أثر موضع البند على الأداء التحصيلي للأفراد وعلى صعوبة البند

د/ ديانا فهمي حمّاد[1]

الملخص

يهدف البحث لدراسة أثر موضع البند على الأداء التحصيلي للأفراد معبراً عنه بدرجات الأفراد وبقدرتهم، كما يهدف إلى دراسة أثر موضع البند على صعوبته، ولتحقيق ذلك تم بناء نموذجين لاختبار تحصيلي يتكون من (٥٠) بنداً، رُتبت البنود في النموذج الأول وفقاً لتسلسل المحتوى التحصيلي، وفي النموذج الثاني تم إزاحة مواضع البنود بأقل من أو يساوي (٣) مواضع.

طُبق النموذجان على عينة حجمها (٢٣٢) طالبة في جامعة أم القرى، وعند اختبار أثر موضع البند على الأداء التحصيلي للأفراد بين النموذجين، أظهرت النتائج وجود أثر دال لموضع البند على درجات وقدرة الأفراد لصالح النموذج الثاني؛ وقبل اختبار أثر موضع البند على صعوبته، أُختزل عدد بنود الاختبار إلى (٣٧) بنداً، بسبب حذف البنود غير الملائمة لضرورات ترتبط بإجراءات تحليل البيانات وفقاً لنظرية الاستجابة للبند Item Response Theory (IRT)، وبسبب مطابقة البنود بين النموذجين لضرورات اختبار الفرض؛ ولم تشر نتائج اختبار هذا الفرض إلى وجود أثر دال لموضع البند على صعوبته بين النموذجين.

الكلمات المفتاحية

الأداء التحصيلي، درجات الأفراد، صعوبة البند، قدرة الأفراد، موضع البند.

---

[1] أستاذ القياس والتقويم التربوي والنفسي المساعد قسم علم النفس / كلية التربية جامعة أم القرى مكة المكرمة - المملكة العربية السعودية

أثر موضع البند على الأداء التحصيلي للأفراد وعلى صعوبة البند

د/ ديانا فهمي حمّاد[٢]

مقدمة

يلجأ الكثير من مطبقي الاختبارات التحصيلية إلى بناء صور متعددة من الاختبار في محاولة للحد من الغش ورفع أمان الاختبار؛ ويعد تغيير ترتيب البنود بين نماذج الاختبار، ممارسة شائعة لتحقيق ذلك مع الحفاظ على محتوى موحد يخضع له جميع المختبرين، هذا التغيير هو ما يُشار إليه بأثر موضع البند بين النماذج الاختبارية؛ وتتباين التوجهات حول الطريقة الأمثل لترتيب بنود الاختبار للحد من تأثيرها المحتمل، فقد اقترح كرونباخ (Cronbach, 1950) ترتيبها من الأسهل إلى الأصعب لتخفيض قلق الاختبار لدى المفحوص وزيادة دافعيته للإجابة على باقي البنود مما يرفع من أدائه التحصيلي، وأشار إلى أن ترتيبها بعكس ذلك يؤدي إلى رفع قلق الاختبار ويؤثر على استجابته على البنود التالية، أما بالش (Balch,1989) فيشير إلى أن ترتيب البنود وفق ترتيب المحتوى يعين الطالب على الحصول على درجات أعلى، لأنه يسمح لذاكرة الطالب باسترجاع المعلومات حسب طريقة ترميزها خلال عملية المذاكرة، وهو ما تدعمه نتائج (Carlson & Ostrosky,1992) و(Russell, Fischer, Fischer, & Premo, 2003).

في الآونة الأخيرة ونتيجة للأبحاث حول هذا الموضوع ظهرت نتائج تشير إلى أن تغير موضع البند بين النماذج الاختبارية المطبقة على الطلاب، قد يؤثر على معالم البند وخصائص الاختبار وعلى أداء المفحوصين، هذه المؤشرات تعد مصدر قلق في ميدان تطوير أدوات الاختبار (Harris, 1991; Hahne, 2008)؛ وذلك لأن المؤسسات التعليمية تعتمد على مخرجات الاختبارات في اتخاذ القرارات الخاصة بتحديد الأداء التحصيلي للأفراد، والاحتياجات التعليمية، وتقييم المناهج الدراسية، ولأغراض التعيين والاختيار والتوجيه والتقييم في هذه المؤسسات.

---

[٢] أستاذ القياس والتقويم التربوي والنفسي المساعد قسم علم النفس / كلية التربية جامعة أم القرى مكة المكرمة - المملكة العربية السعودية

مشكلة البحث

تزداد أهمية تحديد أثر موضع البند على خصائص الاختبار والأفراد، وذلك مع شيوع بناء نماذج من الاختبارات التحصيلية باستخدام نفس البنود مع تغيير ترتيبها بين النماذج للحد من الغش والمحافظة على تكافؤ النماذج ظاهريا (Debeer, & Janssen, 2013)، وقد أظهرت نتائج مسح الدراسات العلمية في هذا المجال وجود اختلاف في أهداف الأبحاث على سبيل المثال (Russell et al. 2003; Qian, 2014; Çokluk, Gül, & Dogan-Gül, 2016)، فهناك دراسات تعرضت لبحث أثر موضع البند على الأداء التحصيلي للأفراد معبرًا عنه بالدرجات في بعضها وبالقدرة في البعض الآخر؛ ودراسات هدفت لبحث هذا الأثر على صعوبة البند؛ ومن بين هذه الدراسات هناك دراسة واحدة هدفت لدراسة أثر موضع البند على صعوبة البند وعلى تقدير قدرة الأفراد معا، وذلك اعتمادًا على بيانات مولّدة وليست حقيقية، كما أظهر مسح الدراسات تباينًا واسعًا في نتائج الأبحاث في هذا المجال، مما يتطلب إجراء بحث لإلقاء مزيدًا من الضوء في هذا المجال.

وعليه فإن مشكلة البحث الحالي تتحدد في الأسئلة التالية:

١) هل توجد فروق دالة في الدرجات التحصيلية عائدة إلى تغير موضع البند بين نموذجي الاختبار التحصيلي؟.

٢) هل توجد فروق دالة في القدرة التحصيلية عائدة إلى تغير موضع البند بين نموذجي الاختبار التحصيلي؟.

٣) هل توجد فروق دالة في صعوبة البند عائدة إلى تغير موضعه بين نموذجي الاختبار التحصيلي؟.

أهداف البحث:

يهدف البحث الحالي إلى ما يلي:

- دراسة أثر تغير موضع البند بين نماذج الاختبار على الأداء التحصيلي معبرًا عنه بالدرجات وبالقدرة.
- دراسة أثر تغير موضع البند بين نماذج الاختبار على صعوبة البند.

أهمية البحث:

تتجلى أهمية البحث فيما يلي:

- تعد الاختبارات التحصيلية أدوات أساسية لتقييم تحقق نواتج التعلم في مختلف المراحل التعليمية، والبحث الحالي يساعد المختصين في مجال بناء الاختبارات على اتخاذ

قرارات صحيحة بشأن تغيير موضع البند والذي أصبح ممارسة شائعة لبناء نماذج متعددة للاختبار، كما قد تساعد نتائج البحث الحالي في تحسين الإجراءات المتبعة لزيادة دقة القياس.

- الدراسات الحديثة في البيئة العربية حول أثر موضع البند على التحصيل وعلى صعوبة البند ليست بالكثيرة، والبحث الحالي يعد إضافة في هذا المجال.
- نتائج البحث الحالي قد تسهم في ترجيح كفة النتائج المتباينة حول أثر موضع البند على التحصيل وعلى صعوبة البند.

مصطلحات البحث:

- أثر موضع البند Item Position Effect:

يقصد به الأثر الناتج عن تغيير مكان البند الاختباري بين النماذج الاختبارية للتحصيل في مقرر ما، ويستخدم معدو الاختبارات التحصيلية هذه الطريقة للحصول على نماذج اختبارية متكافئة وللحد من إمكانية الغش في الاختبارات.

- **الأداء التحصيلي Achievement Performance:**

هو الأداء الذي يعبر عن مقدار ماتعلمه الطالب من معارف ومهارات في مقرر دراسي تقدمه المؤسسة التعليمية التي التحق بها، ويتم تقييم هذا الأداء بأدوات مختلفة منها الاختبارات التحصيلية التحريرية بأنواعها المختلفة؛ ويُعبَّر عن هذا الأداء بالدرجات التحصيلية أو بقدرة الفرد.

- صعوبة البند Item Difficulty:

هي القيمة على متصل القدرة والمقابلة لاحتمال أن تكون الإجابة على البند إجابة صحيحة يساوي ٠.٥؛ ويعدُّ اللوجيت وحدة القياس الأساسية لصعوبة البند وفق النظرية الحديثة للقياس، وتعرَّف صعوبة البند المقاسة باللوجيت بأنها اللوغاريتم الطبيعي لترجيح الإجابة الخاطئة على البند للأفراد اللذين تقع قدرتهم عند نقطة الصفر على متصل القدرة.

الإطار النظري:

١- الأداء التحصيلي Achievement Performance

١-١ مفهوم الأداء التحصيلي

ارتبط التحصيل بالنشاط المعرفي للإنسان، وكان ارتباطه الأكبر بعملية التعلم حيث يعكس آثار الخبرة الناتجة عن عملية التعلم، ومن الملاحظ أن مفهوم التحصيل مضمَّنٌ في مفاهيم ترتبط بمفهوم التعلم، مثل دافعية الإنجاز، أو كفاءة الأداء، هذه المفاهيم تعكس الجانب

السلوكي المتمثل في الأداء المرتبط بتعلم المعارف والمهارات؛ فالتحصيل يتطلب أداءا ظاهرا لعملية التعلم المعرفية بحيث يمكن ترتيب هذا الأداء وتقدير درجاته وإخضاعه لعملية القياس (أبوحطب، ٢٠١١؛ علام، ٢٠٠٧).

١- ٢ قياس الأداء التحصيلي واختبارات المعلم

شاع استخدام الاختبارات التحريرية بنوعيها المقالي والموضوعي لقياس التحصيل، وكان لأسئلة الصواب والخطأ وأسئلة الاختيار من متعدد الموضوعية النصيب الأكبر في الأسئلة التحصيلية، لما تمتعت به الأولى من سهولة الإعداد والثانية من جودة الخصائص السيكومترية؛ وبالإضافة إلى التصنيف السابق، فقد صُنفت الاختبارات التحصيلية إلى اختبارات مقننة واختبارات من وضع المعلم؛ والأخيرة تعتمد جودتها على خبرة المعلم المهنية، وتتعدد أغراض هذا النوع من الاختبارات، فتستخدم لقياس ومتابعة درجة نمو الطالب أثناء المقرر، وللتعرف على الصعوبات التي يواجهها الطالب، ولاتخاذ قرار بشأن نجاح أو رسوب الطلاب في نهاية المقرر (أبوعلام، ٢٠٠٥)، وتكتسب أهميتها من شيوع استخدامها في المراحل التعليمية لتقييم تحصيل الطلاب في المقررات المختلفة.

يعتمد واضع الاختبار في بناء اختباره التحصيلي على معايير جودة صياغة الأسئلة والموجودة في مراجع قياس وتقييم التحصيل ومنها على سبيل المثال (أبوعلام، ٢٠٠٥؛ النبهان، ٢٠١٣) وغيرها، وكما سبق الإشارة إليه يقوم واضع الاختبار ببناء عدة نماذج باستخدام نفس البنود الاختبارية مع تغيير مواضعها بين النماذج، وذلك لرفع أمان الاختبار والحد من ظاهرة الغش مع الحفاظ على تكافؤ النماذج الاختبارية (Debeer, & Janssen, 2013).

٢- نظرية الاستجابة للبند Item Response Theory

ظهرت نظرية الاستجابة للبند للتغلب على مشكلات القياس المصاحبة للنظرية الكلاسيكية، فحققت العديد من متطلبات القياس الموضوعي بداية باستقلالية تقدير بنود المقياس عن تقدير قدرات الأفراد، واستقلالية تقدير قدرة الأفراد عن تقدير البنود، مرورا بالتخلص من تحيز البنود، بالإضافة إلى إمكانية تحقيق دقة قياس عالية لصدق وثبات المقاييس النفسية بحيث لا تتأثر بخصائص العينة أو بطول الاختبار، حيث توفر النظرية قياسا أكثر ثباتاً مع أخطاء أقل في القياس.

وقد قدمت النظرية عدة نماذج احتمالية منها ما يختص بمقاييس التقدير الذاتي متعددة الاستجابة، ومنها ما ارتبط بالمقاييس ثنائية الاستجابة Dichotomous، مثل المقاييس التحصيلية ومقاييس القدرات العقلية؛ هذه النماذج تباينت بتباين عدد بارامترات البنود المضمنة

في النموذج؛ ويعد نموذج راش أحادي المعلم الأكثر استخداما بين النماذج في بناء الاختبارات التحصيلية وتحليل نتائجها (Bachman, 1990).

٢-١ افتراضات النظرية

تقوم النظرية على عدة افتراضات (كروكر والجينا، ٢٠١٧؛ التقي، ٢٠١٣؛ علام، ٢٠٠٨) نوردها فيما يلي:

١) أحادية البعد Unidimetionality: ويشير إلى أن بنود المقياس متجانسة وتقيس سمة أو صفة واحدة فقط.

٢) الاستقلال المحلي Local Independence: ويقصد به أن احتمال الاستجابة على بند ما لا يؤثر على بند آخر، أي أن تقدير صعوبة بند ما في المقياس لايعتمد على تقديرات صعوبة البنود الباقية، ولايعتمد على تقدير قدرة الأفراد المستجيبين على المقياس.

٣) المنحنى المميز للبند Item Characteristic Curve (ICC): تفترض النظرية وجود دالة مميزة لكل بند تمثل احتمالية إجابة الفرد على البند كدالة لقدرته، وتأخذ شكل منحنى الترجيح اللوغاريتمي الذي تزداد فيه احتمالية إجابة الفرد على البند بزيادة قدرته.

٤) التحرر من السرعة Nonspeeded Test: تفترض النظرية أن استجابات الأفراد لا تتأثر بالسرعة، وأن الإخفاق في الإجابة يعود إلى محدودية قدرة الفرد؛ ويعد هذا الافتراض مضمنًا في افتراض أحادية البعد، فوجود تأثير للسرعة على أداء الفرد في الاختبار، يعني أن المقياس يقيس سمتين هما: سرعة الأداء والسمة الكامنة للمقياس (Hammbelton & Swaminthan, 1985).

٢- ٢ النماذج الأساسية لنظرية الاستجابة للبند

تعد النماذج ثلاثية وثنائية وأحادية البارامتر (نموذج راش) هي النماذج الأساسية لنظرية الاستجابة للبند، وجميعها تخضع لافتراضات النظرية الأساسية؛ وتعتمد على دوال لوجستية يتم تمثيلها بالمنحنى المميز للبند (ICC) والذي يختلف نتيجة لاختلاف عدد البارامترات.

يتضمن نموذج راش بارامترًا واحدًا هو بارامتر صعوبة البند، وتكون منحنياته المميزة للبند متوازية نتيجة لتساوي ميولها، مع اختلاف مواضعها على متصل السمة الكامنة (قدرة الفرد)، وفي النموذج ثنائي البارامتر الذي يتضمن بارامتري الصعوبة والتمييز، فإن

المنحنيات المميزة للبند تتباين في ميلها ومكانها على متصل السمة الكامنة، وفي النموذج ثلاثي البارامتر الذي يتضمن بارامترات الصعوبة، والتمييز، والتخمين، فإن المنحنيات المميزة للبنود تختلف أيضا في الجزء المقطوع من محور الصادات (كروكر والجينا، ٢٠١٧؛ التقي، ٢٠١٣؛ علام، ٢٠٠٨).

٢- ٣ اختيار النموذج الأنسب للبيانات

يعتمد اختيار النموذج الأنسب على عاملين:

١)غرض القياس.

٢)حجم العينة.

فإذا كان موقف القياس لايتطلب تخمينًا مثل مسح الاتجاهات فإن نموذج راش يعد كافيًا، وأيضًا يعد استخدامه مناسبًا في الاختبارات التحصيلية المطبقة لأغراض التقويم التكويني للتحصيل أو تقويم طرق التدريس، والتي يكون فيها حجم العينة محدودًا، وذلك لبساطته بسبب انخفاض عدد بارامتراته.

أما إذا كان الغرض من تطبيق الاختبار هو قياس التحصيل على مستوى قومي، أو بهدف اتخاذ قرارات حاسمة مثل تغيير المنهج الدراسي، فإن النموذج الثلاثي يعد الأفضل لدقته، حيث يتضمن بارامترات أكثر، ولإمكانية توفر عينة كبيرة لتحليل بياناته (علام، ٢٠٠٨؛ Hammbelton & Swaminthan, 1985).

الدراسات السابقة

هدف (Russell et al. 2003) إلى دراسة أثر اختلاف ترتيب الأسئلة بين النماذج على الأداء التحصيلي للطلاب، وذلك بتطبيق اختبار من (٣) نماذج، رُتبت الأسئلة في النموذج الأول وفق تسلسل المحتوى، وفي النموذج الثاني رُتبت عكس تسلسل المحتوى، وفي النموذج الثالث كان ترتيب الأسئلة عشوائيا؛ وطبقت النماذج الثلاث على عينة من (١٧٢) طالبًا بحيث خضع كل فرد للنماذج الثلاث على فترات مختلفة خلال الفصل الدراسي، وقد أظهرت النتائج وجود فروق دالة في القدرة التحصيلية للطلاب عائدة إلى اختلاف ترتيب الأسئلة بين النماذج وكان الأداء التحصيلي أقل في النموذج العشوائي، وأعلى في النموذج المرتب وفق تسلسل المحتوى .

صنف دافيس وفردوس (Davis & Ferdous, April 2005) في بحثهما البنود الاختبارية إلى فئتين، فئة تقيس المعرفة وفئة تقيس المهارة في مقرري الرياضيات والقراءة، ثم قسمت كل فئة إلى قسمين (١معرفة،٢ معرفة) (امهارة، ٢مهارة)، ثم قام الباحثان ببناء أربع

نماذج للاختبار، بنود النموذج الأول تتكون من قسمين: (١معرفة +١مهارة)، النموذج الثاني بنوده (١معرفة +٢مهارة)، النموذج الثالث (٢معرفة +١مهارة)، النموذج الرابع (٢معرفة + ٢مهارة)، هذه النماذج طبقت على عينات كبيرة من المفحوصين من طلاب الصف الثالث والخامس، وخرجت الدراسة بنتائج أهمها وجود أثر دال لتغير موضع البند على صعوبته في اختبار الرياضيات للصف الثالث، واختبار الرياضيات والقراءة للصف الخامس، وأشارت نتائج الدراسة إلى أن صعوبة البنود بين النماذج تزداد إذا كان اتجاه إزاحة موضع البند نحو نهاية الاختبار.

طبق بيتي جون الثاني وساكو (Pettijohn II & Sacco, 2007) دراسة على عينة من (٦٦) طالباً وطالبة، منهم (١٧) طالباً و(٤٩) طالبة، من الذين يدرسون مقرر مقدمة في علم النفس؛ بنى الباحثان ثلاثة نماذج للاختبار التحصيلي للمقرر، في النموذج الأول تم ترتيب البنود الاختبارية وفق تسلسل محتوى المقرر، وفي النموذج الثاني تم عكس الترتيب، وفي الثالث كان ترتيب البنود عشوائياً، بالإضافة إلى ذلك طبق الباحثان استبياناً لتحديد وجهة نظر الطالب في صعوبة النموذج الذي خضع له. وباستخدام تحليل التباين أحادي الاتجاه لم تكن هناك فروق ذات دلالة إحصائية بين متوسطات درجات الطلاب عائدة إلى اختلاف نماذج الاختبار، ولا إلى جنس الطالب، أو الوقت اللازم لاستكمال نماذج الاختبار، ولكن تصورات الصعوبة تأثرت بنموذج الاختبار الذي حصل عليه الطالب، حيث كان النموذج العشوائي أكثر صعوبة من نموذج تسلسل المحتوى في نظر الطلاب.

قام هوهينسن وزملاؤه (Hohensinn et. Al, 2008) بدراسة لتحديد أثر موضع البند وذلك من خلال نمذجة هذا الأثر في نموذج خطي لوجستي، وحساب معالم البند الناتجة من تطبيق اختبار تحصيلي في الرياضيات على مستوى وطني، هذا الاختبار يحوي (٤٨) بنداً طبقت على (١٧٩٢)طالبا من طلاب الصف الرابع، وقد تم بناء عدة نماذج للاختبار، يتكون كل نموذج من (٣٥) بنداً تتضمن (٩) بنود مكررة بين النماذج، وتشير النتائج إلى أن إزاحة البنود إلى نهاية الاختبار تزيد من صعوبة الاختبار حيث تكون قيم بارامترات التقدير دالة للبنود الأخيرة.

وفي نفس الإطار، هدف مايرز وزميليه (Meyers, et al., 2009) لدراسة أثر تغيير موضع البند على مستوى صعوبته، وبتحليل بيانات اختبارات مقننة في القراءة والرياضيات للطلاب من الصف الثالث إلى الثامن، هذه الاختبارات تراوح عدد البنود الاختبارية فيها من (٢٦) بند إلى (٤٨) بند، وقد وجدوا أن متوسط التغير في صعوبة البند يزداد بزيادة إزاحة

موضع البند بين نماذج الاختبار إذا تجاوزت الإزاحة (٥) مواضع بين النماذج.

وطبقت دراسة لي وزميليه (Li, et al., 2012) على بيانات مأخوذة من (٣) حزم من البنود الاختبارية في اختبار محوسب للتراخيص الطبية، كل حزمة تتكون من (٥٠) بنداً، تم ترتيبها بين النماذج عشوائيا، وبتطبيق عدة نماذج مختلفة لنموذج راش ونموذج الاستجابة الزمني (Time response model)، خرج الباحثون بأن نموذج راش هو الأفضل في جودة المطابقة بين باقي النماذج، وأن تغيير موضع البند لا يؤثر على صعوبته في النماذج المختلفة.

وهدف ألبانو (Albano, 2013) إلى دراسة التأثيرات المحتملة للتحيز في تقديرات نظرية الاستجابة البند للبنود والأفراد والبارامترات والناتج عن اختلاف موضع البند؛ بواسطة بيانات مأخوذة من تطبيق استطلاعي لاختبار قدرات الجامعيين (GRE) والذي يقيس القدرات الكمية واللفظية، ويتكون من بنود اختيار من متعدد قسمت إلى (٦) نماذج، ثلاثة منها كمية والباقية لفظية، يتضمن كل قسم كمي (٢٨) بنداً، تم توزيعها على (١٣) حزمة مع تغيير ترتيب البنود بين الحزم، طبق الاختبار على مفحوصين يبلغ عددهم (١٧٩٧، ١٨٥٢، ١٨٤٣) مفحوصًا للنماذج: كمي١، كمي٢، كمي٣ على الترتيب، أما الجزء اللفظي فتكون من (٣) نماذج، كلاً منها يتضمن (٣٠) بنداً وزعت على (٧) حزم، يتغير ترتيب البنود داخل الحزم، وبلغ حجم العينة التي خضعت للنماذج اللفظية (١٨١٠، ١٨٩٦، ١٩١٧) فردا على الترتيب، وباستخدام نموذج خطي هرمي عام لتقدير تأثير موضع البند توصل الباحث إلى أن تغير موضع البند يؤثر على صعوبته وذلك في جميع الأجزاء ماعدا لفظي ٢.

تهدف دراسة دي بير وجانسن (Debeer, & Janssen, 2013) إلى استخدام نماذج وصفية وتفسيرية لنظرية استجابة البند للكشف عن أثر تغير موضع البند ونمذجة هذه الآثار في إجراء من خطوة واحدة، بطريقة تسمح بالنظر في التأثيرات الفردية لتغير مواضع البنود على صعوبة البند، وذلك باستخدام دراسة محاكاة للكشف عن أثر موضع البند على معالم نموذج راش من خلال مقارنة المعالم في ثلاثة عينات افتراضية بلغ حجمها (٥٠٠، ١٠٠٠، ٥٠٠٠)، وقد أظهرت نتائج المحاكاة أن موضع البند يؤثر على تحيز تقدير صعوبة البنود، ولا يؤثر على تقدير قدرة الأفراد؛ كما أظهرت الدراسة أن صعوبة البند تزداد كلما أزيح البند إلى نهاية الاختبار.

استكشف (Qian, 2014) في بحثه العوامل المؤثرة على موضع البند في اختبار مقالي، استخدم البحث بيانات اختبار الكتابة الوطني لعام ٢٠٠٧ للصفوف الثامن إلى الثاني عشر، وتضمنت عينة أسئلة الكتابة (٢٠) سؤالاً يقيس الكتابة، كل سؤال يستغرق (٢٥) دقيقة، ويتم تقييم

الاستجابة على مقياس متدرج (غير مرض= صفر، غير مناسب =١، غير مقبول=٢، مقبول=٣، جيد=٤، ممتاز=٥)، تم توزيع الأسئلة على (٤٠)كتيب، كل سؤال يمكن أن يتكرر في (٤) كتيبات، بحيث يكون الأول في كتيبين والثاني في كتيبين آخرين، وكل كتيب يتضمن سؤالين فقط، وبمقارنة الفرق في معامل صعوبة البند محسوبًا بالطريقة التقليدية (المتوسط الموزون بالنسبة للدرجة العظمى) ووفق نظرية الاستجابة للبند، بين الموضع الأول والثاني للبنود، كانت الفروق دالة بين النماذج الاختبارية.

استخدمت (الحاج محمود، يونيو ٢٠١٥) (٦) نماذج اختبارية من نوع الاختيار من متعدد، كل نموذج مكون من (٦٠) بنداً اختباريًا، وذلك على عينة (٤٥٤) من طالبات الجامعة، تم ترتيب البنود في النموذج الأول والثاني والثالث وفق مستوى الصعوبة (من السهل إلى الصعب، ومن الصعب إلى السهل، وعشوائيا)، وتم ترتيب البنود في النموذج الرابع والخامس والسادس وفق تسلسل المحتوى (تسلسلي، عكسي، عشوائي) وبحساب الفروق في الأداء بين النماذج الست، كانت الفروق في متوسطات أداء الطالبات دالة لصالح النموذجين الرابع والخامس مما يشير إلى أن ترتيب البنود وفق تسلسل المحتوى يؤدي إلى ارتفاع أداء الطالبات في الاختبار.

هدفت دراسة تشوكلوك وزميليه (Çokluk et al., 2016) إلى فحص الفروق في دالة تباين البند في ثلاثة نماذج اختبارية، يتكون كلاً منها من (٢٠) بنداً اختباريًا من نوع الاختيار من متعدد، رتبت البنود في النموذج الأول عشوائياً، وفي النموذج الثاني رتبت البنود من الأسهل إلى الأصعب، والنموذج الأخير من الأصعب إلى الأسهل، وذلك اعتمادًا على أساليب النظرية الكلاسيكية للاختبار ((Classic Test Theory (CCT) ونظرية الاستجابة للبند (IRT)، مع أخذ مستويات صعوبة البند في الاعتبار. هذه النماذج طبقت على عينة كلية من (٥٧٨) طالبًا من طلاب الصف السابع؛ وكانت نتائج الدراسة تشير إلى وجود اختلاف جوهري بين الدوال التفاضلية لعدد كبير من البنود الاختبارية، مما يشير إلى أن اختلاف ترتيب البنود يؤدي إلى تباين في الأداء بين طلاب متساوي القدرة على نفس البنود؛ ونتيجة لذلك، فإن تغيير موضع البند يؤدي إلى تغير احتمال الحصول على الإجابة الصحيحة بين الطلاب متساوي القدرة، وأن ترتيب البنود من الأسهل إلى الأصعب يؤدي إلى حصول الطالب على درجات أفضل بصورة أكبر من النموذجين الآخرين.

قام وايرك وزملاؤه (Weirich et. al, 2016) بدراسة كان من أهدافها التحقق من أثر موضع البند على القياسات الكبيرة، حيث طبقت الدراسة على عينة من (٩٤١٠) طالبًا

وطالبة من طلاب الصف التاسع، وببناء اختبار تكون من (٣٨٦) بنداً، منها (٢٢٧) بنداً من نوع الاختيار من متعدد و(١٥٩) بنداً مقالياً، وزعت البنود الاختبارية على (٣١) حزمة، ويتراوح عددها في كل حزمة (بين١١ إلى١٤) بنداً، أعد الباحثون (٣١) كتيبًا يتكون كلاً منها من (٦) حزم تشمل (٧٦) بنداً في مجموعها الكلي. ويتغير موضع الحزم فقط عبر الكتيبات، في حين تظل مواضع البنود داخل كل حزمة ثابتة، تم تحديد (٢٠) دقيقة للإجابة على كل حزمة، وقد تخلل الاختبار تطبيق مقياس آخر لقياس جهد الاختبار، وكان من أهم نتائج الدراسة وجود أثر خطي لموضع البند، مما يشير إلى أن صعوبة البند تزيد بزيادة طول الاختبار، وأن أثر تغير موضع البند يقل بانخفاض طول الاختبار، لكنه لا يختفي حتى في العينات الفرعية ذات الدافعية العالية من المفحوصين مما يشير إلى أهمية بناء اختبارات يكون فيها تغيير مواضع البنود متوازناً بحيث تبقى بارامترات النموذج غير متحيزة.

قدم (Bulut, et al., 2017) نموذجا رياضيا بنائيا بناء على بيانات لاختبار من (٤) نماذج مكونة من نفس الأسئلة، وطبق على عينات كبيرة (كل نموذج طبق على عينة تقترب من ٣٠٠٠فرد)، هذا النموذج يسمح بتقدير صعوبة البند اذا تم تغيير موضعه، وقد أشارت الدراسة إلى أن صعوبة البند تزداد بزيادة إزاحة موضعه.

كان من أهداف بحث (Nagy et al., 2018) دراسة علاقة موضع البند بقدرة الطلاب في اختبار قراءة، تم تطبيقه على عينة من (٢٧٧٤) طالبا وطالبة من طلاب الصف الخامس ، هذا الاختبار يتكون من قطع للقراءة يتبعها بنود اختبارية ثنائية التصحيح، بلغ إجمالي عدد هذه البنود (٥١) بندا اختباريا، وزعت على (٩) نماذج، تراوح تغير موضع البند بين النماذج في أقصاه إلى (٢٢موضع) وفي أدناه إلى (موضعين)، وأشارت النتائج إلى عدم وجود ارتباط لتغيير موضع البند بقدرة الطلاب.

تعليق على الدراسات السابقة:

اختلفت أهداف الدراسات حول أثر موضع البند فهناك دراسات بحثت في أثر موضع البند على أداء المفحوصين، ودراسات بحثت في أثر موضع البند على صعوبته.

- في الأبحاث التي درست أثر موضع البند على الأداء التحصيلي للطلاب معبراً عنه بدرجات الأفراد أو بالقدرة نجد تبايناً في نتائجها حيث أن بعضها خرج بوجود أثر لموضع البند على التحصيل والبعض الآخر خرج بأن موضع البند لايؤثر على التحصيل، فأبحاث (Russell et al. 2003؛ الحاج محمود، يونيو ٢٠١٥؛ Çokluk, Gül, & Dogan-Gül, 2016) رغم أنها خرجت بوجود أثر لتغير موضع البند بين النماذج على الأداء التحصيلي مقاسا إما

بالدرجات أو بالقدرة، إلا أنها اختلفت في طريقة الترتيب المؤثرة على الأداء؛ فدراستي (Russell et al. 2003) و (الحاج محمود، يونيو ٢٠١٥) اللتان تعرضتا لبحث أثر اختلاف موضع البند على الأداء التحصيلي معبرا عنه بالدرجات، ودراسة (Çokluk et al., 2016) الذي تعرض لبحث أثر موضع البند على قدرة الطلاب؛ هذه الدراسات الثلاث اتفقت على أن الترتيب العشوائي يؤدي إلى أداء تحصيلي أقل، وأن الترتيب وفق تسلسل المحتوى هو الأفضل بالنسبة للأداء التحصيلي في البحثين الأولين، أما الدراسة الثالثة فاختلفت معهما في أن نتائجها تشير إلى أن الترتيب من الأسهل إلى الأصعب يؤدي إلى ارتفاع قدرة الطلاب، أما أبحاث (Pettijohn II & Sacco, 2007)، (Debeer, & Janssen, 2013)، (Nagy, Nagengast, Becker, Rose and Frey, 2018) فنتائجها تشير إلى أن موضع البند لا يؤثر على أداء المفحوصين أو قدرتهم.

- وفي الدراسات التي بحثت في أثر موضع البند على صعوبته، أيضا وُجد اختلاف في نتائج الأبحاث التي اهتمت بأثر تغيير موضع البند على صعوبته، فهناك أبحاث أشارت إلى وجود أثر لاختلاف موضع البند على صعوبته، وأبحاث لم تخرج بوجود هذا الأثر، فأبحاث كلاً من: (Davis & Ferdous, April 2005؛ Hohensinn, et al., 2008؛ Meyers, Miller, & Way, 2009؛ Debeer, & Janssen, 2013؛ Qian, 2014) خرجت بوجود تأثير لإزاحة موضع البند بين النماذج على صعوبة البند، بينما توصلت دراسة (Li, Cohen, & Shen, 2012) إلى أن تغيير موضع البند لا يؤثر على صعوبته؛ أما (Albano, 2013) فقد تباينت نتائجه فقد كان موضع البند مؤثرا على صعوبته في جميع نماذجه الاختبارية ماعدا نموذجاً واحدًا لم يظهر لموضع البند فيه تأثير على صعوبته؛ وأشارت معظم الأبحاث المؤيدة لوجود أثر لإزاحة البند على صعوبته، إلى أن إزاحة موضع البند بين النماذج إلى نهاية الاختبار يزيد من صعوبته بسبب الاجهاد (Davis & Ferdous, April 2005؛ Hohensinn et. al, 2008؛ Debeer, & Janssen, 2013)، ويزداد تأثير تغير موضع البند على صعوبته بزيادة إزاحته، فكانت الإزاحة مؤثرة إذا تجاوزت (٥) مواضع بين النماذج في بحث (Meyers, et al., 2009) أو أكثر (Bulut, Quo, & Gierl, 2017)، أما (Qian, 2014) فلم تشر إلا إلى وجود أثر لموضع البند على صعوبته؛ وتوصلت دراسة (Weirich, Hecht, Penk, Roppelt, & Böhme, 2016) إلى أن أثر تغير موضع البند يقل بانخفاض طول الاختبار.

فروض البحث:

نتيجة لاختلاف نتائج الأبحاث السابقة السابق عرضها فإن فروض البحث ستصاغ صياغة صفرية فيما يلي:

١) لا توجد فروق دالة في الدرجات التحصيلية عائدة إلى تغير موضع البند بين نموذجي الاختبار التحصيلي.

٢) لا توجد فروق دالة في القدرة التحصيلية عائدة إلى تغير موضع البند بين نموذجي الاختبار التحصيلي.

٣) لا توجد فروق دالة في صعوبة البند عائدة إلى تغير موضعه بين نموذجي الاختبار التحصيلي.

إجراءات البحث

منهج البحث:

يعتمد البحث على المنهج الوصفي المقارن لاختبار فروض البحث.

عينة البحث:

طُبق البحث على عينة من الطالبات بمرحلة البكالوريوس بكلية التربية بجامعة أم القرى، بلغ حجمها (٢٣٢) طالبة من طالبات كلية التربية السنة الثالثة.

أداة البحث:

هي اختبار تحصيلي في مقرر دراسي مقدم لطالبات البكالوريوس بجامعة أم القرى، قامت ببناء الأداة معدة البحث وتتكون من نموذجين، يتضمن النموذج الأول (٥٠) بنداً اختياريًا ثنائي التصحيح، تم ترتيبها حسب تسلسل المحتوى وذلك وفق ما أشار اليه (Balch,1989؛ Carlson & Ostrosky,1992؛ Russell et al. 2003؛ الحاج محمود، يونيو ٢٠١٥) من أن الطلاب يحصلون على درجات أعلى إذا كانت الأسئلة مرتبة وفق ذلك الترتيب، وتم بناء النموذج الثاني بإزاحة محدودة في موضع البند، حيث أن الإزاحة الكبيرة لموضع البند تؤثر على صعوبته (Bulut, et al., 2017)، وبما لا يتجاوز (٥) مواضع عن النموذج الأول وفق دراسة (Meyers, et al., 2009)، مع الحفاظ على ترتيب المحتوى.

ثبات الأداة:

حُسبت قيمة معامل ألفا كرونباخ لثبات كل نموذج على حدة وللنموذجين معاً، بعد تعديل ترتيب بنود النموذج الثاني وفق النموذج الأول ثم دمج عينة النموذجين، وفيما يلي توضيح لقيم معاملات الثبات لكل نموذج وبعد دمج النموذجين في جدول (١):

جدول (١) : قيم معامل ألفا كرونباخ لثبات الاختبار

رقم النموذج	معامل الثبات
١	٠.٨٤٠
٢	٠.٨٢٩
بعد دمج النموذجين	٠.٨٤٠

يوضح جدول (١) قيم معامل ألفا كرونباخ لنموذجي الاختبار وهي معاملات ثبات مرتفعة حيث بلغت قيمة ألفا للنموذج الأول القيمة (٠.٨٤٠)، وللنموذج الثاني (٠.٨٢٩)، وبعد دمج النموذجين (٠.٨٤٠).

صدق الأداة:

حسبت معاملات الارتباط بين بنود الاختبار وبين مجموع الدرجات للعينة الكاملة بعد دمج النموذجين وتوحيد ترتيب البنود بين النموذجين، فكانت جميع معاملات ارتباط البنود بالمجموع الكلي موجبة ودالة، ماعدا البند (٤٢) حيث بلغ معامل ارتباطه (-٠.١٦٥)، وبمراجعة معامل ثبات المقياس إذا حذف البند، كانت الزيادة في قيمة معامل الثبات (٠.٨٤٤) وهي زيادة طفيفة، فتم الإبقاء على البند.

بالإضافة إلى أن إجراءات تحليل الملائمة المستخدمة في نظرية الاستجابة للبند والتي طبقتها معدة الاختبار للحصول على بارامترات القدرة والصعوبة، يتم فيها حذف البنود غير الملائمة لتحقيق صدق القياس للبنود وللأفراد.

نتائج الدراسة:

اختبار الفرض الأول:

لا توجد فروق دالة في الدرجات التحصيلية عائدة إلى تغير موضع البند بين نموذجي الاختبار التحصيلي.

تم تطبيق اختبار (T) للمجموعات المستقلة للكشف عن الفروق بين متوسط درجات الطالبات والعائدة إلى اختلاف النموذج، والجدول (٢) يوضح أهم مقاييس النزعة المركزية والتشتت لدرجات الطالبات التحصيلية:

جدول ٢: المتوسط الحسابي والانحراف المعياري ومتوسط الخطأ المعياري لدرجات الطالبات في نموذجي الاختبار التحصيلي

رقم النموذج	عدد الطالبات	المتوسط الحسابي	الانحراف المعياري	متوسط الخطأ المعياري
١	١١٦	٣٨.٩٢	٦.٤١٦	٠.٥٩٦
٢	١١٦	٤١.٠٧	٥.٦٧٥	٠.٥٢٧

والجدول (٣) يعرض نتائج اختبار الفروق بين متوسطي درجات الطالبات في

نموذجي الاختبار التحصيلي:

جدول ٣: نتائج اختبار (T) للفروق في درجات الطالبات التحصيلية عائدة إلى تغير موضع البند بين نموذجي الاختبار

قيمة (T)	درجات الحرية (df)	مستوى الدلالة
٢.٦٩٩-	٢٣٠	**٠.٠٠٧

** دالة عند α= ٠.٠١

من جدول (٣) يتضح وجود فروق دالة عند (α= ٠.٠١) بين متوسط درجات الطالبات لصالح نموذج (٢)، هذه الفروق يمكن عزوها إلى تغير موضع البند بين النموذجين.

اختبار الفرض الثاني:

لا توجد فروق دالة في القدرة التحصيلية عائدة إلى تغير موضع البند بين نموذجي الاختبار التحصيلي.

تم حساب قدرة الفرد (θ_i) وفق نظرية الاستجابة للبند (IRT) وذلك باستخدام برنامج (Winsteps)، وفق الإجراءات الآتية:

١)دمج كامل العينة وذلك لأن حذف الأفراد والبنود غير الملائمة يختلف إذا تم معالجة كل نموذج على حدة.

٢)إعادة فصل العينة حسب النماذج.

٣)تطبيق اختبار (T) للمجموعات المستقلة بين متوسطي قدرات الأفراد في النموذجين.

وفيما يلي نتائج ذلك:

أولا: نتائج تحليل البيانات وفق (IRT):

١) بلغ مؤشر الثبات المبدئي للأفراد القيمة (٠.٧٩)، وللبنود القيمة (٠.٩٧) وهي معاملات مقبولة في البرنامج.

٢) حذف الأفراد غير الملائمين، في هذه المرحلة تم استبعاد (٥٠) فردًا غير ملائمين ليكون عدد الأفراد الملائمين (١٨٢) فردًا.

٣) حذف البنود غير الملائمة، في هذه المرحلة تم استبعاد (٧) بنود لا تتسق مع شروط نموذج القياس الموضوعي الذي توفره نظرية الاستجابة للبند (IRT)، ليكون عدد البنود الملائمة (٤٣) بنداً.

٤) بلغ مؤشر الثبات النهائي للأفراد القيمة (٠.٧٤)، وللبنود القيمة (٠.٩٤) ولا تزال المؤشرات ضمن الحدود المقبولة للبرنامج.

ثانيا: نتائج اختبار (T) لمتوسط قدرات الأفراد بين النموذجين:

تم تطبيق اختبار (T) للمجموعات المستقلة للكشف عن الفروق بين متوسط قدرة الطالبات والعائدة إلى اختلاف النموذج، والجدول (٤) يوضح أهم مقاييس النزعة المركزية والتشتت لقدرة الطالبات التحصيلية في نموذجي الاختبار التحصيلي:

جدول ٤: المتوسط الحسابي والانحراف المعياري ومتوسط الخطأ المعياري لقدرة الطالبات في نموذجي الاختبار التحصيلي

رقم النموذج	عدد الطالبات	المتوسط الحسابي	الانحراف المعياري	متوسط الخطأ المعياري
١	٨٥	٢.٢١٢٥	١.٣٥٠٧٢	٠.١٤٦٥١
٢	٩٧	٢.٦٩١٨	١.٣١٢٣١	٠.١٣٣٢٤

يتضمن جدول (٥) الآتي عرضاً لنتائج اختبار الفروق بين متوسطي قدرة الطالبات في نموذجي الاختبار التحصيلي:

جدول ٥: نتائج اختبار (T) للفروق في قدرة الطالبات التحصيلية العائدة إلى تغير موضع البند بين نموذجي الاختبار

قيمة (T)	درجات الحرية (df)	مستوى الدلالة
٢.٤٢٥-	١٨٠	**٠.٠١٦

** دالة عند α= ٠.٠١

يتضح في جدول (٥) وجود فروق دالة عند (α= ٠.٠١) بين متوسط قدرة الطالبات لصالح نموذج (٢)، هذه الفروق يمكن عزوها إلى تغير موضع البند بين نموذجين الاختبار التحصيلي.

اختبار الفرض الثالث:

لا توجد فروق دالة في صعوبة البند عائدة إلى تغير موضعه بين نموذجي الاختبار التحصيلي.

لاختبار هذا الفرض تم حساب صعوبة البند (p_i) وفق نظرية الاستجابة للبند (IRT) وذلك باستخدام برنامج (Winsteps)، حسب الإجراءات الآتية:

١) حساب صعوبة البند لكل نموذج على حدة، فنتج عن حذف البنود في تحليل الملائمة الثاني عدم تناظر البنود بين النموذجين.

٢) تم تحديد البنود غير المتناظرة وحذفها من النموذجين.

٣) تطبيق اختبار (T) للمجموعات المترابطة بين متوسطات الصعوبة للبنود المتناظرة في النموذجين.

وفيما يلي عرضًا لذلك:

أولا: نتائج تحليل البيانات وفق (IRT):

أ) نموذج الاختبار التحصيلي رقم (١):

١) بلغ مؤشر الثبات المبدئي للأفراد القيمة (٠.٨٠)، وللبنود القيمة (٠.٩٤) وهي معاملات مقبولة في البرنامج.

٢) حذف الأفراد غير الملائمين، في هذه المرحلة تم استبعاد (٢٦) فردا غير ملائمين ليكون عدد الأفراد الملائمين (٩٠) فردًا.

٣) حذف البنود غير الملائمة، تم استبعاد (٩) بنود ليكون عدد البنود الملائمة (٤١) بنداً.

٤) بلغ مؤشر الثبات النهائي للأفراد القيمة (٠.٧٦)، وللبنود القيمة (٠.٩١) ولا تزال المؤشرات ضمن الحدود المقبولة.

ب) نموذج الاختبار التحصيلي رقم (٢):

١) بلغ مؤشر الثبات المبدئي للأفراد القيمة (٠.٧٦)، وللبنود القيمة (٠.٩٢) وهي معاملات مقبولة في البرنامج.

٢) حذف الأفراد غير الملائمين، في هذه المرحلة تم استبعاد (٢٣) فردا غير ملائمين، ليكون عدد الأفراد الملائمين (٩٣) فردًا.

٣) حذف البنود غير الملائمة، نتج عن هذا التحليل، استبعاد (٥) بنود ليكون عدد البنود الملائمة (٤٥) بنداً.

٤) بلغ مؤشر الثبات النهائي للأفراد القيمة (٠.٧٠)، وللبنود القيمة (٠.٨٨) ولا تزال المؤشرات ضمن الحدود المقبولة.

ج) نتج عن تحليل كل نموذج على حدة، عدم تناظر في بعض البنود نتيجة لاختلاف البنود المحذوفة غير الملائمة في المرحلة الثانية من التحليل، فتم استبعاد البنود غير المتناظرة بين النموذجين وعددها (١٣) بنداً وهي: البنود (٣، ٨، ١٠، ١٩، ٢٠، ٢١، ٢٥، ٣٠، ٣١، ٣٥، ٤٢، ٤٦، ٤٧)، فكان العدد النهائي للبنود الداخلة في اختبار الفرض الثالث (٣٧) بنداً.

ثانيا: نتائج اختبار (T) لمتوسط صعوبة البنود بين النموذجين:

تم تطبيق اختبار (T) للمجموعات المترابطة للكشف عن الفروق بين متوسط صعوبة البنود بين النموذجين، والجدول (٦) يوضح أهم مقاييس النزعة المركزية والتشتت لصعوبة البنود التحصيلية:

جدول ٦: المتوسط الحسابي والانحراف المعياري ومتوسط الخطأ المعياري لصعوبة البنود بين نموذجي الاختبار التحصيلي

رقم النموذج	عدد البنود	المتوسط الحسابي	الانحراف المعياري	متوسط الخطأ المعياري
١	٣٧	٠,٣٥٩٧-	١,٧٨١٠١	٠,٢٩٢٨٠
٢	٣٧	٠,٥١١٤-	١,٩٨٤٢٠	٠,٣٢٦٢٠

يعرض جدول (٧) نتائج اختبار الفروق بين متوسطي صعوبة البنود بين نموذجي الاختبار التحصيلي:

جدول ٧: نتائج اختبار (T) للمجموعات المترابطة للفروق في صعوبة البند بين نموذجي الاختبار التحصيلي

قيمة (T)	درجات الحرية (df)	مستوى الدلالة
٠,٩٧٢	٣٦	٠,٣٣٨

يتضح من جدول (٧) عدم وجود فروق دالة بين متوسط صعوبة البند عائدة إلى تغير موضع البند بين نموذجي الاختبار التحصيلي.

مناقشة النتائج

أظهر اختبار الفرض الأول والثاني للدراسة الحالية، وجود فروق دالة في أداء الطلاب معبراً عنه بالدرجات أو بالقدرة لصالح النموذج التحصيلي رقم (٢)، هذه الفروق تُعزى إلى تغير موضع البند بين نموذجي الاختبار، وهذه النتائج تتسق مع نتائج أبحاث كلا من: (Russell et al. 2003؛ الحاج محمود، يونيو ٢٠١٥؛ Çokluk et al., 2016) وتختلف النتيجة مع الدراسة الأخيرة في أن البنود في البحث الحالي مرتبة حسب تسلسل المحتوى وليس حسب سهولة البند؛ وتختلف مع نتائج (Pettijohn II & Sacco, 2007؛ Nagy et al., 2018) التي تشير إلى عدم وجود فروق عائدة إلى تغير موضع البند، بين النماذج، وقد يعود ذلك إلى اختلاف العينة بين البحث الحالي والبحثين الآخرين من حيث الحجم والفئة العمرية، كما تختلف عن نتائج دراسة (Debeer, & Janssen, 2013) والتي استخدمت بيانات مولّدة وليست واقعية.

ونتيجة اختبار الفرضين السابقين تعد مثيرة للانتباه، فرغم اتخاذ تدابير مناسبة في ضوء نتائج الدراسات السابقة لبناء نموذجي الاختبار وذلك بترتيب البنود وفق ترتيب المحتوى، وإجراء تغيير محدود في موضع البند بين النموذجين، بهدف التقليل من أثر تغيير موضع البند، إلا أن هذا الأثر لايزال موجودًا بصورة دالة على تحصيل الطالبات، مما يتطلب إجراء دراسات أكثر للكشف عن الأسباب.

لم تشر نتائج الفرض الثالث إلى أن تغيير موضع البند يؤثر على صعوبته وهو ما

يتسق مع نتائج دراسة (Li, et al., 2012) وجزئيا مع نتائج (Albano, 2013)، ولا تتسق في ظاهرها مع أبحاث كلاً من: (Davis & Ferdous, April 2005)، (Hohensinn, et. al, 2008)، (Debeer, & Janssen, 2013)، (Qian, 2014) والتي خرجت بوجود أثر لتغير موضع البند بين النماذج؛ إلا أنها أيضا لا تتعارض معها، حيث أن الأبحاث (Davis & Ferdous, April 2005؛ Hohensinn et. Al, 2008؛ Debeer, & Janssen, 2013؛ Bulut, et al., 2017) أشارت إلى أن صعوبة البند تتأثر بالإزاحة الكبيرة لموضع البند تجاه نهاية الاختبار، وهو ما كان مستبعدا في البحث الحالي، وربما تشير نتيجة هذا الفرض إلى أن أثر تغيير موضع البند بين النماذج على صعوبة البند يبقى محدودا إذا كان تغيير ترتيب البنود بين النماذج صغيرا، حيث لم تصل الفروق بين متوسط صعوبة البنود في النموذجين إلى مستوى الدلالة، وهذا يتسق بشكل ما مع نتائج (Meyers, et al., 2009) التي أظهرت أن أثر تغير موضع البند يصبح واضحاً إذا تجاوزت إزاحة البنود بين النماذج (٥) مواضع، بالإضافة لذلك فربما يعود السبب أيضاً، إلى انخفاض طول الاختبار وهو ما يتسق مع نتائج (Weirich, et. al, 2016,) والتي تشير إلى أن أثر تغير موضع البند يقل بانخفاض طول الاختبار.

خلاصة القول أن نتائج الدراسة الحالية تشير إلى أن تغيير موضع البند بين النماذج المختلفة من الاختبار يؤثر على أداء الطالبات التحصيلي ولا يؤثر على صعوبة البند؛ ورغم أن الدراسات السابقة تشير إلى وجود أثر لموضع البند على صعوبته، فإن نتيجة الدراسة الحالية يمكن أن تكون مؤشرًا لإمكانية الحد من هذا الأثر في الاختبارات غير الطويلة بتحديد عدد مواضع إزاحة البند بين نماذج الاختبار ضمن مدى ضيق؛ كما تشير إلى أن تغيير موضع البند بين نماذج الاختبارات التحصيلية لبناء نماذج متعددة للاختبار، يجب أن يتم التعامل معه بحذر لوجود نتائج تشير إلى وجود تأثير لذلك التغيير على التحصيل.

المراجع

أبوحطب، فؤاد (٢٠١١). *القدرات المعرفية*. القاهرة: مكتبة الأنجلو المصرية.

أبوعلام، رجاء (٢٠٠٥). *تقويم التعلم*. عمان: دار المسيرة.

التقي، أحمد (٢٠١٣). *النظرية الحديثة في القياس*. الأردن: دار المسيرة.

الحاج محمود، فريال محمود (يونيو ٢٠١٥). أثر كل من ترتيب فقرات الاختبار ومستوى القلق على أداء الطلاب في اختبار من نوع الاختيار من متعدد. *مجلة العلوم التربوية والنفسية*، ١٦ (٢)، ٥٨٧- ٦٠٥.

علام، صلاح الدين (٢٠٠٧). *القياس والتقويم في العملية التدريسية*. عمان: دار المسيرة.

علام، صلاح الدين (٢٠٠٨). *نماذج الاستجابة المفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي*. القاهرة: دار الفكر العربي.

كروكر، ليندا وألجينا، جيمس (٢٠١٧). *مدخل إلى نظرية القياس التقليدية والمعاصرة*. (ترجمة هند الحموري وزينات دعنا). عمّان: دار الفكر.

النبهان، موسى (٢٠١٣). *أساسيات القياس في العلوم السلوكية*. عمّان: دار الشروق.

Albano, A. (2013). Multilevel modeling of item position effects. *Journal of Educational Measurement, 50*(4), 408-426.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Balch, W. (1989). Item order affects performance on multiple-choice exams. *Teaching of Psychology, 16*(2), 75-77.

Bulut, O., Quo, Q. and Gierl, M. (2017). A structural equation modeling approach for examining position effects in large-scale assessments. *Large-scale Assessments in Education, 5*(1), 1-20

Carlson, J. & Ostrosky, A. (1992). Item sequence and student performance on multiple-choice exams: Further evidence. *The Journal of Economic Education, 23*(3), p.232-235.

Çokluk, Ö., Gül, E. & Doğan-Gül, Ç. (2016). Examining differential item functions of different item ordered test forms according to item difficulty levels. *Educational Sciences: Theory & Practice, 16*(1), 319-330.

Cronbach, L. (1950). Further evidence on response sets and test Design. *Educational and Psychological Measurement, 10*(1), 3-31.

Davis, J., & Ferdous, A. (April 2005). Using item difficulty and item position to measure test fatigue. Paper presented at *the Annual Meeting of the American Educational Research Association*, Montreal, Quebec, Canada.

Debeer, D. & Janssen, R. (2013). Modeling item-position effects within an IRT framework. *Journal of Educational Measurement, 50*(2), 164-185.

Hahne, J. (2008). Analyzing position effects within reasoning items using

the LLTM for structurally incomplete data. *Psychology Science Quarterly*, *50*(3), 379-390.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer-Nijhoff.

Harris, D. (1991). Effects of passage and item scrambling on equating relationships. *Applied Psychological Measurement*, *15*(3), 247-256.

Hohensinn, C., Kubinger, K., Reif, M., Holocher-Ertl, S., Khorramdel, L. & Frebort, M. (2008). Examining item-position effects in large-scale assessment using the Linear Logistic Test Model. *Psychology Science Quarterly*, *50*, 391-402.

Li, F., Cohen, A. & Shen, L. (2012). Investigating the effect of item position in computer-based tests. *Journal of Educational Measurement*, *49*(4), 362-379.

Meyers, J., Miller, G. & Way, W. (2008). Item position and item difficulty change in an IRT-based common item equating design. *Applied Measurement in Education*, *22*(1), 38-60.

Nagy, G., Nagengast, B., Becker, M., Rose, N. & Frey, A. (2018). Item position effects in a reading comprehension test: An IRT study of individual differences and individual correlates. *Psychological Test and Assessment Modeling*, *60*(2), 165-187.

Pettijohn II, T. & Sacco, M. (2007). Multiple-choice exam question order influences on student performance, completion time, and perceptions. *Journal of Instructional Psychology*, *34*(3), 142-149.

Qian, J. (2014). An investigation of position effects in large-scale writing Assessments. *Applied Psychological Measurement*, *38*(7), 518-534.

Russell, M., Fischer, M., Fischer, C. & Premo, K. (2003). Exam question sequencing effects on marketing and management sciences student performance. *Journal for Advancement of Marketing Education*, *3*(1), 1-10.

Weirich, S., Hecht, M., Penk, C., Roppelt, A. & Böhme, K. (2016). Item position effects are moderated by changes in test-taking effort. *Applied Psychological Measurement*, *41*(2), 115-129.

The effect of item position on individuals' achievement performance and item difficulty

By: Dayana Fahmi Hammad (Ph.D.)

Faculty of Education/ Psychology Department

Umm Al-Qura University

Makkah – Saudi Arabia

Abstract

The research aims to investigate the effect of item position on achievement performance expressed by both Individuals' scores and abilities, on item difficulty.

Two forms of achievement test were built. The first form, consisted of (50) items, which has been ordered according to content sequence, the second form was built by shifting item positions less than or equal to (3) positions. Both forms were conducted on (232) students of Umm Al-Qura University. The results for testing the effect of item position on Individuals achievement performance indicated significant item position effect on Individuals scores and ability parameter in favor of the second form. When testing the effect of item position on item difficulty, the test length was reduced to (37) items due to firstly IRT item fit analysis, and secondly due to matching items between forms, and no significant effect of item position on item difficulty was detected.

Key words

Achievement performance, Individuals scores, Individuals ability, Item difficulty, Item position.