. اختلاف خصائص الفقرات باختلاف عينات المفحوصين؛ حيث إن خصائص فقرات الاختبار تعتمد على مستوى وتوزيع قدرات أفراد هذه العينات؛ إذ تختلف صعوبة الفقرات باختلاف قدرة أفراد العينة. أما تمييز الفقرات فيكون مرتفعًا عندما يكون هناك تباين بين أفراد العينة الواحدة والعكس صحيح. أما معامل ثبات الاختبار فإنه يتأثر بمدى التباين بين أفراد العينة؛ فإذا كان تباين أفراد العينة أكثر من تباين أفراد المجتمع فإن هذا ينعكس على معامل ثبات الاختبار؛ إذ يكون أكبر مما هو عليه.
2. افتراض النظرية الكلاسيكية أن الأخطاء المعيارية في القياس لها نفس التوزيع عند جميع

مستويات القدرة، وهذا يتنافى مع حقيقة أن تباين الخطأ يختلف باختلاف مستويات القدرة.

3. أن أداء الفرد على مجموعة جزئية من فقرات اختبار ما، لا يقدم معلومات بالقدر الذي يقدمه من خلال أدائه على الاختبار ككل.
4. أن مفهوم الثبات للاختبار في النظرية الكلاسيكية يقوم على أساس إمكانية تطوير صورتين متكافئتين من ذلك الاختبار، ومن الصعب إيجاد هذه الصور المتكافئة بسبب أداء الأفراد؛ حيث يختلف باختلاف صور الاختبار.
5. أن الاختبارات المبنية على أساس النظرية الكلاسيكية تكون فقراتها ملائمة للأفراد متوسطي القدرة أكثر من الأفراد ذوي القدرات المرتفعة والمتدنية، لذلك لا تستطيع تقدير قدرات الأفراد ذوي القدرات المرتفعة والمتدنية.

وبالمقارنة بالنظرية الحديثة في القياس نجد أنها قد عملت على تلافي هذا القصور؛ فمثلاً عند استخدام نماذج هذه النظرية في ظل مجتمع كبير من الأفراد تكون تقديرات الخصائص السيكومترية للفقرات مستقلة عن عينة الأفراد التي استخدمت في تقدير هذه الخصائص؛ أي لا تتأثر معالم الفقرات بالعينة المختارة من المجتمع.

وبوجود مجموعة كبيرة نسبيا من فقرات الاختبار التي تقيس نفس السمة أوضحت النظرية الحديثة في القياس أن تقدير قدرات الأفراد تكون متحررة من الفقرات، ومن هنا برزت أهمية اللاتغير (Invariance) في تقدير معالم الفقرات بين مختلف مجموعات الأفراد، كما وفرت النظرية الحديثة في القياس مؤشرًا إحصائيًّا لتقدير درجة الدقة في قياس القدرة لكل فرد.

وتعتمد النظرية الحديثة في القياس على افتراض وجود متصل للسمة؛ حيث يمكن تقدير احتمال الإجابة الصحيحة للفرد عن فقرة الاختبار، إذا علم موقعه على هذا المتصل، كما تفترض أن مقدار هذا الاحتمال يكون دالة متزايدة لموقع الأفراد على متصل السمة، تسمى منحنى خصائص الفقرة (Item Characteristic Curve)، وهذا يعني أن احتمال الإجابة الصحيحة للفرد يزداد بزيادة مقدار السمة الكامنة لديه (Hamblton and Swaminathan, 1985).

كذلك تعتمد النظرية الحديثة في القياس على مجموعة من الافتراضات، التي يجب أن تتوفر أو تتحقق في البيانات المراد تحليلها، وهذه الافتراضات هي أحادية البعد، والاستقلال الموضعي، والتحرر من السرعة في الإجابة كما وضحها Hamblton *et al.* (1991) حيث يعتمد الاختيار المناسب للنموذج على مدى تحقق هذه الافتراضات.

ونظرًا لاختلاف أشكال منحنيات خصائص الفقرة، ونتيجة لاستخدام هذه النظرية بشكل واسع، فلقد تم تطوير نماذج تصف العلاقة بين أداء الأفراد على الاختبار والسمات الكامنة وراء هذا الأداء بطريقة رياضية.

وتصنف نماذج النظرية الحديثة في القياس من حيث طبيعة الاستجابة إلى عدة أصناف منها النماذج ثنائية الاستجابة (Dichotomous Model)؛ حيث تستخدم هذه النماذج عندما تكون الاستجابات ثنائية؛ أي عندما تكون إجابة المفحوص صحيحة يأخذ العلامة واحد، وعندما تكون إجابته خطأ يأخذ العلامة صفر، والبيانات المستخدمة مع هذه النماذج تأخذ من فقرات اختبارات الاختيار من متعدد أو من فقرات اختبارات التكميل، ومن فقرات الاختبارات التي تكون من نوع الصواب والخطأ.

ويمكن أن يبرهن على ملاءمة هذه النماذج لمجموعة من البيانات من خلال مطابقة النموذج للعينة لتحديد ما إذا كان الإطار العام لتوزيع الدرجات في مصفوفة البيانات الأصلية مع الإطار العام المتوقع طبقًا للنموذج المستخدم.

وأشهر النماذج ثنائية الاستجابة هو النموذج اللوجستي أحادي المعلمة، أو ما يشار إليه عادة بنموذج راش، والنموذج اللوجستي ثنائي المعلمة، والنموذج اللوجستي ثلاثي المعلمة، وفيما يلي تعريف مختصر كل نموذج:

- النموذج أحادي المعلمة: حيث يفترض هذا النموذج أن جميع فقرات الاختبار تميز بنفس القدر بين المفحوصين، لكنها تتباين فقط في صعوبتها، ويعد هذا النموذج حالة خاصة من النموذج الثنائي والثلاثي، إلا أنه يتميز بخصائص تجعله أكثر تطبيقًا واستخدامًا لسهولة التعامل معه رياضيًّا مقارنة بالنماذج الأخرى، ويستخدم هذا

النموذج في تحليل الأسئلة ثنائية التدريج، والمعادلة الرياضية لهذا النموذج هي (Hamblton and Swaminathan, 1985):

$$P_i(\theta) = [1 + EXP(-D \times a_i(\theta - b_i)\ ]^{-1}$$

$$, i = 1,2,3,....,n \quad .....(1)$$

حيث:

$P_i(\theta)$: احتمال أن يجيب المفحوص ذو القدرة θ على الفقرة i إجابة صحيحة.

D : عامل التدريج (Scaling Factor)، وهو ثابت لجميع فقرات الاختبار ويساوي (1.7).

a_i : معامل تمييز الفقرة، وهو ثابت لجميع فقرات الاختبار.

b_i : معامل صعوبة الفقرة.

EXP: هي الأساس اللوغاريتمي الطبيعي.

- النموذج ثنائي المعلمة: حيث يقوم على افتراض اختلاف الفقرات في صعوبتها وتمييزها، كما يفترض غياب التخمين، وهو أكثر النماذج واقعية، لأنه من الصعب إيجاد مجموعة من الفقرات لها نفس القدرة التمييزية على مستويات القدرة المختلفة. والمعادلة الرياضية لهذا النموذج هي Hamblton and Swaminathan, 1985):

$$P_i(\theta) = [1 + EXP(-D \times a_i(\theta - b_i)\ ]^{-1}$$

$$, i = 1,2,3,....,n \quad .....(2)$$

حيث:

a_i : معامل تمييز الفقرة، وهو متغير لجميع فقرات الاختبار.

- النموذج ثلاثي المعلمة: حيث يعد امتدادًا للنموذج الثنائي، فهو يضيف معلمًا جديدًا للفقرة هو معامل التخمين c_i والذي يحدد احتمال أن يجيب المفحوص إجابة صحيحة عن الفقرة بالتخمين. وتم الاعتماد على هذا النموذج لأغراض هذه الدراسة، والمعادلة الرياضية لهذا النموذج هي:

$$P_i(\theta) = c_i + (1 - c_i)[1 + EXP(-D \times a_i(\theta - b_i)\ ]^{-1}$$

$$, i = 1,2,3,....,n \quad .....(3)$$

مشكلة الدراسة

هناك الكثير من الأسئلة حول مدى انعكاس إجراءات اختيار المموهات في فقرات الاختيار من متعدد ثنائية التدريج على خصائص الفقرات وخصائص الاختبار الذي يتكون من مجموعة الفقرات، وكذلك على مخرجات التحليل للبرمجيات التي تعتمد نظرية استجابة الفقرة؛ فقد يتم اختيار المموهات وفق أسس منطقية أو وفق أسس إحصائية، ويعتمد الكثير من مطوري الاختبارات على الأسس المنطقية في ضوء محددات معينة مبنية على خبرتهم في محتوى المادة، أو خبرتهم التراكمية في المفحوصين، وقد يتم الاعتماد على فحص جاذبية المموهات إحصائيًّا وخاصة في الاختبارات المقننة.

الأدبيات المتوفرة في القياس لم تشر إلى اختيار المموهات من إجابات المفحوصين أنفسهم مقارنة بالمموهات التي يتم اختيارها وفق أسس منطقية أو إحصائية. وفحص انعكاسات هذا الاختيار على خصائص الفقرات والاختبار يحتاج إلى دراسات تجريبية، وتشكل الدراسة الحالية منطلقًا بحثيًّا في هذا المجال؛ لذا جاءت هذه الدراسة للمقارنة بين ثلاث إستراتيجيات يتم من خلالها بناء مموهات لفقرات اختبار من نوع الاختيار من متعدد، وبيان ما مدى تأثير إستراتيجية الاختيار للمموهات على الخصائص السيكومترية للاختبار وفقراته وفق النظرية الكلاسيكية والنظرية الحديثة في القياس.

ففي الإستراتيجية الأولى يتم بناء مموهات الاختبار من خلال مجموعة من الخبراء والمختصين في تدريس المحتوى المراد قياس الأهداف التعليمية له. أما في الإستراتيجية الثانية فيتم بناء المموهات من خلال التحليل الإحصائي لإجابات الطلبة عن نماذج الاختبار التي تم الحصول عليها من المختصين والخبراء، وأخيرًا في الإستراتيجية الثالثة، يتم بناء المموهات من خلال إجابات الطلبة عن نموذج الاختبار الذي إجابات فقراته على شكل تكميل، ثم يتم بناء المموهات لكل فقرة من إجابات الطلبة التي تم الحصول عليها. وبالتحديد ستحاول هذه الدراسة الإجابة عن الأسئلة التالية:

- هل تختلف قيم معلمة القدرة والخطأ المعياري في تقديرها باختلاف الإستراتيجية المتبعة في اختيار مموهات فقرات الاختبار وفق النظرية الكلاسيكية والنظرية الحديثة في القياس؟

- هل تختلف قيم معالم الفقرة (الصعوبة، التمييز، التخمين) والأخطاء المعيارية في تقديرها باختلاف الإستراتيجية المتبعة في اختيار مموهات فقرات الاختبار وفق النظرية الحديثة في القياس؟

- هل تختلف قيم معالم الفقرة (الصعوبة، التمييز، التخمين) والأخطاء المعيارية في تقديرها باختلاف الإستراتيجية المتبعة في اختيار مموهات فقرات الاختبار وفق النظرية الكلاسيكية؟

وقد انبثق عن الأسئلة السابقة الفرضيات الإحصائية الآتية:

الفرضية الأولى: لا توجد فروق ذات دلالة إحصائية (a=0.05) بين تقديرات قيم معلمة القدرة والأخطاء المعيارية في تقديرها تعزى إلى اختلاف الإستراتيجية المتبعة في اختيار مموهات فقرات الاختبار وفق النظريتين.

الفرضية الثانية: لا توجد فروق ذات دلالة إحصائية (a=0.05) بين تقديرات قيم معالم الفقرة (الصعوبة، والتمييز، والتخمين)، والأخطاء المعيارية في تقديرها تعزى إلى اختلاف الإستراتيجية المتبعة في اختيار مموهات فقرات الاختبار وفق النظرية الحديثة في القياس.

الفرضية الثالثة: لا توجد فروق ذات دلالة إحصائية (a=0.05) بين تقديرات قيم معاملات (الصعوبة، التمييز) تعزى إلى اختلاف الإستراتيجية المتبعة في اختيار مموهات فقرات الاختبار وفق النظرية الكلاسيكية.

أهداف الدراسة

تهدف الدراسة إلى:

- بيان تأثير اختلاف الإستراتيجية المتبعة في بناء مموهات اختبارات الاختيار من متعدد على معالم الفقرات ومعلمة القدرة وفق النظرية الكلاسيكية والنظرية الحديثة في القياس.
- تقديم المقترحات والتوصيات التي تساعد على بناء اختبارات الاختيار من متعدد.

أهمية الدراسة

تتضح أهمية هذه الدراسة في ما يلي:

- دراستها للخصائص السيكومترية للاختبار ولفقراته وفق النظرية الكلاسيكية والنظرية الحديثة في القياس، ومحاولتها التوصل إلى تبريرات علمية باستخدام ثلاث إستراتيجيات مختلفة لبناء مموهات لفقرات اختبار الاختيار من متعدد، وتحديد الإستراتيجية المناسبة لاختيار تلك المموهات.
- يعد اختيار المموهات من أهم العناصر التي يتم مراعاتها في صياغة أسئلة الاختيار من متعدد، وهي المحددة في النهاية لخصائص الفقرات وبالتالي خصائص الاختبار.
- قد تسهم هذه الدراسة في زيادة الاهتمام ببناء اختبارات الاختيار من متعدد، وخاصة الاختبارات التي يتم تطبيقها في المدارس.
- قد تسهم هذه الدراسة في تقديم المقترحات الخاصة ببناء مموهات اختبارات الاختيار من متعدد لذوي الاختصاص في وزارة التربية والتعليم، وباقي المؤسسات التعليمية الأخرى عندما يتم بناء المموهات من خلال أجوبة الطلبة عن تلك الفقرات.

حدود الدراسة

- اقتصرت عينة الدراسة على طلبة الصف التاسع الأساسي، من مدارس تربية محافظة الزرقاء الأولى والثانية خلال العام الدراسي (2010/ 2011).
- اقتصرت الدراسة على موضوع تحليل المقادير الجبرية من مساق الرياضيات الذي يدرسه طلبة الصف التاسع الأساسي.
- استخدم في هذه الدراسة النموذج اللوجستي ثلاثي المعلمة للفقرات.

المصطلحات الإجرائية

*** إستراتيجية اختيار مموهات فقرات اختبار الاختيار من متعدد:** يقصد بها مجموعة من الإجراءات والطرق التي يتم اتباعها في عملية بناء المموهات لفقرات اختبار الاختيار من متعدد، وتتألف من ثلاث إستراتيجيات:

- الإستراتيجية الأولى: يتم بناء مموهات فقرات الاختيار من متعدد من خلال عرض فقرات الاختبار على مجموعة من الخبراء والمختصين في تدريس المحتوى المراد قياس الأهداف التعليمية له، ويطلب منهم وضع أربعة مموهات لكل فقرة تكون الإجابة الصحيحة من ضمنها.
- الإستراتيجية الثانية: يتم بناء المموهات من خلال التحليل الإحصائي لإجابات الطلبة عن النماذج التي تم الحصول عليها من خلال

الإستراتيجية الأولى.
- الإستراتيجية الثالثة: يتم بناء المموهات من خلال إجابات الطلبة عن فقرات الاختبار، بحيث تكون الفقرات على شكل تكميل، دون التغيير في أي متن لأي فقرة من فقرات الاختبار.

*** ثبات الاختبار:**
- وفق النظرية الكلاسيكية: ويقصد به معامل الاتساق الداخلي، ويتم حسابه من خلال معادلة كودر ريتشاردسون (20) (20 - KR).
- وفق النظرية الحديثة في القياس: ويقصد به الثبات الإمبريقي (التجريبي) للاختبار، ويتم استخراج الثبات التجريبي للعلامات حسب طريقة الأرجحية العظمى باستخدام برنامج (BILOG-MG)، من خلال إيجاد ناتج قسمة تباين العلامة الحقيقية على تباين العلامة؛ حيث تباين العلامة هو تباين العلامات المستخرج من خلال الأرجحية العظمى، أما التباين الحقيقي فيتم تقديره عن طريق طرح تباين الخطأ من تباين العلامة.

*** الخصائص السيكومترية للفقرة:**

معامل الصعوبة:
- حسب النظرية الكلاسيكية في القياس: النسبة بين عدد الطلبة الذين أجابوا عن الفقرة الاختيارية إجابة صحيحة وعدد الطلبة الذين حاولوا الإجابة عنها.
- حسب النظرية الحديثة في القياس: نقطة على متصل السمة يكون عندها احتمال إجابة المفحوص للفقرة إجابة صحيحة يساوي $(1 + c_i) / 2$

معامل التمييز:
- حسب النظرية الكلاسيكية في القياس: مؤشر على قدرة الفقرة على التمييز بين الطلبة، ويمكن تقديره من خلال إيجاد معامل ارتباط علامات الطلبة على الفقرة مع علاماتهم على الاختبار، بعد إزالة الفقرة نفسها من الاختبار (Corrected Item Total Correlation).
- حسب النظرية الحديثة في القياس: هو الميل عند النقطة التي يكون عندها احتمال إجابة المفحوص للفقرة إجابة صحيحة يساوي $(1 + c_i) / 2$.

جاذبية المموهات:
- حسب النظرية الكلاسيكية في القياس: تساوي أعداد الطلبة الذين أجابوا عن أي مموه من المموهات، سواء كانوا من الفئة العليا أو الفئة الدنيا (عودة، 2005).
- حسب النظرية الحديثة في القياس: احتمال الإجابة عن الفقرة من قبل الأشخاص ذوي القدرات المتدنية جدًّا (Hamblton and Swaminathan, 1985).

الدراسات السابقة

هناك العديد من الدراسات التي اهتمت بالاختبارات وخاصةً اختبارات الاختيار من متعدد، وكيفية بناء تلك الاختبارات وتحديد أفضل المموهات وعددها في فقرات اختبار الاختيار من متعدد؛ ففي دراسة Landrum *et al.* (1993) كان هدفها معرفة مدى التغير في صعوبة الفقرات عندما يتغير عدد مموهاتها من أربعة إلى ثلاثة. ولتحقيق أهداف دراستهم، اختير 46 طالبًا مسجلين في مساق علم النفس العام في الفصل الدراسي الأول خريف عام (1990) في جامعة (Wisconsin-Patteville) بأمريكا، وطبق عليهم الاختبار النهائي لهذا المساق؛ حيث تكون الاختبار من 250 فقرة غطت جميع المواضيع المكونة لهذا المساق وعددها خمسة مواضيع (50 فقرة لكل موضوع). كما طبق اختبار نهائي لنفس المساق على مجموعة أخرى مكونة من 67 طالبًا من الطلبة المسجلين لنفس المساق في الفصل الدراسي الثاني من العام التالي، وتكون الاختبار من 250 فقرة (50 فقرة لكل موضوع)، منها 144 فقرة مشتركة مع الاختبار الأول في العام 1990.

ولغايات الدراسة حذف أحد مموهات فقرات هذا الاختبار بناءً على الخبرة الذاتية لمدرس المساق ليصبح عدد مموهات فقراته ثلاث فقط. وتم مقارنة أداء المفحوصين على الفقرات المشتركة بين الاختبارين. وبينت النتائج أن أداء المفحوصين على الفقرات ذات المموهات الثلاثة في جميع المواضيع

عدا الموضوع الرابع أفضل من أدائهم على الفقرات ذات المموهات الأربعة لنفس المواضيع، وهذا يعني أن الفقرات ذات المموهات الثلاثة كانت أسهل من الفقرات ذات المموهات الأربعة. وعند تعديل صعوبة الفقرات لأثر عامل التخمين وذلك بطرح احتمال إجابة الفقرة بالصدفة من قيمة صعوبتها تبين أنه لا يوجد فروق إحصائية بين صعوبة الفقرات ذات المموهات الثلاثة وصعوبة الفقرات ذات المموهات الأربعة بالنسبة لفقرات الموضوع الأول والثاني والثالث. أما فقرات الموضوع الرابع والخامس وفقرات الاختبار ككل فكانت الفروق بينها دالة إحصائيًا.

وتكونت دراسة Cizek and Oday (1994) من دراستين مستقلتين؛ حيث هدفت الدراسة الأولى إلى معرفة دقة التحكيم في تحديد المموه غير الفعال مقارنة بالبيانات المتوفرة من تحليل الفقرات. ولتحقيق هذا الهدف قام الباحثان ببناء اختبار في الطب الاختصاصي مكون من 32 فقرة من نوع الاختيار من متعدد، بحيث يكون لفقراته خمسة مموهات من ضمنها الإجابة الصحيحة، وتم عرض الاختبار على لجنة مكونة من 12 خبيرًا في مجلس الامتحانات الطبي، وذلك لتحديد المموه غير الفعال في مموهات كل فقرة. ثم تم تطبيق هذا الاختبار على عينة مكونة من 700 مفحوص؛ حيث أشارت النتائج إلى وجود اتفاق بين طريقتي التحكيم والتجريب لتحديد المموه غير الفعال في 24 فقرة من فقرات الاختبار.

وفي دراستهما الثانية التي هدفت للمقارنة بين طريقة التحكيم وطريقة التجريب في حذف المموهات، تم تطبيق نفس فقرات الاختبار السابق، بحيث تم تقليص عدد المموهات في أول 16 فقرة إلى أربعة مموهات من ضمنها الإجابة الصحيحة باستخدام طريقة التحكيم، وتم تقليص عدد المموهات في الست عشرة فقرة الباقية إلى أربعة مموهات باستخدام طريقة الحذف التجريبي، وتم تطبيق الاختبار على عينة مكونة من 726 مفحوصًا، وأشارت النتائج إلى أنه لا توجد فروق ذات دلالة إحصائية في متوسط صعوبة الفقرات، وكذلك في متوسط معاملات التمييز للفقرات تعود لطريقة الحذف.

وقام أبو جراد (1997) بدراسة هدفت إلى معرفة تأثير طريقة الحذف التجريبي والعشوائي للمموهات في فقرات اختبار التحصيل من نوع الاختيار من متعدد على خصائصه السيكومترية (الثبات والصدق)، وعلى خصائص فقراته (الصعوبة، والتمييز). ولتحقيق أهداف دراسته، تم بناء اختبار تحصيلي لطلبة الصف العاشر في مساق الرياضيات، وتكون الاختبار من 45 فقرة من نوع الاختيار من متعدد ولكل فقرة ستة مموهات من ضمنها الإجابة الصحيحة، وتم تطبيق الاختبار على عينة استطلاعية مكونة من 150 طالبًا وطالبة، وبناء على نتائج العينة الاستطلاعية تم حذف 5 فقرات، ومن ثم تم تكوين ثلاثة نماذج بطريقة الحذف العشوائي (نموذج لفقراته خمسة مموهات من ضمنها الإجابة الصحيحة، ونموذج لفقراته أربعة مموهات من ضمنها الإجابة الصحيحة، ونموذج لفقراته ثلاثة مموهات من ضمنها الإجابة الصحيحة)، وثلاثة نماذج أخرى بطريقة الحذف التجريبي (نموذج لفقراته خمسة مموهات من ضمنها الإجابة الصحيحة، ونموذج لفقراته أربعة مموهات من ضمنها الإجابة الصحيحة، ونموذج لفقراته ثلاثة مموهات من ضمنها الإجابة الصحيحة).

ثم تم تطبيق النماذج الستة على عينة مكونة من 467 طالبًا وطالبة من طلبة الصف العاشر. وأشارت النتائج إلى أن معاملات الصعوبة لفقرات النموذج المكون بطريقة الحذف العشوائي كانت أفضل من معاملات الصعوبة للنموذج المكون بطريقة الحذف التجريبي والذي له نفس العدد من المموهات، ولم تكن هناك فروق ذات دلالة إحصائية بين باقي النماذج تعزى لطريقة الحذف. أما معاملات التمييز فلم يكن هناك فروق ذات دلالة إحصائية بين النماذج ذات الفقرات المتساوية من حيث عدد المموهات تعزى لطريقة الحذف.

وفي دراسة يعقوب (2000) التي هدفت للمقارنة بين ثلاثة نماذج من فقرات الاختيار من متعدد من حيث الصعوبة والتمييز؛ حيث كانت فقرات النموذج الأول لها ثلاثة مموهات، وفقرات النموذج الثاني لها أربعة مموهات، أما فقرات النموذج الثالث فكانت بأربعة مموهات منها المموه «لا شيء مما ذكر». ولتحقيق أهداف دراسته، قام الباحث بكتابة 32 فقرة من نوع الاختيار من متعدد بالصور الثلاث السابقة، ووزعت عشوائيًا

في ثلاثة نماذج اختبار لمساق مقدمة القياس والتقويم كل منها مكون من 75 فقرة، وبعد ذلك وزعت نماذج الاختبار على عينة مكونة من 124 طالبًا وطالبة منهم 65 طالبًا و 59 طالبة من طلاب جامعة اليرموك المسجلين في مساق مقدمة في القياس والتقويم لعام (1997/ 1998). ولقد أشارت النتائج إلى أنه لا يوجد فروق ذات دلالة إحصائية بين قيم صعوبة وتمييز الفقرات ذات المموهات الثلاثة وصعوبة وتمييز الفقرات ذات المموهات الأربعة، كما بينت النتائج أنه لا يوجد فرق دال إحصائيًا بين صعوبة وتمييز الفقرات ذات المموهات الثلاثة والفقرات التي تحتوي المموه «لا شيء مما ذكر». كما بينت النتائج أيضا أن الفرق بين قيم صعوبة الفقرات ذات المموهات الأربعة والفقرات التي تحتوي المموه «لا شيء مما ذكر» لم يكن دال إحصائيًا. أما الفرق بين قيم معاملات تمييز هذين النوعين من الفقرات فقد كان دالاً إحصائيا.

وهدفت دراسة Pelton (2002) إلى مقارنة الدقة والاستقرار في تحديد معلمة الصعوبة ومعلمة القدرة وفق النظرية الكلاسيكية والنظرية الحديثة في القياس عن طريق توليد بيانات بأسلوب المحاكاة. ولقد توصلت الدراسة إلى أن تقديرات القدرة يمكن مقارنتها من خلال النظرية الكلاسيكية والنظرية الحديثة في القياس؛ حيث تختلف تقديرات القدرة حسب كمية المعلومات المتوفرة في بيانات المحاكاة، كما أن تقديرات القدرة تتأثر بأحادية البعد، والاختلاف في صعوبة الفقرات، ودرجة التخمين مقارنة بقدرات المفحوصين، كما توصلت الدراسة إلى أن النموذج اللوجستي الثنائي يعطي تقديرات أكثر دقة من النموذجين اللوجستيين الأحادي والثلاثي عندما يكون حجم العينة معتدلاً (999 مفحوصًا) مع عدد فقرات ملائم (33 فقرة) في ظل انخفاض التخمين على اختيار الفقرات.

وفي دراسة قام بها أمبوسعيدي (2005) هدفت إلى تقصي أثر تغيير موقع المشتت القوي (البديل الخطأ الذي اختاره أكبر نسبة من الطلاب من بين المموهات الخطأ الأخرى)، عند استخدام تقنية إبعاد البديل الخطأ في أسئلة الاختيار من متعدد على التحصيل، وعلى بعض الخصائص السيكومترية (معاملات الصعوبة، والتمييز) لفقرات الاختبار؛ حيث تمثلت تقنية إبعاد البديل الخطأ بالطلب من الطالب حذف الإجابتين الخطأ اللتين يعتقد أنهما بعيدتان كليًا عن أن يكونا من ضمن المموهات الصحيحة، ثم يقوم باختيار البديل الصحيح –حسب وجهة نظره– من البديلين الباقيين. ولتحقيق أهداف دراسته، قام الباحث ببناء اختبار في مساق الفيزياء لطلبة الصف العاشر من نوع الاختيار من متعدد، مكون من 24 فقرة، وتم تصميم نموذجين متشابهين من الاختبار مع اختلاف موقع المشتت القوي بالنسبة إلى البديل الصحيح، كما قام الباحث باستخدام تقنية إبعاد البديل الخطأ في نموذجي الاختبار، وتم عرض مجموعة فقرات الاختبار على مجموعة من المحكمين والخبراء في مساق الفيزياء، وتم حساب ثبات الاختبار باستخدام معادلة كودر ريتشاردسون (KR-20)، والتي أعطت قيمة (0.76)، وقد تم تحديد المشتت القوي من خلال تطبيق الاختبار على عينة مشابهة لعينة الدراسة قبل التطبيق على العينة الأصلية. وتم تطبيق أداة الدراسة على عينة مكونة من 133 طالبة من طالبات التعليم العام؛ تم اختيارهن عشوائيًا من مدرستين من مدارس التعليم العام في قطاع ولاية السيب بمحافظة مسقط بسلطنة عُمان. وقد أشارت نتائج الدراسة إلى عدم وجود فروق ذات دلالة إحصائية بين المتوسطات الحسابية لدرجات الطالبات تعزى إلى نموذج الاختبار. كما أشارت النتائج إلى عدم وجود فروق ذات دلالة إحصائية بين المتوسطات الحسابية لمعاملات الصعوبة، ومعاملات التمييز تعزى إلى نموذج الاختبار. كما أوصى الباحث بتدريب الطلبة على استخدام إبعاد المموهات الخطأ.

وأجرى Rodriguez (2005) دراسة هدفت إلى تحليل 27 دراسة تجريبية تهتم بتحديد العدد الأمثل لمموهات فقرات اختبارات الاختيار من متعدد وتأثير ذلك على معلمة الصعوبة والتمييز، وأظهرت النتائج انخفاضًا طفيفًا في صعوبة الفقرات نتيجة تخفيض عدد المموهات من أربعة مموهات إلى ثلاثة مموهات. ومن ناحية أخرى، أصبحت الفقرات سهلة للغاية عندما أصبحت تلك الفقرات تحوي مموهين اثنين فقط. كما أوضحت النتائج أن تخفيض

عدد المفردات إلى مفردتين فقط في الفقرات يؤدي إلى انخفاض تمييز تلك الفقرات، لكن لم يلاحظ ذلك عندما تم تخفيض عدد المفردات من ثلاثة مفردات إلى أربعة مفردات. وأشارت الدراسة إلى أن تخفيض عدد المفردات بشكل أساسي يؤدي إلى عدم مصداقية النتائج، بدون أن تتأثر تلك المصداقية عندما تم تخفيض عدد المفردات من أربعة مفردات إلى ثلاثة مفردات. وإن أخذ هذه النتائج في الاعتبار مع متغيرات أخرى مثل الوقت المطلوب لبناء الاختبار وإدارته جعلت الباحث يستنتج أن الاختبارات التي تحوي ثلاثة مفردات هي أفضل الاختبارات.

كما قام Shizuka *et al*. (2006) بدراسة هدفت إلى مقارنة بين الخصائص السيكومترية لفقرات اختبار استيعاب المقروء باللغة الإنجليزية، الذي يؤهل الطلبة للالتحاق بالجامعات اليابانية، وتكون الاختبار من 38 فقرة ذات الاختيار من متعدد، وكانت الفقرات تحوي أربعة مفردات، وفي مرحلة لاحقة تم خفض عدد المفردات من أربعة إلى ثلاثة من خلال حذف المفردة الأقل تمييزًا، وتم بناء الاختبار في نموذجين، النموذج الأول تكون من فقرات لها ثلاثة مفردات، والنموذج الثاني تكون من فقرات لها أربعة مفردات. وتم تطبيق نموذجي الاختبار على عينتين مختلفتين من الطلبة، وتم تقدير معلمة الصعوبة للفقرات في كلا النموذجين؛ حيث أشارت النتائج إلى عدم وجود فروق ذات دلالة إحصائية في تقديرات معلمة صعوبة الفقرات لكلا النموذجين، وإلى عدم وجود تغيير يستحق الذكر في دقة وعدد المفردات المُميزة. وأوصى الباحثون باستخدام الاختبارات التي تحوي ثلاثة مفردات لتوفير الوقت والمال في تطوير الاختبارات وإدارتها.

وهدفت دراسة قام بها أبو فودة (2008) إلى بيان أثر تعلم مهارة حكمة الاختبار على افتراض أحادية البعد، والتقديرات المختلفة لنظرية استجابة الفقرة، ولتحقيق هدف الدراسة تم اختيار مجموعتين (تجريبية، وضابطة)، أعطيت المجموعة التجريبية برنامجًا تدريبيًّا في تعلم مهارة حكمة الاختبار، وقام الباحث بإعداد اختبار تحصيلي لطلبة الصف التاسع الأساسي مكون من 44 فقرة ذات أربعة مفردات، ثم جرى صياغة فقرات الاختبار في نموذج آخر مخالف لقواعد الصياغة، وجرى تطبيق نموذجي الاختبار على مجموعتي الدراسة التجريبية والضابطة، في عينة مكونة من 2400 طالب من الذكور من مدارس تربية محافظة إربد الأولى والثانية، بواقع 1200 طالب لكل مجموعة ((600) طالب لكل نموذج اختبار)، وأظهرت نتائج الدراسة أن تعلم الطلبة مهارة حكمة الاختبار تؤدي إلى إعطاء تقديرات متحيزة إلى أعلى لمعاملات الثبات. أما قيم معاملات الثبات التجريبي فقد كانت متحيزة إلى أعلى في نموذج الاختبار المخالف لقواعد الصياغة في مجموعة الدراسة الضابطة، وأظهرت النتائج وجود أثر لتعلم حكمة الاختبار على التقديرات الخاصة بمعلمة القدرة، وفيما يتعلق بالأخطاء المعيارية لتقديرات معلمة القدرة، فقد أظهرت النتائج عدم وجود فروق ذات دلالة إحصائية تعزى إلى مجموعة الدراسة. وأظهرت النتائج المتعلقة بالتقديرات الخاصة بمعلمة الصعوبة ومعلمة التمييز عدم وجود فروق ذات دلالة إحصائية تعزى إلى مجموعة الدراسة أو إلى التفاعل بين مجموعة الدراسة ونموذج الاختبار، بينما كان هناك فرق ذو دلالة إحصائية يعزى إلى نموذج الاختبار، لصالح نموذج الاختبار المخالف لقواعد الصياغة، بينما كان هناك فرق ذو دلالة إحصائية بين تقديرات معلمة التخمين يعزى إلى مجموعة الدراسة، لصالح مجموعة الدراسة الضابطة، ووجود فرق ذي دلالة إحصائية يعزى إلى نموذج الاختبار، لصالح نموذج الاختبار المخالف لقواعد الصياغة، ووجود فرق ذي دلالة إحصائية يعزى إلى التفاعل بين مجموعة الدراسة ونموذج الاختبار.

وأجرى Baghaei and Amrahi (2011) دراسة هدفت إلى تحديد العدد الأمثل من المفردات لفقرات اختبارات الاختيار من متعدد وللمقارنة بين إحصائيات الفقرات وثبات الاختبار، ولتحقيق هدف هذه الدراسة قام الباحثان بإعداد ثلاثة نماذج (A,B,C) متكافئة، يحوي كل نموذج 30 فقرة مختلفة فقط في عدد المفردات للفقرات في كل نموذج؛ ففي النموذج (A) كان للفقرات خمسة مفردات من ضمنها الإجابة الصحيحة، وللنموذج (B) أربعة مفردات بعد حذف أحد المفردات

في النموذج (A) عشوائيًّا، وللنموذج (C) ثلاثة مموهات بعد حذف أحد المموهات عشوائيًّا في النموذج (B). وتم تطبيق نماذج الاختبار الثلاثة على عينة عشوائية مكونة من 78 طالبًا و 102 طالبة من طلبة مرحلة البكالوريوس في تخصص اللغة الإنجليزية، وقام الباحثان بحساب الأوساط الحسابية لعلامات الطلبة على نماذج الاختبار الثلاثة، وكذلك حساب الأوساط الحسابية للأخطاء المعيارية، وحساب قيم ثبات نماذج الاختبار، وتقدير معلمة الصعوبة ومعلمة التمييز لفقرات النماذج. وأشارت النتائج إلى عدم وجود فروق ذات دلالة إحصائية بين الأوساط الحسابية لعلامات الطلبة وللأخطاء المعيارية على النماذج الثلاثة، كما أظهرت النتائج عدم وجود فروق ذات دلالة إحصائية بين قيم معاملات الثبات للنماذج الثلاثة، وأظهرت النتائج عدم وجود فروق ذات دلالة إحصائية بين تقديرات معلمة الصعوبة للفقرات في نماذج الاختبار. أما معلمة التمييز فأشارت النتائج إلى وجود فروق ذات دلالة إحصائية؛ حيث كانت تقديرات معلمة التمييز للنموذج الذي تحوي فقراته ثلاثة مموهات هي الأعلى، وأوصى الباحثان بالاعتماد على اختبارات اختيار من متعدد التي تحوي فقراتها ثلاثة مموهات توفيرًا للجهد وللوقت.

الطريقة والإجراءات

مجتمع الدراسة:

تكون مجتمع الدراسة من جميع طلبة الصف التاسع الأساسي الذكور والإناث في المدارس الحكومية التابعة لمديرية تربية الزرقاء الأولى والثانية، خلال الفصل الدراسي الثاني للعام (2010/ 2011)، ويبلغ عدد الطلبة الذكور 5245 طالبًا، وعدد الإناث 5401 طالبة بحسب إحصائيات قسم التخطيط في مديريتي تربية الزرقاء الأولى والثانية.

عينة الدراسة

تكونت عينة الدراسة من 943 طالبًا وطالبة من طلبة الصف التاسع الأساسي، في مديريتي التربية والتعليم لمنطقة الزرقاء الأولى والثانية؛ إذ تم اختيار مجموعة من المدارس التي تحوي الصف التاسع الأساسي في هاتين المديريتين؛ حيث اختيرت 13 مدرسة عشوائيًّا، منها 6 مدارس للذكور و 7 مدارس للإناث.

أداة الدراسة

لتحقيق أهداف الدراسة، تم إعداد اختبار تحصيلي من نوع الاختيار من متعدد بهدف قياس تحصيل طلبة الصف التاسع في مساق الرياضيات (موضوع تحليل المقادير الجبرية)، مكون من 44 فقرة؛ حيث تم عرض الفقرات على مجموعة من المحكمين لإبداء آرائهم ومقترحاتهم من خلال استبانة تم إعدادها لهذا الغرض. ومن خلال تقديرات المحكمين، وبعد المناقشة معهم تم تعديل بعض فقرات الاختبار الغامضة، وقد تراوحت الأوساط الحسابية لتقديرات المحكمين ضمن تدريج ليكرت على فقرات الاستبانة المتعلقة بالحكم على فقرات الاختبار ما بين (3.89–4.42)، وللحكم على جدول المواصفات ما بين (4.45–4.51). أما الوسط الحسابي لتقديرات المحكمين للحكم على مدى تغطية فقرات الاختبار ككل لجدول المواصفات فقد كان (4.43).

أدوات الدراسة

تم بناء المموهات في فقرات الاختبار بثلاث طرق مختلفة:

الطريقة الأولى: تم توزيع نموذج فقرات الاختبار (بدون مموهات)، على مجموعة تتكون من عشرة معلمين، وكانت مجموعة المعلمين التي تم توزيع نموذج فقرات الاختبار عليهم ممن يدرِّسون مساق الرياضيات للصف التاسع الأساسي في المدارس الحكومية التابعة لتربية الزرقاء الأولى والثانية، وتم الطلب من كل معلم كتابة جميع المموهات الممكنة لكل فقرة من فقرات الاختبار، بالإضافة إلى الإجابة الصحيحة للفقرة؛ حيث كان عدد مموهات فقرات الاختبار التي تم الحصول عليها من المعلمين تختلف من حيث العدد لكل فقرة، وبعد ذلك تم حصر جميع المموهات لكل فقرة ووضعها في نموذج واحد فقط، بحيث تحتوي كل فقرة من فقرات النموذج على جميع المموهات التي تم الحصول عليها، وتم عرض هذا النموذج على ثلاثة معلمين من ذوي الخبرة، ويعملون في

مدرسة الملك عبد الله للتميز؛ حيث كان هؤلاء المعلمون من خارج مجموعة المعلمين التي قامت بوضع المموهات، لاختيار أفضل ثلاثة مموهات بالإضافة إلى الإجابة الصحيحة للفقرة بناءً على خبرتهم، وسمي هذا النموذج نموذج (A).

الطريقة الثانية: بعد حصر مموهات كل فقرة من فقرات الاختبار، التي تم الحصول عليها من خلال الطريقة الأولى، تم تكوين أربعة نماذج للاختبار، لكل نموذج أربعة مموهات بالإضافة إلى الإجابة الصحيحة، من خلال التوزيع العشوائي لمموهات كل فقرة على فقرات النماذج الأربعة للاختبار، مع ملاحظة أن هناك تكرارًا لبعض المموهات لكل فقرة من فقرات الاختبار بين النماذج الأربعة للاختبار، وأن هناك اختلافًا في موقع الإجابة الصحيحة لنماذج الاختبار الأربعة. بعد ذلك تم تطبيق النماذج الأربعة على عينة عشوائية بواقع 52 طالبًا وطالبة لكل نموذج، وكان توزع العينة كما يلي: 3 شعب للطلاب بواقع 112 طالبًا من طلاب مدرسة حي معصوم الثانوية للبنين التابعة لمديرية تربية الزرقاء الأولى، و 3 شعب للطالبات بواقع 96 طالبة من مدرسة هند بنت عتبة الثانوية للبنات التابعة لمديرية تربية الزرقاء الثانية. وتم تطبيق نماذج الاختبار الأربعة في نفس الوقت، وبشكل عشوائي على طلبة الشعبة الواحدة، ثم تم تصحيح أوراق الاختبار، وإجراء التحليلات الإحصائية لتحديد أفضل ثلاثة مموهات (الأكثر تمييزًا) لكل فقرة من فقرات الاختبار، بالإضافة إلى الإجابة الصحيحة للفقرة، وسمي هذا النموذج نموذج (B).

الطريقة الثالثة: تم توزيع فقرات الاختبار (على شكل تكميل) على عينة عشوائية تكونت من 138 طالبًا وطالبة من طلبة الصف التاسع، منهم 64 طالبًا و 74 طالبة؛ حيث توزعت العينة على 4 شعب، شعبتين للطلاب في مدرسة المغيرة الثانوية للبنين التابعة لمديرية الزرقاء الأولى، وشعبتين للطالبات في مدرسة «أبو الزيغان» الثانوية للبنات التابعة لمديرية الزرقاء الثانية، بعد أن تم إبلاغ الطلاب عن الاختبار، وعلى أن تكون علامة هذا الاختبار جزءًا من علامة الاختبار الأول للطلاب، وتم تصحيح أوراق الاختبار من قبل الباحثين؛ حيث تم رصد أخطاء الطلاب أثناء إجابتهم عن فقرات الاختبار، وقد تمت الاستعانة بمعلمتين ممن يدرِّسون مساق الرياضيات للصف التاسع؛ حيث قامت كل معلمة بتصحيح أوراق الاختبار ورصد أخطاء الطلاب في كل فقرة من فقرات الاختبار، ثم قام الباحثان بإجراء مقارنات بين أخطاء الطلاب في كل فقرة من فقرات الاختبار، وتحديد الأخطاء الأكثر تكرارًا، لتضمين تلك الأخطاء لمموهات فقرات الاختبار، ليتم تشكيل النموذج الثالث للاختبار، وسمي هذا النموذج نموذج (C). وتم تقدير فعالية المموهات في فقرات الاختيار من متعدد، باستخدام المؤشر الإحصائي المسمي معامل فعالية المموه لتقدير فعالية المموهات حسب النظرية الكلاسيكية في القياس (عودة، 2004).

تطبيق أداة الدراسة على عينة استطلاعية بصورتها الأولية

تم تطبيق نماذج الاختبار الثلاثة التي تم تشكيلها على عينة استطلاعية مكونة من 200 طالب، موزعين على أربع شعب من طلاب مدرسة سعيد بن المسيب الأساسية للبنين، وبعد ذلك صححت الأوراق، وأعطي صفر للإجابة الخطأ ودرجة واحدة للإجابة الصحيحة، وبهذا تكون العلامة الكلية للمفحوص هي مجموع الإجابات الصحيحة عن فقرات الاختبار، وأدخلت البيانات إلى برنامج (SPSS) لتحليلها والحصول على الإحصاءات اللازمة، فقد تراوحت قيم صعوبة الفقرات في النماذج الثلاثة، (0.22–0.60)، (0.19–0.67)، (0.21–0.66) على التوالي، أما قيم معاملات التمييز لفقرات النماذج الثلاثة، المحسوبة على أساس إيجاد معامل الارتباط بين نتائج المفحوصين في كل فقرة ونتائجهم على الاختبار ككل، فتراوحت بين (0.27–0.70)، (0.16–0.60)، (0.16–0.60) على التوالي، وكانت جميع معاملات الصعوبة والتمييز للفقرات مقبولة لأغراض الدراسة حسب إحصائيات (Choi and Bachman, 1992). وتم تقدير بعض المعايير الإحصائية لمعالم الفقرات (الصعوبة، والتمييز، والتخمين)، كما في الجدول (1).

جدول (1): بعض من المعايير الإحصائية لمعالم الفقرات (الصعوبة، والتمييز، والتخمين) لنماذج الاختبار الثلاثة

معالم الفقرات	المعايير الإحصائية	النموذج A	النموذج B	النموذج C
معلمة الصعوبة	أقل قيمة	0.09-	0.07-	0.04-
	أكبر قيمة	2.42	5.05	4.66
	المتوسط الحسابي	0.74	1.60	1.62
	الانحراف المعياري	0.41	0.97	0.92
معلمة التمييز	أقل قيمة	0.86	0.76	0.76
	أكبر قيمة	6.71	3.01	2.96
	المتوسط الحسابي	2.73	1.59	1.58
	الانحراف المعياري	1.43	0.62	0.61
معلمة التخمين	أقل قيمة	0.12	0.14	0.14
	أكبر قيمة	0.36	0.48	0.49
	المتوسط الحسابي	0.24	0.27	0.27
	الانحراف المعياري	0.06	0.08	0.08

ثبات نماذج الاختبار (A,B,C)

تم تقدير معامل ثبات الاختبار للنماذج المعدة باستخدام النظرية الكلاسيكية والنظرية الحديثة في القياس باستخدام معامل كرونباخ ألفا، ونظرية استجابة الفقرة باستخدام برنامج (BILOG–MG) لتقدير الثبات التجريبي (Empirical Reliability) كما في الجدول (2).

جدول (2): معاملات الثبات لنماذج الاختبار(A,B,C) وفق النظرية الكلاسيكية والنظرية الحديثة في القياس

النموذج	النظرية الكلاسيكية	النظرية الحديثة
A	0.94	0.92
B	0.85	0.88
C	0.87	0.84

ومن الجدول (2) نجد أن قيم معاملات الثبات لنماذج الاختبار المقدرة وفق النظرية الكلاسيكية والنظرية الحديثة في القياس مرتفعة ومتقاربة عندما تم تقديرها لكل نموذج وفق النظريتين، وهذه القيم تدل على أن جميع نماذج الاختبار جيدة لأغراض الدراسة.

دلالات الصدق لنماذج الاختبار الثلاثة (A,B,C) بالصورة الأولية باستخدام النظرية الكلاسيكية في القياس:

تم التحقق من دلالات صدق الاختبار باستخدام النظرية الكلاسيكية في القياس، وذلك باستخدام الطريقتين التاليتين:

- صدق المحتوى (Content Validity): وذلك من خلال الخطوات والإجراءات التي اتبعها الباحثان في بناء الاختبار التحصيلي؛ إذ قام الباحثان بدراسة موضوع تحليل المقادير الجبرية، وإعداد النواتج التعليمية، وبناء جدول للمواصفات، ثم عرضت الصورة الأولية للاختبار على مجموعة من الخبراء للأخذ برأيهم.
- صدق المحك (Criterion Validity): وذلك من خلال حساب معامل ارتباط بيرسون بين أداء العينة الاستطلاعية على النماذج الثلاثة (A,B,C)، وعلاماتهم في الاختبار الأول في مادة الرياضيات؛ إذ كانت قيمته (0.84) للنموذج (A)، (0.67) للنموذج (B)، (0.64) للنموذج (C).

التحقق من افتراض أحادية البعد

تم استخدام طريقة المحاور الأساسية

(Principal Axis Factoring)، مع استخدام طريقة (Quartimax Rotation) للعوامل التي كانت الجذور الكامنة لها أكبر من واحد للبيانات الناتجة من استجابات العينة الاستطلاعية على نماذج الاختبار الثلاثة للتحقق من افتراض أحادية البعد.

جمع البيانات

تم تطبيق نماذج الاختبار الثلاثة على عينة الدراسة، وقد تم توزيع نماذج الاختبار (A,B,C)؛ إذ يأخذ الطالب الأول في الجلوس النموذج A، والطالب الذي بجانبه النموذج B، والطالب الذي يليه النموذج C وهكذا، إلى أن يتم توزيع نماذج الاختبار على جميع الطلبة المفحوصين داخل الشعبة المختارة. وبعد ذلك تم فرز كل نموذج من نماذج الاختبار على حدة، وتصحيحه حسب مفتاح التصحيح، وبعد أن تمت إجراءات التصحيح أعطي صفر للفقرة الخطأ ودرجة واحدة للفقرة الصحيحة، وبهذا تكون العلامة الكلية للمفحوص هي مجموع الإجابات الصحيحة عن فقرات الاختبار، وبعد ذلك أدخلت البيانات إلى البرنامج الإحصائي (SPSS)، وبرنامج (BILOG–MG) لإجراء التحليلات الإحصائية اللازمة للإجابة عن فرضيات الدراسة.

مطابقة الأفراد والفقرات

تم تحليل البيانات الخام باستخدام برنامج (BILOG–MG)، بالاعتماد على استجابات الأفراد على نماذج الاختبار الثلاثة، وقد أفرزت نتائج التحليل، وباستخدام اختبار مربع كاي عند مستوى الدلالة ($\propto$=0.01) تبين عدم مطابقة استجابات 27 مفحوصًا على النماذج الثلاثة، ومن ثم تم إعادة التحليل لاختبار مدى مطابقة الفقرات للنموذج اللوجستي ثلاثي المعلمة، وتبين عدم مطابقة بعض الفقرات في النماذج الثلاثة للنموذج اللوجستي ثلاثي المعلمة، فقد اكتفى الباحثان بإسقاط الفقرات الثلاثة المشتركة فقط في النماذج الثلاثة، وبناء على ذلك تم اعتماد عينة المفحوصين المكونة من 917 طالبًا وطالبة، وعينة الفقرات 41 فقرة في كل نموذج، كأساس في إجراء التحليلات الإحصائية المختلفة.

المعالجات الإحصائية

من أجل تَعَرُّف تأثير طريقة اختيار مموهات فقرات الاختبار على مواقع الأفراد على تدريج متصل القدرة للنماذج الثلاثة، تم إيجاد تقدير قدرات الأفراد والأخطاء المعيارية في تقديرها. وبما أن برنامج (BILO–MG) يقوم بتحويل علامات القدرة إلى وسط مقداره صفر، وانحراف معياري مقداره واحد، قام الباحثان بحساب معاملات ارتباط بيرسون بين تقديرات القدرة الناتجة من النموذج اللوجستي ثلاثي المعلمة والعلامة الخام؛ إذ تم حساب معاملات ارتباط سبيرمان بين تدريج مواقع المفحوصين على متصل القدرة:

- تقدير قدرات الأفراد (θ) وذلك باستخدام برنامج (BILOG–MG) لنماذج الاختبار (A,B,C).
- تدريج المفحوصين على متصل القدرة للنماذج الثلاثة.
- حساب معامل ارتباط سبيرمان بين تدريجات قدرات المفحوصين على متصل القدرة على نماذج الاختبار الثلاث (A,B,C).
- لاختبار الفروق بين قيم معاملات الارتباط المترابطة التي ربما قد تعود لطريقة اختيار مموهات فقرات الاختيار من متعدد، تم استخدام المؤشر الإحصائي (F_{obs}) المقترح من قبل Levy (1978)، الذي يتبع توزيع (F) بدرجات حرية تساوي (k-1,q-k-1) وتحسب قيمة (F_{obs}) من العلاقة التالية:

$$F_{obs} = \frac{q-k+1}{(q-1)(k-1)} \times T^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots(4)$$

حيث:

k : عدد معاملات الارتباط، q :عدد أفراد العينة، T^2: إحصائي Hotellings.

ثم تم إيجاد الخطأ المعياري في تقدير معلمة القدرة لكل فرد على كل نموذج، وفحص دقة تقديرات قدرات الأفراد، واستخدم الباحثان تحليل التباين الأحادي ذا القياسات المتكررة للحصول على القيم.

تقدير معالم الفقرات والأخطاء المعيارية في تقديرها

تم تقدير معالم الصعوبة، ومعالم التمييز، ومعالم التخمين لفقرات نماذج الاختبار الثلاث (A,B,C)، باستخدام برنامج (BILO-MG)، وحساب

المتوسطات الحسابية لكل من المعالم المقدرة، ومن ثم استخدام تحليل التباين الأحادي ذي القياسات المتكررة، وذلك لاستخراج دلالة الفروق على كل من معلمة الصعوبة، ومعلمة التمييز، ومعلمة التخمين، والأخطاء المعيارية في تقديرها، ومن ثم تم استخدام اختبار توكي للمقارنات الثنائية.

النتائج

- النتائج المتعلقة بفرضية الدراسة الأولى: لا توجد فروق ذات دلالة إحصائية (a=0.05) بين تقديرات قيم معلمة القدرة والأخطاء المعيارية في تقديرها تعزى إلى اختلاف الإستراتيجية المتبعة في اختيار مموهات فقرات الاختبار وفق النظريتين.
- معلمة القدرة وفق النموذج اللوجستي الثلاثي المعلمة:

تم تقدير معلمة القدرة وفق النموذج اللوجستي الثلاثي المعلمة باستخدام طريقة الأرجحية العظمى الهامشية (MML) بالاعتماد على برنامج (BILOG-MG)، وقد تراوحت في نموذج الاختبار (A) بين (-3.33-2.35)، وفي نموذج الاختبار (B) ما بين (-3.00-2.91). أما في نموذج الاختبار (C) فقد تراوحت ما بين (-3.23-2.09)، ولفحص فرضية الدراسة الأولى تم حساب معامل الارتباط سبيرمان بين تدريج مواقع تدريج الأفراد على متصل القدرة، كما في الجدول (3).

جدول (3): معاملات الارتباط سبيرمان بين تدريج مواقع الأفراد على متصل القدرة لنماذج الاختبار (A,B,C)

النموذج	A	B	C
A	1	0.61	0.59
B	-	1	0.55
C	-	-	1

وتم استخدام الإحصائي (F_{obs}) المقترح من قبل Levy (1978) للكشف عن الفروق بين عدة معاملات ارتباط مترابطة، الذي يتبع توزيع (F) بدرجات حرية تساوي (k-1,q-k-1)؛ حيث بلغت قيمة (F_{obs}) المحسوبة (0.06)، وقيمة (F) الحرجة (3) عند مستوى الدلالة (∝=0.05)، ويلاحظ عدم وجود فروق ذات دلالة إحصائية بين قيم معاملات الارتباط لموقع تدريج المفحوصين على متصل القدرة يعزى للطريقة المتبعة في بناء مموهات فقرات الاختيار من متعدد.

ولقد تم إيجاد الخطأ المعياري لتقديرات القدرة لكل مفحوص، والناتجة من تطبيق نماذج الاختبار، ويبين الجدول (4) المتوسطات الحسابية والانحرافات المعيارية للأخطاء المعيارية في تقدير القدرة.

جدول (4): المتوسطات الحسابية والانحراف المعياري للأخطاء المعيارية لتقديرات معلمة القدرة في نماذج الاختبار (A,B,C)

النموذج	عدد المفحوصين	المتوسط الحسابي	الانحراف المعياري
A	908	0.68	1.55
B	908	0.49	0.64
C	908	0.89	2.41

وتم استخدام تحليل التباين الأحادي ذي القياسات المتكررة للكشف عن وجود فروق ذات دلالة إحصائية (∝=0.05)، بعد أن تم حذف إجابات 9 مفحوصين من النماذج الثلاثة، وذلك لأن مقدار الخطأ المعياري لهما عالٍ جدًّا عند حساب مقدار الخطأ في برنامج (BILOG-MG)، وهذا يعني أن مقدار المعلومات التي يعطيها الاختبار عن هذه القدرات تكون ضئيلة جدًّا، ويبين الجدول (5) نتائج التحليل.

جدول (5): تحليل التباين الأحادي ذو القياسات المتكررة للأخطاء المعيارية لتقديرات القدرة للاختبارات (A,B,C)

مصدر التباين		مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة (F) المحسوبة	الاحتمال
بين الأفراد		3527.79	907	3.89	-	-
داخل الأفراد	بين النماذج	72.48	2	36.24	15.35	0.00
	الخطأ	4282.00	1814	2.36		
الكلي		7882.27	2723	-	-	-

يلاحظ من الجدول (5) وجود فروق ذات دلالة إحصائية بين تقديرات الخطأ المعياري لتقديرات القدرة تعزى للطريقة المتبعة في اختيار مموهات فقرات الاختيار؛ إذ بلغت قيمة (F) المحسوبة (15.35)، باحتمال (0.00)، وهي دالة عند مستوى الدلالة (∝=0.05)، وللكشف عن الفروق الثنائية لتقديرات الخطأ المعياري لتقديرات القدرة، تم استخدام اختبار توكي، ويبين الجدول (6) نتائج التحليل.

جدول (6): نتائج اختبار توكي للكشف عن الفروق الثنائية لتقديرات الخطأ المعياري لتقديرات القدرة

النموذج	A 0.68	B 0.49	C 0.89
A	-	*0.19	-*0.21
B	-	-	-*0.40

*: ذات دلالة إحصائية عند مستوى الدلالة (∝=0.05).

يلاحظ من الجدول (6) وجود فروق ذات دلالة إحصائية في تقديرات الأخطاء المعيارية لتقديرات معالم القدرة بين النموذجين (A) و (B)، وقد كانت قيم الأخطاء في النموذج (A) أعلى، وعند مقارنة النموذجين (A) و(C)، كانت قيم الأخطاء في النموذج (C) أعلى، وعند مقارنة النموذجين (B) و(C)، كانت قيم الأخطاء المعيارية في النموذج (C) أعلى.

النتائج المتعلقة بقدرات الأفراد مقدرة بالعلامة الظاهرية وفق النظرية الكلاسيكية:

لقد تم حساب قدرات الأفراد مقدرة بالعلامة الظاهرية وفق النظرية الكلاسيكية، ثم تم إيجاد الأوساط الحسابية والانحرافات المعيارية لعلامات المفحوصين على نماذج الاختبار، كما في الجدول (7).

جدول (7): الأوساط الحسابية والانحرافات المعيارية لعلامات المفحوصين على نماذج الاختبار (A,B,C)

النموذج	الوسط الحسابي	الانحراف المعياري
A	22.81	3.06
B	21.22	3.07
C	21.18	3.10

وتم استخدام تحليل التباين الأحادي ذي القياسات المتكررة للكشف عن وجود فروق ذات دلالة إحصائية بين المتوسطات الحسابية لعلامات المفحوصين الخام، كما في الجدول (8).

جدول (8): تحليل التباين الأحادي ذو القياسات المتكررة لعلامات المفحوصين الخام الناتجة من تطبيق نماذج الاختبار

مصدر التباين		مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة (F) المحسوبة	الاحتمال
بين الأفراد		141923.88	907	_	_	_
داخل الأفراد	بين النماذج	1573.318	2	786.66	14.08	0.00
	الخطأ	101328.02	1814	55.86		
الكلي		244825.22	2732	_	_	_

يلاحظ من الجدول (8) وجود فروق ذات دلالة إحصائية تعزى لنماذج الاختبار، وللكشف عن الفروق الثنائية بين المتوسطات الحسابية لعلامات المفحوصين على النماذج، تم استخدام اختبار توكي، والنتائج موضحة في الجدول (9).

جدول (9): نتائج اختبار توكي للكشف عن الفروق الثنائية للمتوسطات الحسابية لعلامات المفحوصين على النماذج

النموذج	A 22.81	B 21.22	C 21.18
A	_	*1.59	*1.63
B	_	_	0.03

*: ذات دلالة إحصائية عند مستوى الدلالة (∝=0.05).

يلاحظ من الجدول (9) أن هناك فروقًا ذات دلالة إحصائية بين علامات الطلبة على النموذج (A) والنموذج (B)، وقد كانت المتوسطات الحسابية لعلامات المفحوصين على النموذج (A) أعلى. كما أن هناك فروقًا ذات دلالة إحصائية بين علامات الطلبة على النموذج (A) والنموذج (C)، وقد كانت المتوسطات الحسابية لعلامات المفحوصين على النموذج (A) أعلى. كما يلاحظ عدم وجود فروق ذات دلالة إحصائية بين علامات الطلبة على النموذجين (B) و (C).

النتائج المتعلقة بفرضية الدراسة الثانية: لا توجد فروق ذات دلالة إحصائية (α=0.05) بين تقديرات قيم معالم الفقرة (الصعوبة، والتمييز، والتخمين)، والأخطاء المعيارية في تقديرها تعزى إلى اختلاف الإستراتيجية المتبعة في اختيار مموهات فقرات الاختبار وفق النظرية الحديثة في القياس.

معلمة الصعوبة والخطأ المعياري في تقديرها

تم حساب المتوسطات الحسابية والانحرافات المعيارية لتقديرات معلمة الصعوبة وللأخطاء المعيارية في تقديرها لفقرات النماذج، كما في الجدول (10).

جدول (10): المتوسطات الحسابية والانحرافات المعيارية لتقديرات معلمة الصعوبة والأخطاء المعيارية في تقديرها لفقرات نماذج الاختبار

معالم الفقرات	المعايير الإحصائية	عدد الفقرات	النموذج A	النموذج B	النموذج C
معلمة الصعوبة	أقل قيمة	41	0.44-	0.43-	0.41-
	أكبر قيمة	41	0.92	1.12	1.07
	المتوسط الحسابي	41	0.31	0.45	0.46
	الانحراف المعياري	41	0.29	0.38	0.37
الأخطاء المعيارية	أقل قيمة	41	0.01	0.03	0.04
	أكبر قيمة	41	0.23	0.29	0.28
	المتوسط الحسابي	41	0.09	0.10	0.09
	الانحراف المعياري	41	0.01	0.01	0.01

واستخدم تحليل التباين الأحادي ذو القياسات المتكررة؛ لاستخراج دالة الفروق على معلمة الصعوبة وللأخطاء المعيارية في تقديرها، ويبين الجدول (11) نتائج التحليل.

جدول (11): تحليل التباين الأحادي ذو القياسات المتكررة للمتوسطات الحسابية لتقديرات معلمة الصعوبة وللأخطاء المعيارية في تقديرها لفقرات نماذج الاختبار (A,B,C)

المعلمة	مصدر التباين		مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة (F) المحسوبة	الاحتمال
معلمة الصعوبة	بين الفقرات		13.51	40	0.338	-	-
	داخل الفقرات	بين النماذج	0.63	2	0.317	21.13	0.00
		الخطأ	1.18	80	0.015		
	الكلي		15.32	122	-	-	-
الخطأ المعياري	بين الفقرات		0.291	40	0.007	-	-
	داخل الفقرات	بين النماذج	0.002	2	0.001	3.33	0.080
		الخطأ	0.024	80	0.0003		
	الكلي		0.317	122	-	-	-

يلاحظ من الجدول (11) وجود فروق ذات دلالة إحصائية بين المتوسطات الحسابية لتقديرات معلمة الصعوبة تعزى للطريقة المتبعة في اختيار مموهات فقرات الاختبار، كما يلاحظ عدم وجود فروق ذات دلالة إحصائية بين المتوسطات الحسابية للأخطاء المعيارية لتقديرات معلمة الصعوبة تعزى للطريقة المتبعة في اختيار مموهات فقرات الاختبار. وللكشف عن الفروق الثنائية بين المتوسطات الحسابية لتقديرات معلمة الصعوبة لفقرات النماذج الثلاثة تم استخدام اختبار توكي، كما في الجدول (12).

جدول (12): نتائج اختبار توكي للكشف عن الفروق الثنائية للمتوسطات الحسابية لتقديرات معلمة الصعوبة لنماذج الاختبار

النموذج	A 0.31	B 0.45	C 0.46
A	–	-0.14	*-0.15
B	–	–	-0.01

*: ذات دلالة إحصائية عند مستوى الدلالة (∝=0.05).

يلاحظ من الجدول (12) أن هناك فروقًا ذات دلالة إحصائية بين تقديرات معلمة الصعوبة للنموذجين (A) و(C)، وقد كانت تقديرات معلمة الصعوبة لفقرات النموذج (C) أعلى، ولم يكن هناك فروق ذات دلالة إحصائية بين المتوسطات الحسابية لتقديرات معلمة الصعوبة للنموذجين (A,B)، وللنموذجين (B,C).

معلمة التمييز والخطأ المعياري في تقديرها

تم حساب المتوسطات الحسابية والانحرافات المعيارية لمعاملات التمييز وللأخطاء المعيارية في تقديرها لفقرات نماذج الاختبار، ويبين الجدول (13) نتائج التحليل.

جدول (13): المتوسطات الحسابية والانحرافات المعيارية لتقديرات معلمة التمييز وللأخطاء المعيارية في تقديرها لفقرات نماذج الاختبار

معالم الفقرات	المعايير الإحصائية	عدد الفقرات	النموذج A	النموذج B	النموذج C
معلمة التمييز	أقل قيمة	41	1.13	1.01	1.10
	أكبر قيمة	41	3.45	5.19	4.94
	المتوسط الحسابي	41	2.16	2.58	2.64
	الانحراف المعياري	41	0.68	1.17	1.12
الأخطاء المعيارية	أقل قيمة	41	0.18	0.20	0.22
	أكبر قيمة	41	0.52	0.78	0.84
	المتوسط الحسابي	41	0.34	0.42	0.43
	الانحراف المعياري	41	0.09	0.18	0.18

واستخدم الباحثان تحليل التباين الأحادي ذا القياسات المتكررة؛ لاستخراج دالة الفروق على معلمة التمييز والأخطاء المعيارية في تقديرها، ويبين الجدول (14) نتائج التحليل.

جدول (14): تحليل التباين الأحادي للمتوسطات الحسابية لتقديرات معلمة التمييز وللأخطاء المعيارية في تقديرها لفقرات نماذج الاختبار

المعلمة	مصدر التباين		مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة (F) المحسوبة	الاحتمال
معلمة التمييز	بين الفقرات		109.85	40	2.75	–	–
	داخل الفقرات	بين النماذج	5.49	2	2.74	17.13	0.00
		الخطأ	13.07	80	0.16		
	الكلي		128.41	122	–	–	–
الخطأ المعياري	بين الفقرات		2.14	40	0.061	–	–
	داخل الفقرات	بين النماذج	0.23	2	0.113	18.83	0.00
		الخطأ	0.50	80	0.006		
	الكلي		2.87	122	–	–	–

يلاحظ من الجدول (14) وجود فروق ذات دلالة إحصائية بين المتوسطات الحسابية لتقديرات معلمة التمييز تعزى للطريقة المتبعة في اختيار مموهات فقرات الاختبار، وللكشف عن الفروق الثنائية بين المتوسطات الحسابية لتقديرات معلمة التمييز لفقرات النماذج الثلاثة، تم استخدام اختبار توكي، كما في الجدول (15).

جدول (15): نتائج اختبار توكي للكشف عن الفروق الثنائية للمتوسطات الحسابية لتقديرات معلمة التمييز لنماذج الاختبار

النموذج	A 2.16	B 2.58	C 2.64
A	–	-0.42	*-0.48
B	–	–	-0.04

*: ذات دلالة إحصائية عند مستوى الدلالة (∝=0.05).

يلاحظ من خلال الجدول (15) أن هناك فروقًا ذات دلالة إحصائية بين تقديرات معلمة التمييز للنموذج (A) والنموذج (C)، وقد كانت تقديرات معلمة التمييز لفقرات النموذج (C) أعلى، ولم يكن هناك فروق ذات دلالة إحصائية بين المتوسطات الحسابية لتقديرات معلمة التمييز للنموذجين (A,B)، وللنموذجين (B,C).

كما يلاحظ من الجدول (14) وجود فروق ذات دلالة إحصائية بين المتوسطات الحسابية للأخطاء المعيارية لتقديرات معلمة التمييز تعزى للطريقة المتبعة في اختيار مموهات فقرات الاختبار، وللكشف عن تلك الفروق الثنائية تم استخدام اختبار توكي، كما في الجدول (16).

جدول (16): نتائج اختبار توكي للكشف عن الفروق الثنائية لتقديرات المتوسطات الحسابية للأخطاء المعيارية لمعلمة التمييز لنماذج الاختبار

النموذج	A 2.16	B 2.58	C 2.64
A	–	*-0.08	*-0.09
B	–	–	-0.01

*: ذات دلالة إحصائية عند مستوى الدلالة (∝=0.05).

يلاحظ من الجدول (16) أن هناك فروقًا ذات دلالة إحصائية بين متوسطات الأخطاء المعيارية لتقديرات معلمة التمييز للنموذجين (A) و(B)؛ أي أن الأخطاء المعيارية في النموذج (B) أعلى، كما أن هناك فروقًا ذات دلالة إحصائية بين متوسطات الأخطاء المعيارية لتقديرات معلمة التمييز للنموذجين (A) و(C)؛ أي أن الأخطاء المعيارية في النموذج (C) أعلى، وعدم وجود فروق ذات دلالة إحصائية بين متوسطات الأخطاء المعيارية لتقديرات معلمة التمييز للنموذجين (B) و(C).

معلمة التخمين والخطأ المعياري في تقديرها

أما معلمة التخمين والخطأ المعياري في تقديرها، فقد تم حساب المتوسطات الحسابية لمعلمة التخمين وللأخطاء المعيارية في تقديرها لفقرات نماذج الاختبار، والنتائج موضحة في الجدول (17).

جدول (17): المتوسطات الحسابية والانحرافات المعيارية لتقديرات معلمة التخمين وللأخطاء المعيارية في تقديرها لفقرات نماذج الاختبار

معالم الفقرات	المعايير الإحصائية	عدد الفقرات	النموذج A	النموذج B	النموذج C
معلمة التخمين	أقل قيمة	41	0.109	0.114	0.128
	أكبر قيمة	41	0.356	0.352	0.392
	المتوسط الحسابي	41	0.2248	0.2232	0.2244
	الانحراف المعياري	41	0.0570	0.5112	0.5576
الأخطاء المعيارية	أقل قيمة	41	0.026	0.019	0.019
	أكبر قيمة	41	0.062	0.073	0.072
	المتوسط الحسابي	41	0.03756	0.03632	0.04180
	الانحراف المعياري	41	0.00838	0.01380	0.05222

وتم استخدام تحليل التباين الأحادي ذي القياسات المتكررة؛ لاستخراج دالة الفروق على معلمة التخمين والأخطاء المعيارية في تقديرها، ويبين الجدول (18) نتائج التحليل.

جدول (18): تحليل التباين الأحادي للمتوسطات الحسابية لتقديرات معلمة التخمين وللأخطاء المعيارية في تقديرها لفقرات نماذج الاختبار

المعلمة	مصدر التباين		مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة (F) المحسوبة	الاحتمال
معلمة التخمين	بين الفقرات		0.233	40	0.007	-	-
	داخل الفقرات	بين النماذج	0.000051	2	0.000026	0.026	0.972
		الخطأ	0.071	80	0.001		
	الكلي		0.3014	122	-	-	-
الخطأ المعياري	بين الفقرات		0.015	40	0.00038	-	-
	داخل الفقرات	بين النماذج	0.000282	2	0.000142	9.103	0.000
		الخطأ	0.001	80	0.000015		
	الكلي		0.0160282	122	-	-	-

يلاحظ من الجدول (18) عدم وجود فروق ذات دلالة إحصائية بين المتوسطات الحسابية لتقديرات معلمة التخمين تعزى للطريقة المتبعة في اختيار مموهات فقرات الاختبار، كما يلاحظ وجود فروق ذات دلالة إحصائية بين المتوسطات الحسابية للأخطاء المعيارية لتقديرات معلمة التخمين. وللكشف عن تلك الفروق الثنائية، تم استخدام اختبار توكي، الجدول (19).

جدول (19): نتائج اختبار توكي للكشف عن الفروق الثنائية للمتوسطات الحسابية لتقديرات الأخطاء المعيارية لمعلمة التخمين لنماذج الاختبار

النموذج	A 0.03756	B 0.03632	C 0.04180
A	-	0.00124	*-0.00424
B	-	-	*0.00548

*: ذات دلالة إحصائية عند مستوى الدلالة (∝=0.05).

يلاحظ من الجدول (19) أن هناك فروقًا ذات دلالة إحصائية بين متوسطات الأخطاء المعيارية لتقديرات معلمة التخمين للنموذجين (A) و(C)، وقد كانت الأخطاء المعيارية في النموذج (C) أعلى، كما أن هناك فروقًا ذات دلالة إحصائية بين متوسطات الأخطاء المعيارية لتقديرات معلمة التخمين للنموذجين (B) و(C)، وقد كانت الأخطاء المعيارية في النموذج (B) أعلى، وكذلك أظهر التحليل عدم وجود فروق ذات دلالة إحصائية بين متوسطات الأخطاء المعيارية لتقديرات معلمة التخمين للنموذجين (A) و(B).

- النتائج المتعلقة بفرضية الدراسة الثالثة: لا توجد فروق ذات دلالة إحصائية (α =0.05) بين تقديرات قيم معاملات (الصعوبة، التمييز) تعزى إلى اختلاف الإستراتيجية المتبعة في اختيار مموهات فقرات الاختبار وفق النظرية الكلاسيكية.

معلمة الصعوبة

تم حساب المتوسطات الحسابية والانحرافات المعيارية لمعاملات الصعوبة لفقرات نماذج الاختبار (A,B,C)، كما في الجدول (20).

جدول (20): المتوسطات الحسابية والانحرافات المعيارية لمعاملات الصعوبة لفقرات نماذج الاختبار (A,B,C)

النموذج	أدنى قيمة	أعلى قيمة	الوسط الحسابي	الانحراف المعياري
A	0.35	0.73	0.55	0.09
B	0.29	0.73	0.51	0.11
C	0.30	0.70	0.51	0.11

وتم استخدام تحليل التباين الأحادي ذي القياسات المتكررة؛ لاستخراج دالة الفروق على معلمة الصعوبة، ويبين الجدول (21) نتائج التحليل.

جدول (21): تحليل التباين الأحادي ذو القياسات المتكررة للمتوسطات الحسابية لتقديرات معلمة الصعوبة لفقرات نماذج الاختبار

مصدر التباين		مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة (F) المحسوبة	الاحتمال
بين الفقرات		1.295	40	0.032	–	–
داخل الفقرات	بين النماذج	0.044	2	0.022	22	0.000
	الخطأ	0.046	80	0.001		
الكلي		1.385	122	–	–	–

يلاحظ من الجدول (21) وجود فروق ذات دلالة إحصائية بين المتوسطات الحسابية لتقديرات معلمة الصعوبة تعزى للطريقة المتبعة في اختيار مموهات فقرات الاختبار، وللكشف عن تلك الفروق تم استخدام اختبار توكي، كما في الجدول (22).

جدول (22): نتائج اختبار توكي للكشف عن الفروق الثنائية للمتوسطات الحسابية لتقديرات معلمة الصعوبة لنماذج الاختبار

النموذج	A 0.55	B 0.51	C 0.51
A	–	*0.04	*0.04
B	–	–	0.0

*: ذات دلالة إحصائية عند مستوى الدلالة (α=0.05).

يلاحظ من الجدول (22) أن هناك فروقًا ذات دلالة إحصائية بين تقديرات معلمة الصعوبة للنموذج (A) والنموذج (B)، وقد كانت تقديرات معلمة الصعوبة لفقرات النموذج (B) أعلى، كما أن هناك فروقًا ذات دلالة إحصائية بين تقديرات معلمة الصعوبة للنموذج (A) والنموذج (C)، وقد كانت تقديرات معلمة الصعوبة لفقرات النموذج (C) أعلى، وكذلك يلاحظ عدم وجود فروق ذات دلالة إحصائية بين تقديرات معلمة الصعوبة للنموذج (B) والنموذج (C).

معلمة التمييز

تم حساب المتوسطات الحسابية والانحرافات المعيارية لمعاملات التمييز لنماذج الاختبار الثلاثة، كما في الجدول (23).

جدول (23): المتوسطات الحسابية والانحرافات المعيارية لمعاملات التمييز لنماذج الاختبار الثلاثة

النموذج	أدنى قيمة	أعلى قيمة	الوسط الحسابي	الانحراف المعياري
A	0.31	0.53	0.42	0.50
B	0.31	0.57	0.44	0.74
C	0.31	0.59	0.45	0.70

وتم استخدام تحليل التباين الأحادي ذي القياسات المتكررة؛ لاستخراج دالة الفروق على معلمة التمييز، ويبين الجدول (24) نتائج التحليل.

جدول (24): تحليل التباين الأحادي ذو القياسات المتكررة للمتوسطات الحسابية لتقديرات معلمة التمييز لفقرات نماذج الاختبار

مصدر التباين		مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة (F) المحسوبة	الاحتمال
بين الفقرات		0.453	40	0.011	–	–
داخل الفقرات	بين النماذج	0.009	2	0.005	5	0.002
	الخطأ	0.054	80	0.001		
الكلي		0.516	122	–	–	–

يلاحظ من الجدول (24) وجود فروق ذات دلالة إحصائية بين المتوسطات الحسابية لتقديرات معلمة التمييز تعزى للطريقة المتبعة في اختيار مموهات فقرات الاختبار، وللكشف عن تلك الفروق، تم استخدام اختبار توكي، كما في الجدول (25).

جدول (25): نتائج اختبار توكي للكشف عن الفروق الثنائية للمتوسطات الحسابية لتقديرات معلمة التمييز لنماذج الاختبار

النموذج	A 0.42	B 0.44	C 0.45
A	–	-0.02	*-0.03
B	–	–	*-0.01

*: ذات دلالة إحصائية عند مستوى الدلالة (∝=0.05).

يلاحظ من الجدول (25) أن هناك فروقًا ذات دلالة إحصائية بين تقديرات معلمة التمييز للنموذجين (A) و(C)، لصالح النموذج (C)،

وقد كانت تقديرات معلمة التمييز لفقرات النموذج (C) أعلى، كما أن هناك فروقًا ذات دلالة إحصائية بين تقديرات معلمة التمييز للنموذجين (B) و(C)، وقد كانت تقديرات معلمة الصعوبة لفقرات النموذج (C) أعلى، وكذلك عدم وجود فروق ذات دلالة إحصائية بين تقديرات معلمة التمييز للنموذجين (A) و(B).

مناقشة النتائج

مناقشة النتائج المتعلقة بفرضية الدراسة الأولى:

من فحص النتائج المتعلقة بتدريج قدرات المفحوصين على متصل القدرة على النماذج الثلاثة تبين وجود علاقة قوية بين تلك التدريجات في جميع الحالات، وعند فحص العلاقة الارتباطية أظهر التحليل عدم وجود فروق ذات دلالة إحصائية بين معاملات ارتباط سبيرمان لقيم تدريج المفحوصين على متصل القدرة، ويعزى ذلك إلى أن مموهات الفقرات تم بناؤها بالاعتماد على طرق جيدة في اختيار المموهات، بحيث تكون مموهات الفقرات أكثر تجانسًا، مما ساهم في زيادة قوة العلاقة الارتباطية بين قيم تدريجات المفحوصين على متصل القدرة. ويمكن القول إنه لا يوجد تأثير لطريقة اختيار المموهات لفقرات الاختبارات على مواقع المفحوصين على متصل القدرة.

أما الأخطاء المعيارية لتقديرات معلمة القدرة، فقد كانت مرتفعة عند إجابة المفحوصين عن نموذج الاختبار (C)، وهذه النتيجة منطقية كون الطالب قد يفكر في عدة جوانب لإجابة الفقرة، مما يزيد من قيم الأخطاء المعيارية لتقديرات معلمة القدرة.

وأشارت النتائج المتعلقة بتقديرات قدرات الأفراد المقدرة بالعلامة الخام وفق النظرية الكلاسيكية إلى وجود فروق ذات دلالة إحصائية لصالح النموذج (A)، ولعل سبب ارتفاع المتوسط الحسابي لعلامات الطلبة وفق هذا النموذج (A) يعزى إلى خبرة الطالب في أخذ الاختبارات من نوع الاختيار من متعدد من إعداد المعلم.

مناقشة النتائج المتعلقة بفرضية الدراسة الثانية:

- مناقشة النتائج الخاصة بمعلمة الصعوبة والأخطاء المعيارية في تقديرها:

نجد من النتائج أن فقرات النموذج (C) تقديرات عالية لمعاملات الصعوبة؛ أي أن فقراتها أصعب من فقرات النموذجين الآخرين (A,B). وعند فحص دلالة الفرق على معلمة الصعوبة، تبين وجود فروق ذات دلالة إحصائية بين المتوسطات الحسابية لتقديرات معلمة الصعوبة بين النموذجين (A,C)، لصالح النموذج (C)؛ أي أن تقديرات معلمة الصعوبة لفقراته أعلى، وربما قد يعود ذلك إلى أن إدخال الأخطاء التي كان يقع بها الطلبة ضمن مموهات الفقرات، يعمل على تضليل الطالب في تَعَرُّف الإجابة الصحيحة. ويمكن القول إن فقرات نموذج الاختبار (C) الذي تم بناء مموهاته من خلال إجابات الطلبة، ستزيد من مستوى صعوبتها من منظور نظرية استجابة الفقرة، وتتفق هذه النتيجة مع نتائج دراسة أبو جراد (1997) من منظور النظرية الكلاسيكية، وتتناقض هذه النتيجة مع نتائج دراسة Cizek and Oday (1994) ودراسة أمبوسعيدي (2005) من منظور النظرية الكلاسيكية اللتين أظهرتا عدم وجود فرق ذي دلالة إحصائية بين المتوسطات الحسابية لمعاملات الصعوبة.

أما الأخطاء المعيارية لتقديرات معلمة الصعوبة فقد كانت الأعلى في النموذج (B)، مقارنة بالنموذجين (A,B)، وهذا يدل على أن فقرات هذا النموذج أكثر دقة في قياس تحصيل الطلبة، علاوة على أن متوسط الأخطاء المعيارية لتقديرات الصعوبة القليل يشير إلى دقة تقدير معلمة الصعوبة، وعند فحص دلالة الفروق في الأخطاء المعيارية لتقديرات معلمة الصعوبة تبين عدم وجود فروق ذات دلالة إحصائية تعزى لأثر طريقة اختيار المموهات على التقديرات الخاصة بالأخطاء المعيارية لتقديرات معلمة الصعوبة.

- مناقشة النتائج المتعلقة بمعلمة التمييز والأخطاء المعيارية في تقديرها:

بينت النتائج أن تقديرات معاملات التمييز لفقرات النموذج (C) كانت الأعلى؛ أي أن نوعية وجودة الفقرات مناسبة، وبالتالي تميز بشكل كبير بين المفحوصين، وعند فحص دلالة الفرق على معلمة التمييز، تبين وجود فروق ذات دلالة إحصائية بين المتوسطات الحسابية لتقديرات معلمة التمييز للنموذجين (A,C)، لصالح النموذج (C)؛

أي أن تقديرات معلمة التمييز لفقرات النموذج (C) أعلى، ولم يكن هناك فروق ذات دلالة إحصائية بين المتوسطات الحسابية للنموذجين (A,B)، وللنموذجين (B,C) تعزى لطريقة اختيار المموهات، وتدل هذه النتائج على أن طريقة اختيار مموهات فقرات الاختيار من متعدد تؤثر في التقديرات الخاصة بمعلمة التمييز للفقرة، وهذا ما يتناقض مع دراسة (Cizek and Oday, 1994) ودراسة أمبوسعيدي (2005).

أما عن الأخطاء المعيارية لتقديرات معلمة التمييز، فقد كان المتوسط الحسابي للأخطاء المعيارية لفقرات النموذج (C) الأعلى مقارنة بالنموذجين الآخرين، وكان أقل متوسط للأخطاء المعيارية لتقديرات معلمة التمييز لفقرات النموذج (A)، وهذا يعني أن عملية اختيار مموهات فقرات اختبار الاختيار من متعدد المبنية على الأسس المنطقية، وعلى الخبرة، تعمل على تقليل الأخطاء المعيارية لتقديرات معلمة التمييز.

- مناقشة النتائج المتعلقة بمعلمة التخمين والأخطاء المعيارية في تقديرها:

لقد كانت تقديرات معلمة التخمين في النماذج الثلاثة متقاربة إلى حد كبير، ويمكن عزو ذلك إلى الدقة في بناء المموهات في الفقرات في النماذج الثلاثة، وإلى تطبيق نماذج الاختبار الثلاثة في نفس الجلسة، وعلى نفس المفحوصين، وأن المفحوص الذي يلجأ إلى التخمين على فقرة ما في أحد النماذج، فإنه يلجأ أيضًا إلى التخمين على نفس الفقرة في النموذجين الآخرين، لأن الفقرات لها نفس المتن، ولكن بمموهات مختلفة، وعند فحص دلالة الفروق على معلمة التخمين، تبين عدم وجود فروق ذات دلالة إحصائية بين المتوسطات الحسابية لتقديرات معلمة التخمين تعزى لطريقة اختيار المموهات، ويعزى ذلك إلى تقارب تقديرات معلمة التخمين في النماذج الثلاثة.

أما الأخطاء المعيارية لتقديرات معلمة التخمين، فكانت الأعلى في نموذج الاختبار (c) مقارنة بالنموذجين الآخرين، وكانت أقل ما يمكن لفقرات نموذج الاختبار (B)، وعند فحص دلالة الفروق على الأخطاء المعيارية، تبين وجود فروق ذات دلالة إحصائية بين النموذجين (A,C)، لصالح النموذج (A)؛ أي أن الأخطاء المعيارية لمعلمة التخمين في النموذج (A) أعلى، وكذلك تبين وجود فروق ذات دلالة إحصائية بين النموذجين (B,C) لصالح النموذج (B)، ولم تكن هناك فروق بين النموذجين (A,B)، ويفسر ذلك بالاعتماد على التقارب بين الأخطاء المعيارية لتقديرات معلمة التخمين لفقرات النموذجين (A,B)، وهذا يدل على أن هناك تأثيرًا لطريقة اختيار المموهات على التقديرات الخاصة بالأخطاء المعيارية لتقديرات معلمة التخمين لفقرات نماذج الاختبار الثلاثة.

مناقشة النتائج المتعلقة بفرضية الدراسة الثالثة:

معلمة الصعوبة: من النتائج وجد أن فقرات النموذج (C) كانت الأصعب مقارنة بالنموذجين الآخرين، ويعزى ذلك إلى أن المموهات التي تم بناؤها لفقرات النموذج (C) كانت من خلال إجابات الطلبة، ويمكن أن يقع الطلبة في نفس الأخطاء أثناء إجابتهم عن فقرات الاختبار، وعند فحص دلالة الفروق على معلمة الصعوبة، تبين وجود فروق ذات دلالة إحصائية بين المتوسطات الحسابية لتقديرات معلمة الصعوبة، بين النموذجين (A,B) لصالح النموذج (B)؛ أي أن النموذج (B) يعطي تقديرات أعلى لمعاملات الصعوبة؛ أي أن فقراته أسهل، ويعود السبب في وجود الفرق الدال إلى الاختلاف في قيم معاملات صعوبة الفقرات في هذا النموذج، وأظهرت النتائج عدم وجود فروق بين النموذجين (B,C)، ويعزى ذلك إلى أن إجابات المفحوصين عن الفقرات المتناظرة في النموذجين متقاربة إلى حد كبير.

معلمة التمييز: بينت النتائج أن القدرة التمييزية لفقرات النموذج (C) كانت الأعلى، وكذلك كانت فقرات النموذج (C) هي الأصعب مقارنة بالنموذجين الآخرين؛ أي أن مموهات تلك الفقرات كانت أكثر فاعلية، وساهمت بشكل كبير في إبراز التباين بين المفحوصين، وعند فحص دالة الفروق على معلمة التمييز، تبين وجود فروق ذات دلالة إحصائية بين المتوسطات الحسابية لتقديرات معلمة التمييز، بين النموذجين (A,C) لصالح النموذج (C)، وبين النموذجين (B,C) لصالح النموذج (C)؛ أي أن النموذج (C) يعطي تقديرات عالية لمعامل التمييز، وهذا يدل على أن فقراته مناسبة للتمييز بشكل جيد بين الطلبة

المفحوصين، وتعزى الفروق الدالة إحصائيًّا بين معاملات التمييز لنماذج الاختبار الثلاثة إلى اختلاف قيم صعوبتها، وإلى اختلاف توزيع قيم معاملات التمييز بين النماذج.

الخلاصة والتوصيات

نلاحظ مما سبق أن تقديرات قدرات المفحوصين على متصل القدرة على النماذج الثلاثة متقاربة، ولا يوجد فروق ذات دلالة إحصائية بين معاملات الارتباط لمواقع تدريج المفحوصين على متصل القدرة. أما قدرات المفحوصين المقدرة بالعلامة الخام، فقد كان هناك فروق ذات دلالة إحصائية بين النماذج تعزى لطريقة اختيار المموهات، وكان أعلى التقديرات في النموذج (C)، كما أن هناك تباينًا في خصائص الفقرات بين النماذج الثلاثة. وفيما يتعلق بمعلمة الصعوبة، ومعلمة التمييز فقد كان هناك فروق ذات دلالة إحصائية تعزى لطريقة اختيار المموهات بين النماذج الثلاثة، وفق النظرية الكلاسيكية والنظرية الحديثة في القياس، وكان أعلى التقديرات للمعلمتين في النموذج (C). أما معلمة التخمين فلم يكن هناك فروق ذات دلالة إحصائية بين النماذج الثلاثة تعزى لطريقة اختيار المموهات.

أما الأخطاء المعيارية لمعالم الفقرات (الصعوبة، والتمييز، والتخمين)، فقد كان هناك تباين في الأخطاء المعيارية لمعلمة التمييز، ومعلمة التخمين تعزى لطريقة اختيار المموهات. أما فيما يتعلق بالأخطاء المعيارية لمعلمة الصعوبة فلم يكن هناك فروق ذات دلالة إحصائية تعزى لطريقة اختيار المموهات؛ حيث كان أعلى التقديرات للمعلمتين في النموذج (C). ونظرًا للتباين في خصائص الفقرات وقدرات المفحوصين، بين النماذج الثلاثة للاختبار، فإن الباحثين يوصيان باستخدام هذه الإستراتيجيات مع مواضيع علمية أخرى، ومراحل تعليمية مختلفة.

المراجع

أبو جراد، حمدي يونس. 1997م. أثر الحذف التجريبي والعشوائي للمموهات في فقرات اختبار التحصيل من نوع الاختيار من متعدد على خصائصه السيكومترية. رسالة ماجستير غير منشورة، كلية التربية والفنون، جامعة اليرموك، إربد، الأردن.

أبو فودة، باسل خميس. 2008. أثر تعلم حكمة الاختبار على افتراض أحادية البعد والتقديرات المختلفة لنظرية استجابة الفقرة. رسالة دكتوراه غير منشورة، كلية التربية، جامعة اليرموك، إربد، الأردن.

أمبوسعيدي، عبدالله بن خميس. 2005. أثر تغيير موقع المشتت القوي (Strong Distractor) عند استخدام تقنية إبعاد البدائل الخطأ في صعوبة الفقرات لأسئلة الاختيار من متعدد في الفيزياء وتمييزها. مجلة العلوم التربوية والنفسية، كلية التربية، جامعة البحرين، البحرين، مج (6)، ع (3)، ص ص 85-105.

عودة، أحمد سليمان. 2004. القياس والتقويم في العملية التدريسية. بدون رقم الطبعة، دار الأمل، إربد، الأردن.

يعقوب، إبراهيم محمد عيسى. 2000. أثر اختلاف عدد البدائل والبديل «لا شيء مما ذكر» على الخصائص السيكومترية لفقرات الاختيار من متعدد. مجلة الأستاذ، كلية التربية، ابن رشد للعلوم الإنسانية، جامعة بغداد، بغداد، العراق، ع (17)، ص ص 152-178.

Baghaei, P., and Amrahi, N. 2011. The effects of number of options on the psychometric characteristic of multiple choice. Psychological Test and Assessment Modeling. 53(2): 192-211.

Budescu, D., and Nevo, B. 1985. Optimal number of options: an investigation of the assumption of proportionality. Journal of Educational Measurement. 22(3): 183-196.

Choi, I., and Bachman, F. 1992. An investigation into the adequacy of three IRT models for data from two EFL reading tests. Language Testing. 9: 51 – 78.

Cizek, G. J., and Oday, D. M. 1994. Further investigation of nonfunctioning options in multiple-choice test item. Educational and Psychological Measurement. 54(4): 861-872.

Crehan, K. D., and Haladna, T. M. 1991. The validity of two item writing rules. Journal of Experimental Education. 59(2): 183-192.

Ebel, R. L., and Frisbie, D. A. 1986. Essentials of Educational Measurement. (Fourth edition). Prentice-Hall, New Jersey.

Gay, L. 1980. Educational Evaluation and Measurement. E. Merril, Ohio.

Haladyna, T., Downing, S., and Rodriguez, M. 2002. A Review of multiple choice item writing guidelines for classroom assessment. Applied Measurement in Education. 15(3): 309 - 334.

Hamblton, R., and Swaminathan, H. 1985. Item Response Theory: Principles and Applications. Nijhoff publishing, Boston.

Hambleton, R., Swaminathan, H., and Rogers, H. 1991. Fundamentals of Item Response Theory. sage publication, Thousand Oaks, CA.

Landrum, R. E., Cashin, J. R., and Thesis, K. S. 1993. More evidence in favor of three-option multiple-choice tests. Educational and Psychological Measurement. 53: 771-778.

Levy, K. 1978. A procedure for testing the tquality of K dependent correlation coefficients. Journal of Statistical Computation and Simulation. 7: 181-187.

Murphy, K. R., and Hofer, C. O. 1994. Psychological Testing Principles. Third edition, Prentice-Hall, New Jersey.

Pelton, T. 2002. The accuracy of unidimensional measurement model in the presence of devitions from the underlying assumptions. Unpublished doctorel dissertation, Bringham University, USA.

Rodriguez, M. C. 2005. Three options are optimal for multiple-choice items: A metaanalysis of 80 years of research. Educational Measurement: Issues and Practice. 24(2): 3-13.

Shizuka, T., Takeuchi, O., Yashima, T., and Yoshizawa, K. 2006. A comparison of three- and four-option English tests for university entrance selection purposes in Japan. LanguageTesting. 23(1): 35-57.

Straton, R. G., and Catts, R. M. 1980. A comparison of two, three and four choices item tests given a fixed total number of choices. Educational and Psychological Measurement. 40: 357-365.

The Effect of Distracters Selection Method in Multiple Choice Tests Items on their Psychometric Properties Based on Classical Measurement and Modern Measurement Theories

Nayel Odeh Al- Kaabnh[1] and Ahmad Sulaiman Oude[2]

(1) Huraimla Community College, Shaqra University, Shaqra, Kingdom of Saudi Arabia
(2) Department of Educational Psychology, Faculty of Education,Yarmouk University, Irbed, Jordan

ABASTRACT

The objective of this study was to investigate the effect of distracters selection method in multiple choice test items on their psychometric properties (difficulty, discrimination, and guessing) and ability parameters estimation based on classical and modern measurement theories. To achieve the objective of the study, the researchers prepared an achievement test for 9 th graders consisting in its final format of 41 items. Three forms of items distracters (A, B, C) were used. SPSS and Bilog-MG programmes were used to estimate statistical indicators.

Findings revealed no effect for selection method of multiple test item distracters on examinees ranks on continuum ability scaling. Significant differences were found for standard error estimations of ability parameters estimation means. Significant differences were found between examinees abilities estimated using raw score due to the method used in item distractors construction. Significant differences were found between difficulty parameter estimation and discrimination parameter estimation. No significant differences were found in standard errors estimation for difficulty parameter. No significant differences were found in guessing estimation parameter. Significant differences were found between standard error for guessing parameter estimation. Significant differences between difficulty coefficients and discrimination coefficients values were found according to the Classical Theory.

Key Words: Classical measurement theory, Distractors' selection strategy, Modern measurement theory.