

أثر اختلاف قدرة الأفراد على دقة تقدير معالم المُفردات والأفراد في أسئلة الاختيار من متعدد وفق النموذج اللوجستي الثلاثي المعلم

خضراء عبد الله البلوي*

الملخص: هدفت هذه الدراسة إلى التحقق من أثر اختلاف قدرة الأفراد على دقة تقدير معالم المُفردات والأفراد، في أسئلة الاختيار من متعدد، وفق النموذج اللوجستي الثلاثي المعلم. ولتحقيق الغرض من الدراسة؛ تم توليد استجابات ثلاثة أشكال من القدرة، باستخدام برنامج WinGen3 على (60) مُفردة من نوع الاختيار من متعدد، ثنائياً الاستجابية، وفق النموذج اللوجستي ثلاثي المعلم، وبلغ عدد أفراد العينة (3000) فرد، بواقع (1000) فرد لكل شكل من أشكال القدرة، وتم تحليل البيانات المولدة باستخدام البرمجيات Spss، Xcalibre. وبعد إجراء التحليل الإحصائي المناسب المتمثل باستخدام اختبار تحليل التباين الأحادي، وحساب حجم الأثر، ودقة التقدير، توصلت الدراسة إلى وجود فروق ذات دلالة إحصائية عند مستوى دلالة (0.05) في دقة تقدير معالم المُفردات والأفراد، تُعزى لاختلاف شكل قدرة الأفراد (مرتفعة، متوسطة، منخفضة)، حيث كانت تقديرات الصعوبة والتخمين والقدرة أكثر دقة لصالح القدرة المتوسطة، وتقديرات التمييز أكثر دقة لصالح القدرة المنخفضة، كما أشارت النتائج إلى تباين كمية المعلومات التي قدمتها كل مُفردة باختلاف قدرة الأفراد، وكانت النتائج لصالح القدرة المنخفضة؛ إذ إنها قد زوّدت المُفردات بأكبر كم من المعلومات.

الكلمات المفتاحية: المحاكاة، قدرة الأفراد، النموذج اللوجستي الثلاثي المعلم، الاختيار من متعدد، معالم المفردات.

أثر اختلاف قدرة الأفراد على دقة تقدير معالم المُفردات والأفراد في أسئلة

الاختيار من متعدد وفق النموذج اللوجستي الثلاثي المُعلم

1. المقدمة

تُعَدُّ الاختبارات التحصيلية أحد أهم أساليب قياس التحصيل الدراسي، والأكثر انتشاراً في المجال التربوي، وتهدف لقياس وتقويم أداء الطلاب، ومستوى تعلمهم؛ فهي تعمل على إعطاء قياسات مناسبة للكثير من المخرجات التربوية.

وللاختبارات التحصيلية التي يتم تطبيقها عدّة صور، منها الاختبارات الشفهية، والاختبارات الأدائية، واختبارات الورقة والقلم، والنوع الأخير هو السائد في المؤسسات التعليمية عموماً عاملاً [1].

وتنقسم اختبارات الورقة والقلم إلى نوعين؛ فإما أن تكون أسئلةً مقاليتية، وإما أن تكون أسئلةً موضوعية، ويعتمد الاختيار على الهدف الذي وُضع الاختبار من أجله، وكذا ظروف الأفراد الذين سيُجرى لهم الاختبار. وتعدُّ أسئلة الاختيار من متعدد من أفضل أنواع الفقرات الموضوعية، والأكثر انتشاراً، وهي المفضّلة لدى المُعلّمين؛ لأنها تتميز بسهولة التصحيح، والموضوعية، وتوفّر تغطية جيّدة للمادّة الدراسية، وتحديد نتائج التعلم بدرجة عالية، على الرغم من الجهد الكبير والوقت الطويل الذي يحتاجه المُعلّم في إعدادها، وما تتطلبه من مهارة من قبل واضعها.

ومن صيغ اختبارات الاختيار من متعدد: أن يكون لكلِّ فقرة بديلان أو ثلاثة بدائل، أو أربعة بدائل، أو خمسة بدائل، وهذه البدائل تحوي إجابة واحدة صحيحة، وبقية البدائل تكاد تُشبه هذه الإجابة الصحيحة، والهدف منها إبعاد تفكير الطلبة الذين لا يعرفون الإجابة الصحيحة عنها، وتبين مدى نقص المعلومات اللازمة للإجابة الصحيحة، ومن أهميتها أنها تبين مواطن الضعف، والأخطاء الشائعة لدى الطلبة؛ ليتسنى معالجة هذه الأخطاء لاحقاً الطبري [2].

وعلى الرغم من امتلاك فقرات الاختيار من متعدد لبعض الخصائص، فإن هناك ما يحددها، وهو العدد الكافي من البدائل، بحيث تكون مقبولة للطلبة، وتناسب المرحلة التعليمية والمعلومات التي يعرفونها، وقد توصّلت بعبارة [3] إلى أن عدد البدائل يتناسب طردياً مع مستوى الطلبة الدراسي، حيث المرحلة الابتدائية تناسبها البدائل الثلاثة، والمرحلة المتوسطة تناسبها البدائل الأربعة، بينما المرحلة الثانوية تصلح لها البدائل الخمسة.

ومن جانب آخر، يُعنى علماء القياس النفسي والتربوي بموضوعية القياس، أو ما يسمّى باستقلالية القياس، ويعني أن تقدير معالم المُفردات - أي: صعوبتها وتميزها - يكون مستقلاً عن خصائص الأفراد الذين تمّ استخدامها في تقدير هذه المعالم، وأن تقدير قدرة الفرد يكون مستقلاً عن عينة المُفردات التي تطبّق عليه رينولدز وليفنجستون [4].

والجدير بالذكر أن النظرية التقليدية في القياس قد خدّمت المتخصّصين في القياس النفسي والتربوي بشكل جيّد لفترة طويلة؛ ولكن هذه النظرية عانت من بعض جوانب الضعف، ومنها: اعتماد معالم مُفردات الاختبار؛ كالصعوبة والتميز والتخمين، على عينة الأفراد الذين يؤدّون الاختبار، والتي تؤثر بدورها على مستويات القدرة، كما أن درجات الأفراد في الاختبار تعتمد على عينة المُفردات التي اشتمل عليها الاختبار

الطبري [2].

لذا: اعتمد الاتجاه المعاصر في القياس النفسي والتربوي، على نظرية الاستجابة للمُفردة، أو نظرية السمات الكامنة، التي تهدف إلى تقدير جميع إحصاءات المُفردة والفرد، وتفترض أنه يمكن التنبؤ بأداء الأفراد، أو تفسير أدائهم في اختبار نفسي أو تربوي معيّن، بناءً على خاصية معيّنة مميزة لهذا الأداء، تسمّى السمات، ولصعوبة ملاحظة هذه السمات بشكل مباشر؛ فإننا نستدلُّ عليها من خلال استجابات الأفراد الملاحظة على فقرات الاختبار، الذي يقيس السمة أو القدرة المطلوبة، وهذا هو سبب تسمية هذه النظرية بنظرية السمات الكامنة [5].

وحقّ يتحقّق القياس الموضوعي؛ فقد اهتمت نظرية السمات الكامنة بتطوير النماذج الاحتمالية؛ لتحديد العلاقة بين أداء الفرد على مُفردات الاختبار، وبين القدرات الكامنة وراء هذا الأداء، ويمكن تصنيف هذه النماذج إلى نوعين: الأول: نماذج متعدّدة التدرج (Polytomous Models)، والثاني: نماذج ثنائية التدرج (Dichotomous Models)، التي تكون استجابتها إما صحيحةً أو خاطئةً.

وتعدُّ النماذج ثنائية التدرج من أشهر النماذج استخداماً في بناء الاختبارات العقلية والتحصيلية، وتوجد في الوقت الحالي ثلاثة نماذج من النوع ثنائي التدرج، تختلف في عدد بارامترات أو معالمها، ويمثّل النموذج ثلاثي المُعلم الشكل العامّ للنماذج اللوجستية؛ لأنه يحتوي على المعالم الثلاثة الممكنة للمُفردة، وهي: (الصعوبة، والتميز، والتخمين)، وهو النموذج المناسب عند تحليل بيانات الفقرات الموضوعية، خاصةً من نوع الاختيار من متعدد عاملاً [1].

وقد أشار عاملاً [6] إلى أن نظرية الاستجابة للمُفردة لا تشترط التوزيع الطبيعي للسمة، أو القدرة المقاسة على غرار النظرية التقليدية، ويتّسجم هذا مع بعض المواقف الاختبارية التي يتمّ فيها التعامل مع عينات غير عشوائية، فقد يتمّ التعامل مع أفراد ذوي قدرة مرتفعة أو منخفضة؛ ممّا يعني توزيعاً غير طبيعيّ قد يحدث، ومن هنا يمكن طرح أسئلة كثيرة، منها: ما مدى تأثير مُفردات الاختبار وقدرات الأفراد بشكل توزيع القدرة؟

ونظراً لأهمية الاختبارات في الكشف عن القدرات الحقيقية للأفراد، وخاصةً في المجال التعليمي، والحاجة لإيجاد مقاييس تتمتع بنوع من الثبات والدقة، وانطلاقاً من توجّه الدقة في عملية القياس للحيلولة دون اتّخاذ قرارات خاطئة على مستوى الفرد والمجتمع، وظهور تساؤلات، منها: ما مدى تأثير مُفردات الاختبار وقدرات الأفراد بشكل توزيع القدرة؟ كونها أحد المتغيرات التي قد تؤثر على تقديرات معالم المُفردات وقدرة الأفراد - برزت الحاجة لإجراء هذه الدراسة؛ لتقصّي أثر اختلاف قدرة الأفراد على دقة تقدير معالم المُفردات والأفراد في أسئلة الاختيار من متعدد، وفق النموذج ثلاثي المُعلم.

2. مشكلة الدراسة

من القضايا المهمة في نظرية الاستجابة للمفردة دقة تقدير معالم المفردات والأفراد وخاصة حينما يتعلق الأمر بالاختبارات التحصيلية، وقد

الجانب النظري؛
أهميّة التحقّق من دقّة إجراءات تحليل بيانات أسئلة الاختيار من متعدّد؛
كوّنها تقيس جميع المستويات العقلية للطالب، وتسعى هذه الدراسة إلى
إيجاد المدى الأمثل للقدرة التي تناسب تحليل هذا النوع من الأسئلة (الاختيار
من متعدّد).

دعم القاعدة النظرية للبحوث المهتمّة بنماذج نظرية الاستجابة للمُفردة في
مجال تحليل أسئلة الاختيار من متعدّد.

وتتمثّل أهميّة الجانب العملي في:

أنه قد تفيد نتائج الدراسة متخصصي القياس والتقويم التربوي، في الكشف
عن مدى تأثير قدرات الأفراد في دقّة تقدير معالم المُفردات، وفقّ التّمودج
الثلاثي المُعلّم، عند تحليل أسئلة الاختيار من متعدّد.

تطبيق التّمودج ثلاثي المُعلّم في تقدير معالم المُفردات من صعوبة، وتمييز،
وتخمين؛ وفي تقدير قدرة الأفراد، في ظلّ ثلاث عيّنات متحرّرة في القدرة
(مرتفعة، متوسّطة، منخفضة).

د. حدود الدراسة

تمّ استخدام المحاكاة بطريقة المونتي كارلو (monte Carlo) باستخدام
برنامج (WINGEN3-3) لتوليد البيانات، المستخدم في علم القياس والتقويم،
حيث إنه منذ ظهور نظرية الاستجابة للمُفردة، تطوّرت برامج التوليد
المتعلّقة بها، واعتمدت البحوث العلمية على أسلوب التوليد، وكانت دراسة
المونتي كارلو من أكثرها شيوعاً؛ وذلك بهدف فحص افتراضات النظرية،
ومعادلة درجات الاختبار، وغيرها، وذكر Harwell; Stone; Hsu & Kirisci [7]
مدى فائدة التوليد بطريقة المونتي كارلو حيث يتّم التوصل من خلالها إلى
معلومات دقيقة حول كلّ نمودج انبثق عن هذه النظرية، وحلّ كثيراً من
المشكلات المرتبطة بالقياس والإحصاء، وتميّز كذلك بقدرته على معالجة
المتغيّرات، ودراسة تأثيرها مع بعضها البعض، بالإضافة إلى أنها أقلّ تكلفةً
ماديّةً ومعنويّةً من الدراسات التي تعتمد على تطبيق الأدوات الدراسية على
الأفراد، ومن خلاله يتّم التوصل لنتائج يمكن تعميمها.
تمّ استخدام التّمودج ثلاثي البارامتر في تحليل البيانات، واستخراج مؤشّرات
حول معالم المُفردات والأفراد.

اقتصرت الدراسة على المقارنة بين فئات القدرة الثلاث: (مرتفعة، متوسّطة،
منخفضة).

ه. مصطلحات الدراسة

قدرة الأفراد: "مقدار ما يمتلكه الفرد من السمة التي يقيسها تبعاً لنظرية
الاستجابة للمُفردة" Hambleton & Swaminathan [5].

معالم المُفردات: المقصود بها "معالم الصعوبة والتمييز والتخمين، وفقّ
النّمودج ثلاثي المُعلّم" علام [6].

معلّم القدرة: "هي الخصائص المتعلّقة بالتوزيع التكراري للقدرة، من حيث
شكل التوزيع، ومتوسّطه الحسابي، وانحرافه المعياري، والتواءه، وتفلطحه"
علام [6].

النّمودج ثلاثي المُعلّم: "هو أحد نماذج نظرية الاستجابة للمُفردة ثنائيتة
التدرج، وهذا النّمودج يقدر ثلاثة معالم للمُفردة، وهي: معلّم الصعوبة،
ومعلّم التمييز، ومعلّم التخمين" علام [1].

معلّم الصعوبة: "هو عبارة عن نقطة على منّصل القدرة التي تقابل احتمال
(50٪) للإجابة على المُفردة إجابةً صحيحة" [8].

تباينت وجهات نظر الباحثين حول العوامل المؤثرة على دقة التقديرات،
فمنهم من يرى أن لحجم العينة وطول الاختبار أثر على دقة التقدير، وكذلك
تباينت الآراء في أثر انتهاك الافتراضات التي تستند عليها النظرية الحديثة وأثر
ذلك على التقديرات، ولكن هناك مواقف كثيرة قد تتطلب تقييم أفراد
يختلفون في مستوى القدرة، خصوصاً في الاختبارات المقننة والتكيفية، وقد
يؤثر ذلك على خصائص المفردات مما يستدعي دراسة أثر اختلاف القدرة
على معالم المفردات، وذلك تمشياً مع التوجه في تقييم مدى دقة وكفاءة
نظرية الاستجابة للمفردة في تقدير معالم المفردات ودقة القياس.

أ. أسئلة الدراسة

تحاول الدراسة الإجابة عن السؤال الرئيس التالي:

ما أثر اختلاف مستوى قدرة الأفراد على دقّة تقدير معالم المُفردات والأفراد
في أسئلة الاختيار من متعدّد؟

ويتفرّع من هذا السؤال الأسئلة التالية:

هل توجد فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha \leq 0.05$) في دقّة
تقدير معلّم الصعوبة، وفقّ النّمودج ثلاثي المُعلّم، تُعزى لاختلاف قدرة
الأفراد (مرتفعة، متوسّطة، منخفضة)؟

هل توجد فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha \leq 0.05$) في دقّة
تقدير معلّم التمييز، وفقّ النّمودج ثلاثي المُعلّم، تُعزى لاختلاف قدرة الأفراد
(مرتفعة، متوسّطة، منخفضة)؟

هل توجد فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha \leq 0.05$) في دقّة
تقدير معلّم التخمين، وفقّ النّمودج ثلاثي المُعلّم، تُعزى لاختلاف قدرة الأفراد
(مرتفعة، متوسّطة، منخفضة)؟

هل توجد فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha \leq 0.05$) في دقّة
تقدير قدرة الأفراد، وفقّ النّمودج ثلاثي المُعلّم، تُعزى لاختلاف قدرة الأفراد
(مرتفعة، متوسّطة، منخفضة)؟

هل توجد فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha \leq 0.05$) في كميّة
المعلومات التي تقدّمها كلّ مُفردة، وفقّ النّمودج ثلاثي المُعلّم، تبعاً لاختلاف
قدرة الأفراد (مرتفعة، متوسّطة، منخفضة)؟

ب. أهداف الدراسة

تهدف الدراسة إلى ما يلي:

الكشف عن أثر اختلاف قدرة الأفراد على دقّة تقدير معلّم الصعوبة، وفقّ
النّمودج ثلاثي المُعلّم.

الكشف عن أثر اختلاف قدرة الأفراد على دقّة تقدير معلّم التمييز، وفقّ
النّمودج ثلاثي المُعلّم.

الكشف عن أثر اختلاف قدرة الأفراد على دقّة تقدير معلّم التخمين، وفقّ
النّمودج ثلاثي المُعلّم.

الكشف عن أثر اختلاف قدرة الأفراد على دقّة تقدير القدرة، وفقّ النّمودج
ثلاثي المُعلّم.

الكشف عن أثر اختلاف قدرة الأفراد على كميّة المعلومات التي تقدّمها كلّ
مُفردة، وفقّ النّمودج ثلاثي المُعلّم.

ج. أهمية الدراسة

إن أهميّة أيّ دراسة تنبثق من مقدار ما تُضيفه إلى المعرفة من الناحيتين:
النظرية والعملية، لذا؛ فإن أهميّة هذه الدراسة تظهر فيما يلي:

أثر اختلاف قدرة الأفراد على دقة تقدير معالم المُفَرَّدات والأفراد في أسئلة الاختيار من متعدد

خضراء البلوي

أما الشرفين [14] فقد أجرى دراسة هدفت إلى الكشف عن أثر طريقة تقدير معالم الفقرة، وقدرات الأفراد، على قيم معالم الفقرة، والخصائص السيكو مترية للاختبار، في ضوء اختلاف حجم العينة. ولتحقيق هدف الدراسة؛ تمّ بناء اختبار تحصيلي في الفيزياء تكوّن من (33) فقرة من نوع الاختيار من أربعة بدائل، وطبّق على عيّنة مكوّنة من (1000) طالب وطالبة، وحلّلت النتائج وفق النّمودج اللوجستيّ ثلاثي المَعْلَم باستخدام برمجية Bilogmg، وأشارت النتائج إلى أن دقة تقديرات مَعْلَم القدرة تزداد في حالة عيّنة الأفراد ذوي القدرة العالية، وعيّنة الأفراد ذوي القدرة المتدنيّة، عند استخدام طريقة بيز التوقّع (EAP). في حين تزداد الدقة عند مستويات الأفراد ذوي القدرة المتوسّطة، باستخدام طريقة الأرجحية العظمى (MLE) بغضّ النّظر عن حجم العينة.

كما أجرى بني عطا [15] دراسة هدفت إلى التحقّق من أثر اختلاف شكل توزيع القدرة على معالم الفقرة، ودالة معلومات الاختبار، تحت شروط اختبار بطول (60) مُفَرَّدة، وأربعة أشكال من توزيعات القدرة (طبيعيّ، ملتبس نحو اليسار، ملتبس نحو اليمين، ثنائي المنوال)، تمّ توليدها وفق النّمودج اللوجستيّ ثلاثي المَعْلَم باستخدام برنامج WinGen3، وباستخدام برنامج Bilogmg تمّ تحليل الاستجابات. وقد كشفت نتائج الدراسة عن وجود أثر لاختلاف شكل توزيع القدرة على معالم الفقرات، ودالة المعلومات للاختبار، حيث أنتج التوزيع ثنائي المنوال أعلى معاملات تمييز للمُفَرَّدات، وأعلى دالة معلومات، بينما قدّر التوزيع المتلوي نحو اليمين مَعْلَم الصعوبة بدقة، وقدّر التوزيع المتلوي نحو اليسار مَعْلَم التخمين بدقة، مقارنةً بأشكال التوزيع الأخرى.

ودراسة الصبح [16] التي هدفت إلى المقارنة بين دقة تقدير القدرة باختلاف طول الاختبار، وشكل توزيع مَعْلَم القدرة للبيانات المولّدة والحقيقيةّة. ولتحقيق هدف الدراسة؛ تمّ استخدام بيانات الاختبار الوطنيّ لمادّة العلوم للصف الثامن كمصدر للبيانات الحقيقيةّة، وتمّ توليد استجابات مماثلة لخصائص العينة الحقيقيةّة وفق النّمودج اللوجستيّ ثلاثي المَعْلَم، كما تمّ تحديد حجم العينة بألف مفحوص، وطول الاختبار (25-50) مُفَرَّدة، وتمّ تحليل البيانات باستخدام برنامج Bilogmg وتوصّلت الباحثة إلى عدم وجود فروق عمليّة في دقة تقدير مَعْلَم القدرة، وذلك باختلاف طول الاختبار، واختلاف شكل توزيع القدرة (ملتبس نحو اليسار، طبيعيّ، ملتبس نحو اليمين).

دراسة الحواري [17] وهدفت إلى الكشف عن أثر طول الاختبار، وشكل توزيع القدرة في تقديرات القدرة للأفراد، وتقديرات معالم الصعوبة للمُفَرَّدات، وفق النّمودج اللوجستيّ الأحاديّ المَعْلَم. ولتحقيق الهدف من الدراسة؛ استخدمت الباحثة ثلاثة أشكال من توزيعات القدرة (ملتبس التواءً سالبًا، طبيعيّ، ملتبس التواءً موجبًا)، وتمّ توليد بيانات (1000) مفحوص، بطول اختبار (30-60)، وحلّلت البيانات ببرنامج Bilogmg. وقد أشارت النتائج لوجود أثر لشكل توزيع القدرة وطول الاختبار، حيث كانت تقديرات قدرة الأفراد أكثر دقةً لصالح شكل التوزيع المتلوي السالب والموجب، ولطول الاختبار (30) مُفَرَّدة، كما أشارت النتائج إلى أن طول الاختبار (30) مُفَرَّدة، وشكل التوزيع المتلوي للقدرة قد زوّد بمعلوماتٍ أكثر مَعْلَم الصعوبة.

مَعْلَم التمييز: "هو نسبة ميل منحني خصائص الفقرة، عند نقطة الانقلاب التّقريريّ، بحيث كلّما كانت زاوية الانقلاب كبيرة، دلّ ذلك على تميّز عالٍ للمُفَرَّدة" [8].

مَعْلَم التخمين: "احتمال حدوث الاستجابة الصحيحة بمحض الصدفة، ويتحدّد بالخطّ المقارب الأدنى في المنحنى المميّز للمُفَرَّدة" أبو حطب، عثمان، صادق [9].

3. الدراسات السابقة

أجرى ساس وشسملت ووكر [10] دراسة كان الهدف منها قياس أثر حجم العينة، والتوزيع المتلوي للقدرة، وطريقة التقدير على معالم الفقرة، ومَعْلَم القدرة للفرد، وذلك وفق النّمودج اللوجستيّ الثلاثي المَعْلَم. ولتحقيق هدف الدراسة؛ استخدم الباحثون حجم عيّنة (500-1000) فرد، وشكلين من توزيع القدرة (ملتبس سالب وموجب، وطبيعيّ)، وقد بيّنت النتائج أن تقديرات معالم الفقرات أكثر دقةً في التوزيع الطبيعيّ للقدرة، وأقلّ دقةً في التوزيع المتلوي (سالب، موجب).

ودراسة عابنة [11]، التي هدفت إلى التعرف على أثر حجم العينة، وطريقة انتقائها، وعدد الفقرات، وطريقة انتقائها، على دقة تقدير معالم الفقرة، والقدرة لاختبار قدرة عقليّة باستخدام نظرية الاستجابة للفقرة. ولتحقيق أهداف الدراسة؛ تمّ إعداد اختبار قدرة عقليّة، تكوّن من (71) فقرة من نوع الاختيار من متعدّد، وطبّق الاختبار على عيّنة من طلاب وطالبات الصف السابع، بلغ عددهم (1000) طالب وطالبة، وتمّ تقدير معالم الأفراد والمُفَرَّدات باستخدام برنامج Bilogmg وفق النّمودج اللوجستيّ ثلاثي المَعْلَم. وأظهرت النتائج أن دقة تقدير مَعْلَم التمييز يزداد بزيادة تباين قدرات المفحوصين، ولصالح القدرة المتدنيّة، وتزداد الدقة في تقدير مَعْلَم الصعوبة والقدرة عندما يكون مدى قدرة المفحوصين متوافقًا مع مدى صعوبة الفقرات، ولصالح العينة العشوائية، كما تزداد الدقة في تقدير مَعْلَم التخمين عند استخدام عيّنة من ذوي القدرة المنخفضة في معايرة الفقرات.

ودراسة الرشيد [12]، التي كان الهدف منها فحص اللاتغير في معالم الفقرات باختلاف معالم القدرة، واللاتغير في معالم القدرة باختلاف معالم المُفَرَّدات، عند استخدام النّمودج اللوجستيّ ثلاثي المَعْلَم. ولتحقيق أغراض الدراسة؛ تمّ بناء اختبار تحصيلي في مادّة الفيزياء مكوّن من (60) فقرة اختيار من متعدّد، وبلغ حجم العينة (1693) طالبًا وطالبة، وتمّ تقدير معالم المُفَرَّدات والأفراد باستخدام برنامج Bilogmg، كذلك تمّ تقسيم العينة إلى أربع مجموعات حسب مَعْلَم القدرة (ذوي قدرات عشوائية، ذوي قدرات مرتفعة، ذوي قدرات متوسّطة، ذوي قدرات متدنيّة)، وكانت النتائج عدم تحقّق افتراض اللاتغير في تقدير معالم الفقرات باختلاف معالم القدرة.

وكشفت دراسة الشرفين والصبح [13] أن الأفراد ذوي القدرة المرتفعة أكثر دقةً في تقدير مَعْلَم القدرة، وأن ذوي القدرة المتدنيّة أكثر دقةً في تقدير مَعْلَم التمييز والتخمين، وأن الأفراد ذوي القدرة المرتفعة والمتوسّطة أكثر دقةً في تقدير مَعْلَم الصعوبة، وذلك في دراستهما التي هدفت للتحقّق من أثر بنية فقرات الاختيار من متعدّد (معكمة البنية، ومخالف لقواعد صياغة الأسئلة)، ومستوى القدرة لدى الأفراد على دقة التقديرات لمعالم الفقرات والأفراد وفق النّمودجين الثنائيّ والثلاثي المَعْلَم، وبلغت عيّنة الدراسة (1057) طالبًا وطالبة، وتمّ بناء اختبار تحصيلي في مادّة الرياضيات من نوع الاختيار من أربعة بدائل، وحلّلت النتائج باستخدام برنامج Bilogmg.

تشارك به مجموعة من الاختبارات وليس جميعها، والعامل الخاص الذي يميّز اختباراً عن مجموعة من الاختبارات فرج [20].

وقد حاول كاتل التوفيق بين نمودجّي ثرستون وسبيرمان، حيث افترض بنية هرمية للقدرات العقلية، حيث يوجد العامل العام في جميع مقاييس القدرة، بينما القدرات الأولية تنتمي إلى مستوى أدنى في تنظيمه الهرمي، كما توصل إلى قابلية العامل العام للقسم إلى عاملين، أطلق على أحدهما "الذكاء السائل"، وعلى الآخر "الذكاء المتبلور"، ويتحدّد الذكاء السائل باختبارات تقيس قدرة الفرد على اكتساب المعرفة؛ للاختبارات التي تقيس عاملي الاستدلال والاستقراء والقدرة المكانية، أما الذكاء المتبلور، فيتحدّد باختبارات التي تقيس آثار التدريس.

ولعب الاختلاف في القدرات العقلية دوراً مؤثراً في حياة الفرد؛ فقد نجد اختلافاً في القدرات العقلية عند الأفراد تؤثر على تحصيلهم العلمي، ومن أشهر وسائل قياس القدرات العقلية: الاختبارات؛ حيث تُعدّ أحد أهم أدوات القياس التي يستخدمها المشتغلون بالقياس والتقويم التربوي والنفسي؛ للتعرف على مقدار ما يمتلكه الأفراد من السمة المقاسة، ويمكن تعريف الاختبار النفسي والتربوي بأنه أداة قياس مقننة يتمّ تصميمها للحصول على قياس موضوعي لعينة من سلوك الأفراد، ويتمّ معايرة أدائهم بمستوى أداء محدد، ولا يُعدّ قياس القدرات والسمات غايةً بحذ ذاتها؛ إنما يُعدّ وسيلة للتعرف على الفروق الفردية بين الأفراد، والمقارنة بينهم، وتصنيفهم إلى مستوياتهم المختلفة، المنخفضة والمتوسطة والمرفعة، وهذا هو الهدف من الاختبارات التحصيلية، أبو حطب؛ عثمان؛ صادق [9].

ويرى بيكر [21] أن هناك هدفين يتحققان من تقدير قدرة المفحوص؛ أولهما: هو تقييم المفحوص تبعاً لقدرة الحقيقية التي تكمن وراء استجابته لقياس الاختبار، والهدف الثاني: هو مقارنة المفحوصين ببعضهم البعض لغايات التصنيف والابتعاث.

وتعتمد الاختبارات النفسية والتربوية في أساليب بنائها وتحليل بياناتها الدالة على السمة والقدرة المقاسة على نظريات تقوم على أسس ومبادئ وافتراسات معينة، فمنذ البدايات الأولى للقياس النفسي والتربوي، كان الاعتماد على النظرية الكلاسيكية للاختبارات، ووفق هذه النظرية فإنه يمكن التعبير عن قدرة الأفراد من خلال الدرجة الحقيقية التي تتضح من خلال أدائهم على الاختبار الطبري [2].

وعلى الرغم من شيوع استخدام هذه النظرية لسنوات، فإنها قد عانت من جوانب ضعف متعددة، تمثلت في تأثر الخصائص السيكو مترية للمفردات بعينة الأفراد الذين طبق عليهم الاختبار، وتأثر قدرات الأفراد بالخصائص السيكو مترية للمفردات، كما أن هذه النظرية لم تستطع التغلب على الكثير من المشكلات السيكو مترية المعاصرة؛ كتحيّز المفردات، وبنوك الأسئلة [5].

وقد أدت جهود العلماء في محاولتهم لمواجهة جوانب ضعف النظرية الكلاسيكية، إلى تطوير نظرية معاصرة في القياس النفسي والتربوي، أطلق عليها نظرية الاستجابة للمفردة (IRT) علام [1]. وتعدّ هذه النظرية ثورة في القياس النفسي والتربوي، وبواسطتها تمّ التغلب على الكثير من المشكلات التي تعرّض لها القياس في النظرية الكلاسيكية.

تبين -على حد علم الباحثة- ندرة الدراسات التي تناولت أثر اختلاف توزيع القدرة على معالم المفردات والأفراد، كما نجد أن غالبيتها اهتمت بدراسة توزيعات القدرة المتطرفة بمصاحبة متغيرات مثل طول الاختبار، وحجم العينة، وطريقة التقدير، بالإضافة إلى اختلافها في شكل توزيع القدرة، فمنها ما قسم القدرة لأربعة أشكال، بينما تم تقسيم القدرة في دراسات أخرى إلى ثلاثة أشكال. كما تباينت النتائج التي توصلت إليها الدراسات السابقة بالرغم من استخدامها لنفس برنامج تحليل البيانات bilogmg. لذا؛ جاءت هذه الدراسة استكمالاً لجهود الدراسات السابقة، وتركز على أثر اختلاف مستويات القدرة على معالم المفردات والأفراد في أسئلة الاختبار من متعدّد، باستخدام النمودج ثلاثي المعلم، وباستخدام بيانات مولدة، وبثلاث توزيعات للقدرة (مرتفعة، متوسطة، منخفضة).

وتتميّز الدراسة الحالية عن الدراسات السابقة في استخدامها لبرنامج Xcalibre، الذي يُعدّ من بين البرامج عالية التخصص في تحليل نظرية الاستجابة للمفردة، حيث إنه استناداً على دراسة Wang [18]؛ فإن برامج التحليل الإحصائي تختلف في دقة التقدير، وجودة المخرجات، كما أن قيم معالم المفردات والأفراد المقدرة ببرنامج (Xcalibre) أقرب إلى القيم الحقيقية، مقارنة ببرنامج (bilogmg) [19].

4. الإطار النظري

تُعدّ القدرات والاستعدادات من المواضيع التي عُني بها علماء النفس والتربية، ويختلف الأفراد في قدراتهم العقلية العامة، كما يختلفون في قدراتهم واستعداداتهم العقلية المختلفة، وقد اهتم علماء النفس والتربية بالقدرات العقلية للأفراد، والتمييز بينهم حسب مقدرتهم على الإدراك والاستنتاج والتفكير المجرد، وأطلقوا على هذه القدرات العقلية مصطلح "الذكاء"، وتعددت تعريفات الذكاء بتعدد علماء النفس، ومدارس علم النفس، حيث عرّف سبيرمان الذكاء بأنه "القدرة التي تتضمن تعلم العلاقات والارتباطات"، وعرّف بينيه الذكاء باعتباره حكماً جيداً وفهماً جيداً واستدلالاً جيداً فرج [20].

وقد دفع الاهتمام بالقدرات العقلية علماء القياس إلى تطوير الاختبارات التي تُعدّ الوسيلة الأساسية لقياس القدرات العقلية، ويعود الفضل للعالم بينيه في قياس القدرات العقلية وتقديرها بالاختبارات، حيث أعدّ أشهر اختبار للذكاء عام 1905، وقد مرّ هذا الاختبار بمراحل متعدّدة، وتعديلات مختلفة، وتجدد الإشارة إلى أن اختبار بينيه وجميع مقاييس القدرة العامة الأخرى، تعطي للمفحوص درجة واحدة فقط على الاختبار، ثم جاء سبيرمان بنظريته التي تقول بأن الذكاء يتكوّن من عاملين، هما: عامل عام موجود في كلّ النشاطات العقلية، وعامل خاص موجود في بعضها الآخر، بالإضافة إلى العامل العام أبو حطب؛ عثمان؛ صادق [9].

وفي عام 1938 أكّد ثرستون على وجود سبعة عوامل مختلفة سمّاها "القدرات العقلية الأولية"، وهي: القم اللفظي، والطلاقة اللغوية، والقدرة المكانية، والتذكّر، والاستقراء، والسرعة الإدراكية، والعددية، وهذه القدرات العقلية مستقلة وغير مرتبطة جميعها بعامل عام.

ويرى فيرنون أن النشاط العقلي في اختبار ما يُعتبر نتاجاً لأربعة عوامل، هي: العامل العام الذي تشارك به جميع الاختبارات، والعامل الطائفي الذي

أثر اختلاف قدرة الأفراد على دقة تقدير معالم المُفردات والأفراد في أسئلة الاختيار من متعدد

خضراء البلوي

ويعني أن مُفردات الاختبار تكون متجانسة فيما بينها، وتُقاس نفس السمة: أي: إن ثَمَّةَ سِمَةً واحدة تُفسَّر أداء الفرد على الاختبار كروكر والجينا [22].

افتراض الاستقلال المحلي أو الموضوعي local independence:

ويُقصد بهذا الافتراض أن تكون استجابات الفرد لمختلف مُفردات الاختبار مستقلةً استقلالاً إحصائياً؛ ممَّا يعني عدم تأثر استجابة الفرد لإحدى المُفردات باستجابته للمُفردات الأخرى.

افتراض التحرُّر من السرعة Speededness:

تفترض نماذج الاستجابة للمُفردة (IRT) أن عامل السرعة في الأداء ليس له دورٌ في الإجابة على المُفردات؛ أي: إن إخفاق الفرد في الإجابة على مُفردات الاختبار يعود إلى إخفاق قدرته، وليس إلى تأثير السرعة في الإجابة، أو عدم وصوله للمُفردات لقصر الوقت.

المنحنى المميِّز للمُفردة ICC Item characteristic curve:

ويُعدُّ أحد أهم المفاهيم المركزيَّة لنظرية الاستجابة للمُفردة (IRT)، ويمثِّل العلاقة الرياضيّة التي تربط بين احتمال إجابة الفرد إجابةً صحيحةً على المُفردة (i)، وبين القدرة أو السمة التي تقيسها مُفردات الاختبار الذي يحتوي على هذه المُفردة، وتكمن أهميَّة منحنى خصائص المُفردة في أنه يسمَح برؤية أن احتمال استجابة الفرد استجابةً صحيحةً على المُفردة يعتمد على السمة الكامنة، كروكر والجينا [22]، وأنه كلما كانت قدرة الأفراد أكبر، زادت فُرصهم للإجابة الصحيحة على المُفردة، والعكس بالنسبة لذوي القدرة الأقل.

5. الطريقة والإجراءات

أ. منهج الدراسة

تمَّ استخدام المنهج التجريبي، بالاعتماد على المحاكاة simulation study (23) hakim، ويُتيح تصميم الدراسة الحاليُّ بطريقة المونتي كارلو monte Carlo MCM للمتغيِّر المستقلِّ (مستويات القدرة المرتفعة، والمتوسطة، والمنخفضة) أن تؤثر بشكل مباشر - وبدون تدخل عواملٍ أخرى - في المتغيِّرات التابعة، وهي تقديرات معالم المُفردات، وتقديرات معالم الأفراد.

ب. مجتمع الدراسة

يتكوَّن مجتمع الدراسة من جميع الأفراد الذين تمتدُّ قدراتهم من (-3) إلى (+3)، وجميع الاختبارات أحاديَّة البُعد ثنائيَّة الاستجابة (0، 1) التي لها نفس توزيع المعالم، التي تمَّ تحديدها في الدراسة.

ج. عينة الدراسة

تمَّ استخدام عينة عشوائية من الأفراد بحجم (3000) فرد، بواقع (1000) فرد لكلِّ من القدرة المرتفعة، والمتوسطة، والمنخفضة، وذلك حسب توصية كروكر والجينا [22] ولورد [24] بأن يتراوح حجم العينة من (200) إلى (1000) فرد، عند استخدام نماذج نظرية الاستجابة للمُفردة، وكذلك استناداً إلى دراسة Drasgow [25]؛ وذلك للوصول لدقة في التقديرات عند استخدام النموذج ثلاثي البارامتر.

وبلغ طول الاختبار (60) مُفردةً ثنائيَّة الاستجابة، ومطابقة للنموذج اللوجستي ثلاثي المعلم، وتمَّ اختيار هذا الطول للاختبار استناداً إلى إشارة الشرفين [14] بأن هذا الطول يتناسب مع الاختبارات التي تُعدُّ في المدارس والجامعات والاختبارات الموضوعية المقتنَّة، إضافةً إلى ما أشار إليه بني عطا [26] بأن هذا الطول مناسبٌ للوصول إلى دقة في التقديرات عند استخدام النموذج ثلاثي المعلم، متفقاً مع دراسة سوامينتان وجيفورد [27] التي أكَّدت بأن دقة التقدير تزداد بزيادة طول الاختبار.

وتعرَّف هذه النظرية بأنها نموذج للقياس العقلي يوضِّح أن استجابات الأفراد للمُفردات في أيِّ اختبار يمكن تفسيرها بواسطة سمةٍ كامنة، وأن كلِّ فرد يمتلك قدرًا معيَّنًا من أيِّ سمةٍ كامنة.

والجدير بالذكر أن هذه النظرية قد قدَّمت طريقة في انتقاء المُفردات تتميز بتقديم معالم ثابتة للمُفردة، هي: معلَّم الصعوبة ويُرمز له بالرمز (b)، والتمييز ويُرمز له بالرمز (a)، والتخمين ورمزه (c)، بالإضافة إلى تمتُّعها بخاصية اللأ تغاير، التي تعني عدم تغَيُّر معالم المُفردات بتغَيُّر عينة الأفراد المختبرين، وعدم تغَيُّر تقديرات القدرات بتغَيُّر المُفردات، بشرط أن تقيس هذه المُفردات القدرة ذاتها [4].

كما أن دقة تقدير المستويات المختلفة من المُفردات تعتمد على مناسبة مجموعة المُفردات لمستويات القدرة المراد تقديرها؛ فالمُفردات الصعبة تناسب مستويات القدرة المرتفعة؛ بمعنى: أن الدقة تكون عاليةً، كما أن مجموعة المُفردات السهلة تناسب مستويات القدرة المنخفضة، حيث تكون درجة الدقة عاليةً كذلك [5].

ويبن هامبيلتون وسوامينتان [5] أن نظرية الاستجابة للمُفردة (IRT) تساعد على التنبؤ بأداء الفرد للاختبار بدلالة واحدة أو أكثر من الخصائص السيكومترية، وذلك من خلال العلاقة بين أداء الفرد الملاحظ على المُفردة، والسمة أو القدرة التي تكمن وراء تلك الاستجابة، ويُشار إلى هذه العلاقة من خلال المنحنى المميِّز للمُفردة (ICC)، الذي يأخذ أحد النماذج الثلاثة الشائعة الاستخدام للبيانات الثنائية التي تقدَّر بالدرجة (1) للاستجابة الصحيحة، و(0) للاستجابة الخاطئة، وهذه النماذج هي:

النموذج اللوجستي أحادي المعلم (IPLM)

ويسمَّى كذلك بنموذج راش، ويعرَّف هذا النموذج بدلالة معلَّم صعوبة المُفردة (b)؛ حيث تتباين مُفرداته في صعوبتها فقط، ويُفترض أن جميع المُفردات تميِّز بين الأفراد بنفس القدر، كما يُفترض عدم تأثر الإجابات بعامل التخمين، علام [1].

النموذج اللوجستي ثنائي المعلم (PLM2):

ويُفترض هذا النموذج أن الفقرات تختلف في صعوبتها (b)، وكذلك في تمييزها (a)، كما يُفترض عدم تأثر الإجابات بعامل التخمين (c) علام [1].

النموذج اللوجستي ثلاثي المعلم (PLM3):

ويُعدُّ النموذج العامُّ للنماذج ثنائيَّة التدرج، والأقلَّ تشدُّداً؛ إذ يسمَح باختلاف المُفردات في صعوبتها وتمييزها، وهذا الملاحظ في بناء الاختبارات، حيث يصعب إيجاد المُفردات التي تميِّز بين مستويات السمة أو القدرة بنفس الدرجة، كما افترض النموذج أحادي المعلم (PLM1).

كذلك يُفترض هذا النموذج تأثر الاستجابات بعامل التخمين، وهذا ما تميِّز به عن النموذج ثنائي المعلم (PLM2)، وهو ما أضافه بيرنوم، وأطلق عليه معلَّم الخطِّ التقاربي الأدنى، وهذا المعلم - كما يعرفه علام [1] - يحدِّد احتمال إجابة الفرد ذي القدرة المنخفضة إجابةً صحيحةً عن طريق التخمين، لذا؛ تبدو أهميته عند مطابقة بيانات الاختبارات من نوع الاختيار من متعدد لهذا النموذج، بينما يرى بيكر [21] أن التوصل للحلِّ الصحيح عن طريق التخمين، لا يرتبط بمستوى قدرة الأفراد.

وتفترض هذه النماذج عددًا من الافتراضات القويَّة في البيانات التي

ستطبق عليها؛ لكي تؤدي إلى نتائج يمكن الوثوق بها، ومن هذه الافتراضات:

افتراض أحاديَّة البُعد Unidimensionality:

تكوّن من مجموعة المُفَرّدات المولّدة تبعًا للمتغيّر المستقل (قدرة الأفراد)، وتبعًا للنموذج المستخدم (النموذج اللوجستي الثلاثي المُعلّم).

إجراءات الدراسة

تم توليد البيانات باستخدام أسلوب المحاكاة، وتعرّف المحاكاة بأنها تقليدٌ يحاكي الواقع العملي، ويُتيح هذا الأسلوب المجال للحصول على بيانات من مختلف المواقف الاختبارية، وتحت مختلف الشروط؛ لإيجاد حلول للمشكلات المتعلقة بالقياس النفسي والتربوي، وبعض القضايا الإحصائية. وقد تمّ استخدام برنامج WinGen3-3 للمحاكاة في الدراسة الحالية، ويُعدّ هذا البرنامج من أشهر برامج المحاكاة، وهو من إنتاج هان وهاميلتون [28]، ويعمل هذا البرنامج على محاكاة معالم المُفَرّدات واستجابات الأفراد باستخدام العديد من نماذج نظرية الاستجابة للمُفَرّدات IRT أحادية البعد، أو متعدّدة الأبعاد، مع شروط مختلفة كيفما تطّلب التطبيق.

وقد تمّت عملية توليد البيانات في البحث الحالي بطريقة المونتي كارلو mc باستخدام برنامج WinGen3، وحسب المراحل التالية:

المرحلة الأولى:

جرى توليد اختبار مكوّن من (60) مُفَرّدة ثنائية الاستجابة (0، 1)، ومطابقة للنموذج اللوجستي ثلاثي المُعلّم، وبلغت مرّات الإعادة Replication لكلّ ظرف لدراسة المحاكاة الحالية (100) إعادة، حيث أشار هارويل وزملاؤه [7] إلى أن دقّة التقديرات تزداد بزيادة عدد مرّات التكرار، بحيث تصبح دقّة هذه التقديرات أكثر استقرارًا، وفُق الشروط التالية:

توليد مُعلّم التمييز للمُفَرّدات وفُق التوزيع المنتظم (0.50 ~ Uniform، 2)، بمتوسّط (1.3) وانحراف معياري (0.4)، وهذه القيمة مماثلة للقيم الحقيقية التي استخدمها هاميلتون وسواميثان [5] حيث أكّدا على أن تكون قيمة مُعلّم التمييز تتراوح ما بين صفر و2 لوجت.

توليد مُعلّم الصعوبة للمُفَرّدات وفُق التوزيع المنتظم (-3 ~ Uniform، +3)، بمتوسّط (1.7)، وانحراف معياري (0.1).

ولكون النموذج المستخدم ثلاثي المُعلّم؛ كان لا بدّ من توليد مُعلّم التخمين للمُفَرّدات، وتمّ ذلك وفُق التوزيع المنتظم (0.10 ~ Uniform، 0.30)، وكما أشارت الأحمدي [29] فإن هذا التوزيع لمُعلّم التخمين يُنتج قيمًا تماثل قيم التخمين للاختبار الموضوعي ثنائي الاستجابة المؤلف من أربعة بدائل.

المرحلة الثانية:

جرى توليد استجابات (1000) فرد بمائة مرّة إعادة، وفُق الظروف التالية: توليد استجابات (1000) فرد من ذوي القدرة المرتفعة وفُق التوزيع المنتظم (1.5 ~ Uniform، +3)، بمتوسّط (2.24)، وانحراف معياري (0.4).

توليد استجابات (1000) فرد من ذوي القدرة المتوسّطة وفُق التوزيع المنتظم (-1.5 ~ Uniform، +1.5)، بمتوسّط (0)، وانحراف معياري (0.8).

توليد استجابات (1000) فرد من ذوي القدرة المنخفضة وفُق التوزيع المنتظم (-3 ~ Uniform، -1.5)، بمتوسّط (-2.2)، وانحراف معياري (0.4).

ومن ثمّ تمّ تجهيز البيانات وإدخالها بالصيغ التي تناسب مع البرامج الإحصائية المختلفة، والمستخدمّة في تحليل بيانات الدراسة الحالية، مع العلم بأن البيانات المولّدة حسب المواصفات السابقة تمّ اعتبارها بيانات لمعالم مُفَرّدات وأفراد حقيقية، وتمّ محاكاة بيانات الدراسة الحالية استنادًا إليها.

الأساليب الإحصائية للدراسة

للإجابة عن أسئلة الدراسة؛ تمّ استخدام الأساليب الإحصائية التالية: لتوليد الاختبار واستجابات عينة الدراسة؛ تمّ استخدام برنامج WinGen3.

وتم استخدام برنامج xcalibre 4.2.2 لتقدير معالم المُفَرّدات والأفراد لعينة الدراسة التي تمّ توليدها، ويوفّر هذا البرنامج مؤشرًا للحكم على حسن مطابقة المُفَرّدات لنماذج نظرية الاستجابة للمُفَرّدات من خلال توفير مؤشر مرتع كاي الإحصائي لحسن المطابقة، وذلك لكل مُفَرّد على حدّ؛ حتى يمكن تقييم ملاءمة كلّ مُفَرّد للنموذج، كما يوفّر معلومات كاملة عن تقديرات قدرة الأفراد ومعالم المُفَرّدات ودالة معلومات الاختبار والمُفَرّدات، بالإضافة إلى مجموعة من الرسوم والتمثيلات البيانية التي تمثّل جميع المعلومات للمُفَرّدات والاختبار ككلّ.

تم استخدام برنامج Spss؛ لإجراء تحليل للبيانات طبقًا للأسئلة البحثية، كما يلي: أولًا: لقياس دقّة تقدير معالم المُفَرّدات والأفراد؛ تمّ حساب التحيز Bias، ويُعدّ أحد مؤشرات دقّة القياس، حيث يمثّل الفرق الواقع بين القيم المقدّرة للمُعلّم والقيم الحقيقية له، وتمّ استخدامه في الدراسة الحالية للإشارة فيما إذا كان الخطأ في النموذج يعكس مبالًا منتظمًا للمغالاة أو للبخس في تقدير المعالم، حيث تعد قيمة التحيز مناسبة عندما تقترب من الصفر، وتشير إلى عدم وجود ميل منتظم للخطأ في تقدير المعلم، وإذا ارتفعت قيمة التحيز عن الصفر، فإن معنى ذلك أن القيم المقدّرة للمُعلّم أعلى من القيم الحقيقية له، وهذه مغالاة في التقدير Overestimating، أما إذا انخفضت قيمة التحيز عن الصفر، فإن التحيز سالب، وهذا يعني أن القيم المقدّرة للمُعلّم أقلّ من القيم الحقيقية له؛ مما يُعدّ بخسًا للتقدير Underestimating، كما تم حساب متوسط تحيز معالم المُفَرّدات بطرح القيم الحقيقية لها من القيم المقدّرة لمعالم المُفَرّدات، ومن ثمّ حسب متوسط التحيز بين القيم المقدّرة والحقيقية لمعالم المُفَرّدات عبر المُفَرّدات، ولمعالم الفرد عبر الأفراد الأحمدي [29].

ثانيًا: إجراء تحليل التباين الأحادي؛ لقياس حجم التأثير لتحيز وكفاءة تقدير معالم المُفَرّدات والأفراد، والحصول على قياسات لحجم التأثير الرئيس لأشكال توزيعات القدرة على تحيز وكفاءة النموذج المستخدم في الدراسة عند تقديره لمعالم المُفَرّدات والأفراد، ثم المتابعة بالاختبارات البعدية الملائمة إذا وجدت تأثيرات ذات دلالة إحصائية، إلّا أنه في حالة أحجام العينات الكبيرة - كما في الدراسة الحالية - بسبب العدد الكبير لمرّات الإعادة، فإن ثمة مقدارًا كبيرًا من القوة.

ولعلاج هذا الأمر؛ فإنه سيتمّ استعمال قياسات حجم التأثير Effect Size، ويعرّف حجم التأثير في حال تحليل التباين عن العلاقة بين المتغيّر المستقل من جهة، والمتغيّر التابع من جهة أخرى؛ أي: إن قيمة حجم التأثير تعبر عن الدرجة التي يمكن التنبؤ بها بالمتغيّر التابع من خلال المتغيّر المستقل أبو جراد [30].

وقد تمّ استعمال حجم التأثير Effect Size ES بديلاً لاختبارات الدلالة؛ حيث يمتاز حجم التأثير بإمكانية تفسيره بشكل مستقل عن حجم العينة، وهو مفيد في تحديد المقدار الفعلي للفروق بين متوسطات المجموعات.

ومؤشّر حجم التأثير الذي تمّ استخدامه هو مؤشّر حجم التأثير لكوهين Cohen d، وقد أشار كوهين [31] إلى معايير للحكم على قيمة حجم الأثر

خضراء البلوي

أثر اختلاف قدرة الأفراد على دقة تقدير معالم المُفردات والأفراد في أسئلة الاختيار من متعدد

فقد تمَّ حساب دقة تقدير معلّم الصعوبة، وذلك بحساب متوسط التحيز لتقديرات معلّم صعوبة المُفردات من خلال جمعها عبر مَرّات الإعادة المائة، ومن ثمَّ أخذ متوسط الاختلاف بين القيم الحقيقيّة والقيم المقدّرة لمعلّم صعوبة المُفردة؛ حيث تُشير القيم الموجبة لتقدير أكبر من القيم الحقيقيّة للمعلّم، بينما تعكس القيم السلبية تقديرًا أقلّ من القيم الحقيقيّة للمعلّم، ويوضّح الجدول (1) نتائج متوسط التحيز لتقديرات معلّم الصعوبة.

المستخرجة بواسطة مؤشّره (d)، حيث اعتبر القيمة (0.8) قيمة كبيرة، واعتبره صغيرًا عند القيمة (0.2)، ومتوسطًا عند القيمة (0.5).

6. النتائج ومناقشتها

فيما يلي عرض لنتائج الدراسة وفقاً لتسلسل أسئلتها على النحو التالي: أولاً: للإجابة عن سؤال الدراسة الأول الذي نصّ على: هل توجد فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha \leq 0.05$) في دقة تقدير معلّم الصعوبة، وفق النّمودج اللوجستي ثلاثي المُعلّم، تُعزى لاختلاف قدرة الأفراد (مرتفعة، متوسطة، منخفضة)؟

جدول 1

تحيز تقديرات معلّم صعوبة المُفردة

الانحراف المعياري	المتوسط	نوع القدرة
.0385018	-2.650413	القدرة المرتفعة
.0150010	.048282	القدرة المتوسطة
.0368933	1.093433	القدرة المنخفضة

مدى قدرة الأفراد متوافقاً مع مدى صعوبة المُفردات. كما تمَّ إجراء اختبار تحليل التباين الأحادي One way ANOVA، حيث كانت جميع التأثيرات ذات دلالة إحصائية عند مستوى الدلالة ($\alpha \leq 0.05$)، إلا أنه لوجود حجم عينة كبير بسبب مَرّات الإعادة (100) إعادة لكل طرف من ظروف الاختبار؛ فقد تمَّ التركيز على حجم التأثير (Effect Size) الذي يعود إلى التفاعل بين معلّم الصعوبة وأشكال القدرة (المرتفعة، والمتوسطة، والمنخفضة). ويُعتبر حجم التأثير (ES) بديلاً لاختبارات الدلالة الإحصائية، وهو مفيد في تحديد مقدار الفروق بين متوسطات المجموعات بشكل فعّال، لذا؛ فهو يزود الباحثين بمعلومات تتجاوز حدود اختبارات الدلالة الإحصائية الأحادي [29]. ويوضّح الجدول (2) حجم تأثير أشكال القدرة الثلاثة على معلّم الصعوبة.

يتضح من الجدول (1) أن قيم متوسط التحيز تشير إلى أن القدرة المرتفعة لها تحيز سالب؛ مما يعني أن القيم المقدّرة لمعلّم صعوبة المُفردة أقلّ في القيمة من القيم الحقيقيّة لمعلّم صعوبة المُفردة؛ أي: إنها تميل لبخس تقدير معلّم الصعوبة، في حين أن القدرة المتوسطة والمنخفضة لهما قيم تحيز موجبة؛ مما يعني أن القيم المقدّرة لمعلّم صعوبة المُفردة تكون أعلى من القيم الحقيقيّة لمعلّم الصعوبة؛ أي: إنها تميل للمغالاة في تقدير معلّم صعوبة المُفردة. إلا أن قيم تحيز القدرة المتوسطة هي الأقل؛ مما يعني أن القدرة المتوسطة هي الأكثر دقة في تقدير معلّم صعوبة المُفردة، وتتفق هذه النتيجة مع ما توصلت إليه دراسة الشريفيين والصبح [13] من أن القدرة المتوسطة أكثر دقة في تقدير معلّم الصعوبة، وقد أشار عابنة [11] إلى أن دقة تقدير معلّم الصعوبة تزداد عندما يكون

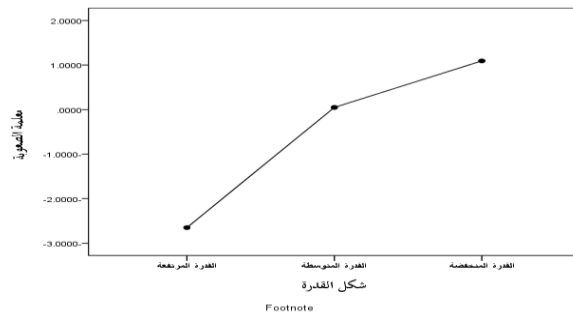
جدول 2

حجم التأثير لأشكال القدرة على صعوبة المُفردات

مصدر التأثير	حجم التأثير	قيمة F	القيمة الاحتمالية	الدلالة الإحصائية
أشكال القدرة الثلاثة	.03850	364868.243	.000	دالة إحصائية

توزيع القدرة ليس لها تأثير كبير على تحيز تقدير معلّم الصعوبة، إلا أن أدقها هي القدرة المتوسطة، ويمكن تفسير ذلك من خلال الشكل (1).

من الجدول (2)، يتضح صغر حجم تأثير أشكال القدرة (المرتفعة، والمتوسطة، والمنخفضة) على تحيز صعوبة المُفردات؛ مما يعني أن أشكال



شكل 1

طبيعة تأثير أشكال القدرة الثلاثة على تحيز تقدير معلّم صعوبة المُفردة

المُفردات التي تقع في مدى قدرات المفحوصين؛ للوصول إلى دقة أكبر في تقدير معالم المُفردات [5]، ويتضح كذلك أن تحيز معلّم الصعوبة يقل عند ذوي القدرة المتوسطة؛ مما يدل على قدرتهم على تقدير معلّم صعوبة المُفردات.

يتضح من خلال الشكل (1) أن القدرة المرتفعة تقلل من تقدير معلّم الصعوبة، وعلى النقيض، فإن القدرة المنخفضة تُغالي في تقدير معلّم الصعوبة؛ ولعلّ السبب في ذلك هو كون اختبار الاختيار من متعدد لا يقع في مدى قدرة ذوي القدرة المنخفضة والمرتفعة؛ مما يدعو إلى ضرورة اختيار

مرّات الإعادة المائة، ومن ثمّ أخذ متوسط الاختلاف بين القيم الحقيقية والقيم المقدّرة لمُعَلِّم تمييز المُفْرَدَة؛ حيث تُشير القيم الموجبة لتقدير أكبر من القيم الحقيقية للمُعَلِّم، بينما تعكس القيم السلبية تقديرًا أقلّ من القيم الحقيقية للمُعَلِّم. ويوضّح الجدول (3) نتائج متوسط التحيز لتقديرات مُعَلِّم التمييز.

جدول 3

تحيز تقديرات مُعَلِّم تمييز المُفْرَدَة

الانحراف المعياري	المتوسط	نوع القدرة
.0226652	-849743	القدرة المرتفعة
.0120111	-522936	القدرة المتوسطة
.0209339	-358923	القدرة المنخفضة

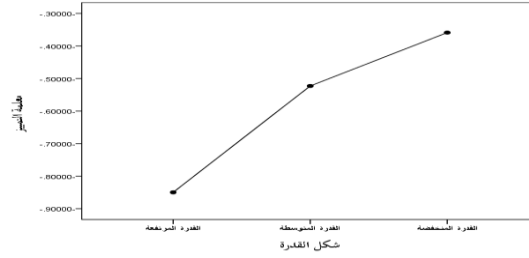
أنتج أعلى معاملات التمييز للمُفْرَدات. كما تمّ إجراء اختبار تحليل التباين الأحادي، حيث كانت جميع التأثيرات ذات دلالة إحصائية عند مستوى الدلالة ($\alpha \leq 0.05$)، إلا أنه لوجود حجم عينة كبير بسبب مرّات الإعادة (100) إعادة لكل طرف من ظروف الاختبار؛ فقد تمّ التركيز على حجم التأثير Effect Size ES الذي يُعتبر بديلاً لاختبارات الدلالة الإحصائية، ويعود إلى التفاعل بين مُعَلِّم التمييز وأشكال القدرة الثلاثة. ويوضّح الجدول (4) حجم تأثير أشكال القدرة الثلاثة على مُعَلِّم التمييز.

جدول 4

حجم التأثير لأشكال القدرة على المُفْرَدات

الدلالة الإحصائية	القيمة الاحتمالية	قيمة F	حجم التأثير	مصدر التأثير
دالة إحصائية	.000	17090.548	.991	أشكال القدرة الثلاثة

القدرة يؤثر على تحيز تقدير مُعَلِّم التمييز بشكل كبير، ويمكن تفسير ذلك من خلال الشكل (2).



شكل 2

طبيعة تأثير أشكال القدرة الثلاثة على تحيز تقدير مُعَلِّم تمييز المُفْرَدَة

متوسطة، منخفضة)؟ فقد تمّ حساب دقة تقدير مُعَلِّم التخمين، وذلك بحساب متوسط التحيز لتقديرات مُعَلِّم التخمين للمُفْرَدات من خلال جمعها عبر مرّات الإعادة المائة، ومن ثم أخذ متوسط الاختلاف بين القيم الحقيقية والقيم المقدّرة لمُعَلِّم تخمين المُفْرَدَة، حيث تُشير القيم الموجبة لتقدير أكبر من القيم الحقيقية للمُعَلِّم، بينما تعكس القيم السلبية تقديرًا أقلّ من القيم الحقيقية للمُعَلِّم.

ويوضّح الجدول (5) نتائج متوسط التحيز لتقديرات مُعَلِّم التخمين.

ثانيًا: للإجابة عن سؤال الدراسة الثاني الذي نصّ على: هل توجد فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha \leq 0.05$) في دقة تقدير مُعَلِّم التمييز، وفق النُمُوذج اللوجستي ثلاثي المُعَلِّم، تُعزى لاختلاف قدرة الأفراد (مرتفعة، متوسطة، منخفضة)؟ فقد تمّ حساب دقة تقدير مُعَلِّم التمييز، وذلك بحساب متوسط التحيز لتقديرات مُعَلِّم تمييز المُفْرَدات من خلال جمعها عبر

يُتضح من الجدول (3): أن قيم متوسط التحيز تشير إلى أن جميع أشكال القدرة لها تحيز سالب؛ مما يعني أن القيم المقدّرة لمُعَلِّم تمييز المُفْرَدات أقلّ في القيمة من القيم الحقيقية لمُعَلِّم التمييز؛ أي: إنها تميل لبخس تقدير مُعَلِّم التمييز، إلا أن القدرة المنخفضة هي الأكثر دقة في تقدير مُعَلِّم تمييز المُفْرَدَة؛ وذلك لكون قيمة متوسط تحيزها هي الأقل، وهذا يتفق مع ما جاء في دراسة عابنة [11] بأن القدرة المنخفضة تقدّر مُعَلِّم التمييز، وتختلف مع نتيجة بني عطا [15] بأن التوزيع ثنائي المنوال

يُتضح من الجدول (4) كبر حجم تأثير أشكال القدرة (المرتفعة، والمتوسطة، والمنخفضة) على تحيز مُعَلِّم تمييز المُفْرَدات؛ مما يعني أن اختلاف أشكال

يُتضح من خلال الشكل (2) أن جميع توزيعات القدرة تميل إلى التقليل من تقدير مُعَلِّم التمييز، إلا أن القدرة المنخفضة أقلها تحيزًا؛ مما يدلّ على دقتها في تقدير مُعَلِّم تمييز المُفْرَدات، يليها القدرة المتوسطة، وأخيرًا المرتفعة، كما يدلنا تقارب قيم التحيز إلى قدرة هذه التوزيعات الثلاث للقدرة على تقدير مُعَلِّم التمييز؛ ولكن بدرجات دقة متفاوتة.

ثالثًا: للإجابة عن سؤال الدراسة الثالث الذي نصّ على: هل توجد فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha \leq 0.05$) في دقة تقدير مُعَلِّم التخمين، وفق النُمُوذج اللوجستي ثلاثي المُعَلِّم، تُعزى لاختلاف قدرة الأفراد (مرتفعة،

تحيزُ تقديرات مَعلم تخمين المُفردة

الانحراف المعياري	المتوسط	نوع القدرة
.0063570	.043271	القدرة المرتفعة
.0013968	.032704	القدرة المتوسطة
.0013477	.036840	القدرة المنخفضة

المنتظم، بينما اختلفت جزئياً مع ما جاء في دراسة عبابنة [11] بأن دقة تقدير مَعلم التخمين تزداد مع ذوي القدرة المنخفضة. كما تم إجراء اختبار تحليل التباين الأحادي، وكانت جميع التأثيرات ذات دلالة إحصائية عند مستوى الدلالة ($\alpha \leq 0.05$)، إلا أنه لوجود حجم عينة كبير بسبب مرآت الإعادة، (100) إعادة لكل ظرف من ظروف الاختبار؛ فقد تم التركيز على حجم التأثير Effect Size ES الذي يُعتبر بديلاً لاختبارات الدلالة الإحصائية، ويعود إلى التفاعل بين مَعلم التخمين وأشكال القدرة (المرتفعة، والمتوسطة، والمنخفضة).

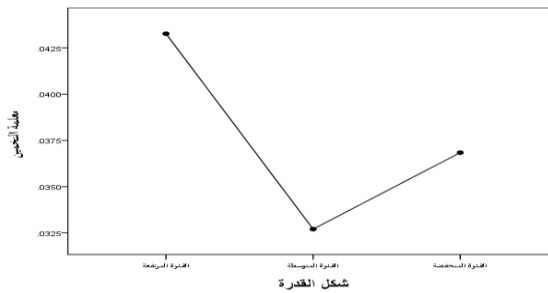
حجم التأثير لأشكال القدرة على تخمين المُفردات

الدلالة الإحصائية	القيمة الاحتمالية	قيمة F	حجم التأثير	مصدر التأثير
دالة إحصائية	.000	192.540	.565	أشكال القدرة الثلاثة

والذي يتضح من خلاله بأن جميع أشكال القدرة (المرتفعة، والمتوسطة، والمنخفضة) تُغالي بشكل طفيف في تقدير مَعلم التخمين، إلا أن القدرة المتوسطة أقلها تحيزاً؛ مما يدل على قدرتها على تقدير مَعلم التخمين، يليها القدرة المنخفضة، وأخيراً المرتفعة، كما يدلنا تقارب قيم التحيز إلى قدرة هذه التوزيعات الثلاث للقدرة على تقدير مَعلم التخمين.

يتضح من الجدول (5) أن قيم متوسط التحيز تشير إلى أن جميع أشكال القدرة لها تحيز موجب؛ مما يعني أن القيم المقدرة لمَعلم تخمين المُفردات أكبر في القيمة من القيم الحقيقية لمَعلم التمييز؛ أي: إنها تميل للمبالغة بشكل طفيف في تقدير مَعلم التخمين، ويتضح تقارب قيم القدرة المتوسطة والمنخفضة، إلا أن القدرة المتوسطة هي الأكثر دقة في تقدير مَعلم التخمين، يليها القدرة المنخفضة ثم المرتفعة، ويتفق هذا مع نتائج دراسة ساس ووكر [10] التي توصلت إلى أن التقديرات تكون أقل دقة عند التوزيع المتناهي السالب والمتناهي الموجب، وأكثر دقة في التوزيع

من الجدول (6) يتضح أن حجم تأثير أشكال القدرة (المرتفعة، والمتوسطة، والمنخفضة) متوسط على تحيز تخمين المُفردات؛ مما يعني أن اختلاف أشكال القدرة يؤثر على تحيز تقدير مَعلم الصعوبة، ويمكن تفسير ذلك من خلال الشكل (3).



شكل 3

طبيعة تأثير أشكال القدرة الثلاثة على تحيز تقدير مَعلم تخمين المُفردة

عبر مرآت الإعادة المائة، ومن ثم أخذ متوسط الاختلاف بين القيم الحقيقية والقيم المقدرة لمَعلم قدرة الفرد؛ حيث تشير القيم الموجبة لتقدير أكبر من القيم الحقيقية للمَعلم، بينما تعكس القيم السلبية تقديرًا أقل من القيم الحقيقية للمَعلم. ويوضح الجدول (7) نتائج متوسط التحيز لتقديرات مَعلم قدرة الأفراد.

رابعاً: للإجابة عن سؤال الدراسة الرابع الذي نص على: هل توجد فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha \leq 0.05$) في دقة تقدير القدرة، وفق النموذج اللوجستي ثلاثي المَعلم، تُعزى لاختلاف قدرة الأفراد (مرتفعة، متوسطة، منخفضة)؟ فقد تم حساب دقة تقدير مَعلم القدرة، وذلك بحساب متوسط التحيز لتقديرات مَعلم القدرة للأفراد، من خلال جمعها

تحيزُ تقديرات مَعلم قدرة الأفراد

الانحراف المعياري	المتوسط	نوع القدرة
.01331987	-2.1837180	القدرة المرتفعة
9.62	-0.00881777	القدرة المتوسطة
.02614698	2.1635470	القدرة المنخفضة

قدرة الأفراد أصغر في القيمة من القيم الحقيقية لمَعلم القدرة، وأن قيمة متوسط تحيز القدرة المرتفعة هي الأكبر؛ أي: إنها تميل لبخس تقدير مَعلم

يتضح من الجدول (7) أن قيم متوسط التحيز تشير إلى أن القدرة المرتفعة والقدرة المتوسطة لهما تحيز سالب؛ مما يعني أن القيم المقدرة لمَعلم

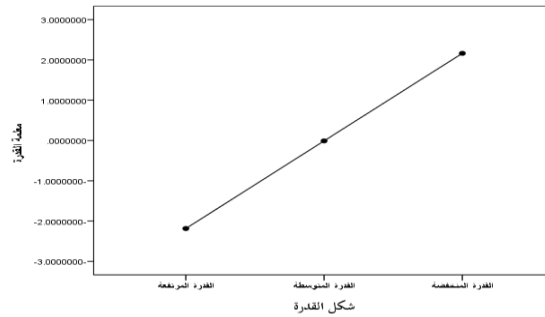
تقدير قدرة الأفراد. كما تم إجراء اختبار تحليل التباين الأحادي، إلا أنه لوجود حجم عينة كبير بسبب مرّات الإعادة، (100) إعادة لكل طرف من ظروف الاختبار؛ فقد تمّ التركيز على حجم التأثير Effect Size ES الذي يُعتبر بديلاً لاختبارات الدلالة الإحصائية، ويعود إلى التفاعل بين معلّم القدرة وأشكال القدرة الثلاث.

ويُتضح من الجدول (8) كبر حجم تأثير أشكال القدرة (المرتفعة، والمتوسطة، والمنخفضة) على تحيُّز معلّم قدرة الأفراد؛ مما يعني أن اختلاف أشكال القدرة يؤثّر على تحيُّز تقدير معلّم القدرة، ويمكن تفسير ذلك من خلال الشكل (4).

جدول 8

حجم التأثير لأشكال القدرة على قدرة الأفراد

مصدر التأثير	حجم التأثير	قيمة F	القيمة الاحتمالية	الدلالة الإحصائية
أشكال القدرة الثلاثة	1	#####	0	دالة إحصائية



شكل 4

طبيعة تأثير أشكال القدرة الثلاثة على تحيُّز تقدير معلّم القدرة

فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha \leq 0.05$) في كمية المعلومات التي تقدّمها كلُّ مُفردة، وفُق النُّموذج اللوجستي ثلاثي المُعلّم، تبعاً لاختلاف قدرة الأفراد (مرتفعة، متوسطة، منخفضة)؛ فقد تمّ تحليل البيانات باستخدام تحليل التباين الأحادي؛ للكشف عن مدى الفروق في المتوسطات الحسابية في تقدير كمية المعلومات التي تقدّمها كلُّ مُفردة باختلاف شكل توزيع القدرة، ويوضّح الجدول (9) نتائج تحليل التباين.

جدول 9

نتائج تحليل التباين لأثر شكل توزيع القدرة في كمية معلومات كل مُفردة

المتوسط	الانحراف المعياري	قيمة F	القيمة الاحتمالية	الدلالة الإحصائية
القدرة المرتفعة	.10575	17845.512	.000	دالة إحصائية
القدرة المتوسطة	.09561			
القدرة المنخفضة	.21411			

يكشف أن ثمة فروقاً ذات دلالة إحصائية بين الأوساط الحسابية لدالة معلومات المُفردة تبعاً لأشكال توزيعات القدرة الثلاثة، حيث أظهرت هذه الفروق أن أعلى فرق كان بين الوسطين الحسابيين لدالة معلومات المُفردات الناتجة من ذوي القدرة المنخفضة والمرتفعة؛ إذ بلغت قيمة الفرق (3.96164) لصالح ذوي القدرة المنخفضة، في حين أظهرت نتائج الفروق أن أقل فرق كان بين الوسطين الحسابيين لدالة معلومات المُفردات الناتجة من ذوي القدرة المنخفضة، والقدرة المتوسطة، حيث بلغت قيمة الفرق (1.78672) لصالح القدرة المنخفضة. ويتسجم هذا مع ما أشار إليه

القدرة، بينما تقترب قيمة متوسط تحيُّز القدرة المتوسطة من الصفر؛ مما يجعل منها الأكثر دقةً في تقدير معلّم قدرة الأفراد، وهذا يؤكد ما توصلت إليه دراسة الشريفيين [14] من أن دقة تقدير قدرة الأفراد تزداد في حالة عينة الأفراد ذوي القدرة المتوسطة، وتختلف هذه النتيجة مع نتائج دراسة كلٍّ من: الشريفيين والصباح [13]، والصباح [16]، حيث توصلت الشريفيين والصباح [13] إلى أن ذوي القدرة المرتفعة هم الأكثر دقةً في تقدير معلّم القدرة، وتوصلت الصباح [16] في دراستها إلى عدم وجود فروقٍ عمليةً في دقة تقدير معلّم القدرة باختلاف توزيع القدرة. كما يتضح من الجدول أن القدرة المنخفضة لها قيمة تحيُّز موجبة مرتفعة؛ مما يعني أن القيم المُقدّرة لمعلّم القدرة تكون أعلى من القيم الحقيقية لمعلّم القدرة؛ أي: إنها تميل للمغالاة في

يُتضح من خلال الشكل (4) أن القدرة المرتفعة تميل إلى بخس تقدير معلّم القدرة بشكل كبير، وتقلّل القدرة المتوسطة في التقدير بشكل طفيف، بينما تُغالي القدرة المنخفضة في تقدير معلّم القدرة؛ ولكن تُعدّ القدرة المتوسطة هي الأقلّ تحيُّزاً، وبالتالي الأدرق من بين أشكال القدرة الثلاثة في تقدير قدرة الأفراد.

خامساً: للإجابة عن سؤال الدراسة الخامس الذي نصّ على: هل توجد

تشير النتائج الواردة في الجدول (9) إلى أنه توجد فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha \leq 0.05$) بين متوسط كمية المعلومات التي تقدّمها كلُّ مُفردة تبعاً لشكل توزيع القدرة؛ مما يشير إلى تأثير دالة معلومات المُفردة بشكل توزيع القدرة، ويتضح أيضاً ارتفاع قيمة متوسط دالة معلومات المُفردة لذوي القدرة المنخفضة، يلها ذوو القدرة المتوسطة ثم المرتفعة.

ولمعرفة أيّ الفروق دالة؟ تمّ إجراء المقارنات البعدية باستخدام اختبار شيفيه Schffe Test، ويوضّح الجدول (10) مواقع هذه الفروق. حيث

أثر اختلاف قدرة الأفراد على دقة تقدير معالم المُفردات والأفراد في أسئلة الاختيار من متعدد

خضراء البلوي

وتختلف هذه النتيجة مع دراسة بني عطا [15] حيث كانت القيمة القصوى لدالة المعلومات لصالح التوزيع الثنائي المنوال، وكذلك تختلف مع دراسة الشريفين والصبح [13] التي وجدت أن القيمة القصوى للمعلومات قد قدّمتها الاختبار عند القدرة المرتفعة.

هامبيلتون وسوامينان [5] من أن المُفردات ذات التمييز العالي تُسهم بقوة أكبر في تأكيد دقة القياس من تلك ذات التمييز المتدني، ويُضيفان أن الاختبار الصعب يقدم معلومات أكبر عند مستويات القدرة المرتفعة، والاختبار السهل يقدم معلومات أكبر عند مستويات القدرة المنخفضة.

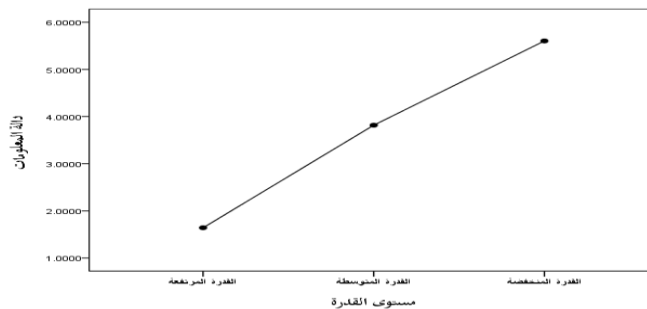
جدول 10

نتائج المقارنات الثنائية بين الأوساط الحسابية لدالة معلومات المفردة لكل توزيع للقدرة

شكل التوزيع	الفروق في المتوسطات	القيمة الاحتمالية	الدلالة الإحصائية
القدرة المتوسطة والقدرة المرتفعة	2.17492*	.000	دالة إحصائية
القدرة المنخفضة والقدرة المرتفعة	3.96164*	.000	دالة إحصائية
القدرة المنخفضة والقدرة المتوسطة	1.78672*	.000	دالة إحصائية

*دالٌّ إحصائيًا عند مستوى الدلالة ($\alpha \leq 0.05$)

كما يتّضح أيضًا من خلال الرسم البياني في الشكل (5) أن الفروق كانت لصالح ذوي القدرة المنخفضة.



شكل 5

نتائج المقارنات الثنائية بين الأوساط الحسابية لدالة معلومات المفردة لكل توزيع للقدرة

- [3] بعاره، وفاء صبيحي (2010). أثر اختلاف عدد بدائل الاختبار من متعدّد وبحسب المستوى الدراسي في الخصائص السيكمومترية للاختبار وفقراته (رسالة ماجستير غير منشورة). جامعة جدارا، الأردن.
- [4] رينولدز، س؛ ليفينغستون، ر (2013). إتقان القياس النفسي الحديث.. النظريات والطرق. (صلاح الدين محمود علام، مترجم). عمان: دار الفكر. (العمل الأصلي نُشر عام 2012).
- [6] علام، صلاح الدين محمود (2005). نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي. القاهرة: دار الفكر العربي.
- [9] أبو حطب، فؤاد؛ عثمان، سيد أحمد؛ صادق، أمال (2008). التقويم النفسي. القاهرة: مكتبة الأنجلو المصرية.
- [11] عباينة، عماد (2004). أثر حجم العينة وطريقة انتقائها وعدد الفقرات وطريقة انتقائها على دقة تقدير معالم الفقرة والقدرة لاختبار قدرة عقلية باستخدام نظرية الاستجابة للفقرة (رسالة دكتوراه غير منشورة). جامعة عمان العربية، الأردن.
- [12] الرشيد، عبد الجليل (2010). فحص اللاتغير في معالم الفقرات باختلاف معالم القدرة واللاتغير في معالم القدرة باختلاف معالم الفقرات (رسالة ماجستير غير منشورة). كلية التربية، جامعة مؤتة، الكرك.
- [13] الشريفين، نضال كمال؛ الصبح، رانيا محمد (2011). أثر بيئة فقرات الاختبار من متعدّد ومستوى القدرة لدى الأفراد على دقة التقديرات لمعالم الفقرات والأفراد وفق نظرية الاستجابة للفقرة. مجلة جامعة أم القرى للعلوم التربوية والنفسية، 3(2)، 45-110.

في ضوء النتائج التي تمّ التوصل إليها، فإن الباحثة توصي بما يلي: الاعتماد على النموذج اللوجستي ثلاثي المعلم عند إعداد الأسئلة التحصيلية. استخدام النموذج الثلاثي في تقدير قدرات الأفراد، ومعالم المُفردات، خاصّة في أسئلة الاختيار من متعدّد؛ لأنه الأكثر دقة؛ لاحتوائه على معلم التخمين الخاص بالأفراد ذوي القدرات المنخفضة. الاهتمام بتوفير البرامج الحاسوبية التي تقوم بتحليل مُفردات الاختبارات في ضوء نظرية الاستجابة للمفردة، وضرورة تدريب العاملين في الميدان التربوي عليها. الاعتماد على ذوي القدرة المتوسطة في تقدير معلم الصعوبة والتخمين وقدرة الأفراد في أسئلة الاختيار من متعدد ذات الأربعة بدائل. الاعتماد على ذوي القدرة المنخفضة في تقدير معلم التمييز، وللحصول على أكبر كم من المعلومات من مفردات اختبار الاختيار من متعدد. إجراء دراسة مماثلة باستخدام بيانات حقيقية، وباستخدام نفس البرنامج المستخدم في هذه الدراسة. وتتوجه الباحثة بجزيل الشكر لسعادة الدكتورة شرف حامد الأحمد، أستاذ مساعد في القياس والتقويم في جامعة طيبة لما قدمته من توجيه ونصائح قيمة ساعدت في إخراج هذه الدراسة للوجود.

المراجع

أ. المراجع العربية

- [1] علام، صلاح الدين محمود (2011). القياس والتقويم التربوي والنفسي؛ أساسياته، وتطبيقاته، وتوجهاته المعاصرة. القاهرة: دار الفكر العربي.
- [2] الطريحي، عبدالرحمن سليمان (2014). القياس النفسي والتربوي؛ نظريته، أسسه، تطبيقاته. الرياض: مكتبة الرشد.

- [7] Harwell, m., Stone, c., Hsu, t.-c., & Kirisci, I (1996). monte carlo studies in item response theory. 20, pp. 101-125.
- [8] Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of Item Response Theory. California: SAGE Publications.
- [10] Sass, D. A., Schmitt, T. A., & Walker, C. M. (2004) An evaluation of BILOG-MG with skewed theta distribution using various estimation procedures: A simulation study. Poster presented at the National Council on Measurement in Education. San Diego, California.
- [18] Wang, Joy (2018). Does it matter which IRT software you use? Yes. Woodbury MN: Assessment Systems Corporation.
- [19] Weiss, David; Minden, Shannon (2012). A comparison of Item Parameter Estimates from Xcalibre 4.1 and Bilog-Mg. Woodbury MN: Assessment Systems Corporation .
- [23] Hakim, C. (2012) Research Design: successful Designs for Social Economics Research. Routledge.
- [24] Lord, F.M. (1980) Applications of item response theory to practical testing problems. New Jersey: Lawrence Erlbaum associates publishers.
- [25] Drasgow, F. (1982) biased test items and differential validity. Psychological bulletin,92(2),526-531.
- [27] Swaminathan, H. & Gifford, J. (1979). Estimation of parameters in the three- parameter latent trait model (Report No. 90). Amherst. Massachusetts: university of Massachusetts, school of education, laboratory of psychometric and evaluation research.
- [28] Han, K. T., & Hambleton, R. K. (2014). User's Manual: WinGen (Center for Educational Assessment Report No. 642). Amherst, MA: University of Massachusetts, School of Education.
- [31] Cohen, J. (1988) Statistical power analysis for the behavioral sciences. New York: Routledge.
- [14] الشريفيين، نضال كمال (2012). أثر طريقة تقدير معالم الفقرة وقدرات الأفراد على قيم معالم الفقرة والخصائص السيكو مترية للاختبار في ضوء تغيّر حجم العينة. المجلة التربوية، 26(104)، 177-238.
- [15] بني عطا، زياد؛ الشريفيين، نضال (2012). أثر اختلاف شكل توزيع القدرة على معالم الفقرة ودالة المعلومات للاختبار. مجلة العلوم التربوية الأردنية، 8(2)، 166-151.
- [16] الصباح، عامر محمود (2014). المقارنة بين دقة تقدير القدرة باختلاف طول الاختبار وشكل توزيع معلّم القدرة تبعاً للنموذج اللوجستي ثلاثي المَعلم باستخدام بيانات حقيقية ومولدة (رسالة ماجستير غير منشورة). جامعة مؤتة، الأردن.
- [17] الحواري، أروى (2015). أثر طول الاختبار وشكل توزيع القدرة في تقديرات قدرة الأفراد وفق نموذج راش في نظرية استجابة الفقرة. مجلة جامعة النجاح للأبحاث (العلوم الإنسانية)، 29(8)، 1488-1464.
- [20] فرج، صفوت (2007). القياس النفسي. القاهرة: مكتبة الأنجلو المصرية.
- [21] بيكر، ف.ب (2010). أسس نظرية الاستجابة للمفردة. (عبد الرحمن الطريبي؛ السيد أبو هاشم؛ سوسن شلبي، مترجمين). الرياض: جامعة الملك سعود للنشر العلمي. (العمل الأصلي نُشر عام 2001).
- [22] كروكر، ل؛ ألجينا، ج (2017). مدخل إلى نظرية القياس التقليديّة والمعاصرة. (زينات دعنا؛ وهند الحموري، مترجمتين). عمان: دار الفكر. (العمل الأصلي نُشر عام 2006).
- [26] بني عطا، زايد صالح (2017). تقصي أثر طول الاختبار وحجم العينة على دقة طرق تقدير معالم الفقرات وقدرات الأفراد في برنامج بيبالوج. المجلة العالمية في الأبحاث التربوية والنفسية، 5(2)، 606-580.
- [29] الأحمدى، شرف حامد (2007). أثر اختلاف طول الاختبار وحجم العينة على دقة واتساق تقدير بارامترات المفردة والفرد في نظرية الاستجابة للمفردة (رسالة دكتوراه غير منشورة). جامعة أم القرى، مكة المكرمة.
- [30] أبو جراد، حمدي يونس (2013). قوة الاختبارات الإحصائية وحجم الأثر في البحوث التربوية المنشورة في مجلة جامعة القدس المفتوحة للأبحاث والدراسات. مجلة العلوم التربوية والنفسية، 14(2)، 368-350.
- ب. المراجع الأجنبية
- [5] Hambleton, R. K., & Swaminathan, H. (1985). Item Response Theory: Principles and Applications. USA: Kluwer-Nijhoff Publishing

THE EFFECT OF DIFFERENT PERSONS' ABILITIES ON THE ACCURACY OF THE ESTIMATION OF ITEMS PARAMETERS AND INDIVIDUALS IN MULTIPLE CHOICE QUESTIONS ACCORDING TO THE THREE PARAMETERS LOGISTIC MODEL

KHADRA ABDULLAH MAHMOUD AL-BALAWI

Master of measurement and educational evaluation, University of Taiba, Researcher and teacher in the Ministry of Education, Medina, KSA

ABSTRACT *This study aimed to investigate the effect of different abilities of individuals on the accuracy of the estimation of items and individuals' parameters in multiple choice exam questions according to the three parameters logistic model. To achieve the study purpose the responses three forms of abilities were generated using WinGen3 program on 60 dichotomous multiple choice items according to the three parameters logistic model. A sample of 3000 participants was divided into 1000 for each type of ability and the generated data was tested using SPSS and Xcalibre programmers. Statistical analysis was performed using Anova and the effect size and the accuracy of estimation were measured. The findings showed that there were significant statistical differences at the level of (0.05) in items and person parameters accuracy estimation attributed to the person's ability differences (high, moderate, low). The difficulty, guessing and ability parameters investigations were more accurate for moderate ability while the discrimination parameters were more accurate for low abilities. The results showed the variance between the size of information offered by each item according to the Person's abilities for low ability as its vocabularies were provided with the most information.*

KEYWORDS: *stimulation, Individual abilities, three parameters logistic model, multiple choice, vocabulary parameters.*