

أثر موقع الفقرات الصعبة في اختبار اختيار من متعدد على خصائصه السيكومترية وأداء الطلبة عليه.

د.زايد صالح بني عطا*

علي عبد الله الشريفيين**

الملخص

هدفت الدراسة إلى الكشف عن أثر موقع الفقرات الصعبة في اختبار اختيار من متعدد على خصائصه السيكومترية وأداء الطلبة عليه. ولتحقيق هدف الدراسة تم بناء اختبار تحصيلي في الحاسوب من نوع الاختيار من متعدد. تكوّن الاختبار بصورته النهائية من (50) فقرة، وزعت في أربعة نماذج، تختلف فقط في موقع الفقرات الصعبة فيها، حيث توزعت الفقرات الصعبة في النموذج الأول في العشر فقرات الأولى، وفي النموذج الثاني في العشر فقرات الوسط، وفي النموذج الثالث في العشر فقرات الأخيرة، وفي النموذج الرابع توزعت الفقرات العشر عشوائياً. تكوّن مجتمع الدراسة من جميع طلاب وطالبات الصف الأول الثانوي في المدارس الحكومية التابعة لمديرية قصبة إربد للعام الدراسي 2012/2013، والبالغ عددهم (8332) طالباً وطالبة، موزعين على (61) مدرسة. وتكونت عينة الدراسة من (519) طالباً وطالبة موزعين على (6) مدارس اختيرت بالطريقة المتيسرة، كما وزعت النماذج عشوائياً على أفراد عينة الدراسة. أظهرت نتائج الدراسة وجود فرق ذي دلالة إحصائية عند مستوى الدلالة ($0.05 =$) بين متوسطي معاملات الصعوبة بين النموذجين الأول والرابع فقط ولصالح النموذج الرابع، وعدم وجود فروق ذات دلالة إحصائية بين النماذج الأخرى. وفيما يتعلق بمعاملات التمييز، فقد بينت النتائج وجود فرق ذي دلالة إحصائية عند مستوى الدلالة ($0.05 =$) بين متوسطي معاملات التمييز بين النموذجين الأول والرابع ولصالح النموذج الرابع، وعدم وجود فروق ذات دلالة إحصائية بين بقية النماذج. كما كشفت نتائج الدراسة أن أعلى معامل ثبات كان لصالح النموذج الرابع، وأدنى معامل ثبات كان للنموذج الأول. وكشفت نتائج الدراسة أيضاً عن وجود فرق ذي دلالة إحصائية عند مستوى الدلالة ($0.05 =$) بين معاملي الصدق المرتبط بمحك الخاصين بالنموذجين الأول والرابع، وعدم وجود فروق ذات دلالة إحصائية بين قيم معاملات الصدق الخاصة بالنماذج الأخرى. وأخيراً أظهرت نتائج الدراسة وجود فرق ذي دلالة إحصائية عند مستوى الدلالة ($0.05 =$) بين متوسطي درجات أداء الطلبة على الاختبار التحصيلي في الحاسوب بين النموذجين الأول والرابع، وعدم وجود فروق ذات دلالة إحصائية بين النماذج الأخرى. الكلمات المفتاحية: صعوبة الفقرات، الخصائص السيكومترية، أداء الطلبة، ترتيب الفقرات، اختبار الاختيار من متعدد.

* أستاذ مشارك في قسم علم النفس الإرشادي والتربوي - كلية التربية - جامعة اليرموك - الأردن.

** وزارة التربية والتعليم - الأردن.

1- خلفية الدراسة

تعدّ الاختبارات التحصيلية من أهم الأدوات للوقوف على مدى ما اكتسبه الطلبة من معارف ومهارات في محتوى معين نتيجة مرورهم بخبرات تعليمية لتقويم تعلمهم. وهناك العديد من العوامل التي تؤثر على طبيعة هذه الاختبارات مثل: لغة الاختبار، شكل السؤال أو نوعه (Item format)، سهولة الفقرات أو صعوبتها، ترابط فقرات الاختبار، صياغة الفقرات وغيرها من العوامل، ولذلك من الصعب الحصول على معلومات تتعلق بمدى تحقق هذه الأهداف أو الأغراض المرجوة منها دون الإعداد الكافي، والتخطيط السليم المبني على نتائج الدراسات، التي بحثت أثر مثل هذه العوامل (Brown, 1983; Nitoko, 2001).

ويرى عوده (2010) بأن الاختبارات التحصيلية يمكن تصنيفها وفق معايير مختلفة، أحدها أن تصنف حسب طبيعة الإجابة المطلوبة، فتصنف إلى: الاختبارات ذات الإجابة المصاغة والاختبارات ذات الإجابة المنتقاة. حيث يتوقف استخدام أحد هذه الأنواع من الفقرات على مجموعة من العوامل منها: مستوى الهدف المراد قياسه، وعمر المفحوص، وغرض الاختبار، وعدد الطلبة، وظروف التطبيق، ومهارة المعلم في صياغة الأسئلة.

وبين برون (Brown, 1983) أن اختيار شكل الفقرات المناسب يعتمد على الخصائص السيكومترية للفقرة، ومهارة معد الاختبار في كتابة الفقرات، والالتزام بالإرشادات المتعلقة بشكل الفقرات وفلسفة التربية.

وتعدّ فقرات الاختبار من متعدد من أكثر أشكال الفقرات التي لها استخدام واسع في الاختبارات التحصيلية، واختبارات الاستعداد، حيث تتكون فقرة الاختبار من متعدد من جزأين هما: الجزء الأول ويمثل متن الفقرة، والذي هو عبارة عن جملة أو سؤال يطرح مشكلة محددة واضحة، أما الجزء الثاني يمثل البدائل، وتعد بمثابة الحلول المقترحة للمشكلة الواردة في متن السؤال (الشريفيين، 2013) ؛ (Gronlund & Linn, 1990). وتتميز فقرات الاختبار من متعدد بأنها تقيس مخرجات التعلم البسيطة بشكل أكثر فاعلية من فقرات الصواب والخطأ، وفقرات الإجابة القصيرة والمزاجية، بالإضافة إلى أنها يمكن أن تقيس مخرجات تعلم متنوعة في مستوى الفهم والتطبيق (Gronlund & Linn, 1990).

1990؛ Aiken, 2003)، إضافة إلى استخدامها بشكل كبير في بناء الاختبارات المقننة، واستخدامها الواسع في الاختبارات المدرسية من إعداد المعلم، وربما يعود ذلك لما يمتاز به هذا النوع من الاختبارات من قدرة على تمثيل المجال السلوكي للسمة موضوع القياس، ولسهولة تصحيحها ولموضوعيتها، غير أن هذه الاختبارات لا تخلو من أخطاء القياس، والتي تؤثر على دقة التقييم والقرارات المترتبة عليها (Ben-Simone, Bedescu & Nevo, 1997).

وعلى الرغم مما تمتاز به الاختبارات من نوع الاختيار من متعدد من ميزات، إلا أن هنالك عدد من الانتقادات قد وجهت إليها منها: عدد البدائل، التخمين، غموض وسطحية الفقرات، ومشكلة التحيز لموقع البدائل، وموقع الإجابة الصحيحة، والتي بدورها تؤثر على استجابة المفحوصين (Blunch, 2004; Eble & Frisbie, 1986; McMillan, 1984). وما تجدر الإشارة إليه بأن مستوى أداء المفحوص في الاختبار من نوع الاختيار من متعدد، يتوقف على ما يعرف بصعوبة الفقرة، وقدرتها التمييزية، وفاعلية المموهات، ومدى تحرر الفقرة بشكل عام من خلل صياغتها، ودرجة الوضوح في صياغة المتن، ومدى تجانس بدائل الفقرة، ودرجة علاقتها مع متن السؤال (Weiten, 1984).

لقد اعتبرت قضية ترتيب الفقرات في ورقة الاختبار بالإضافة إلى العوامل الأخرى مثل عدد البدائل، وشكل الفقرة والمموهات من القضايا السيكمومترية، التي اهتم بها البحث السيكمومتري لمعرفة الأثر الناتج منها على الخصائص السيكمومترية للفقرات، وللاختبار ومستوى أداء المفحوصين عليه. حيث أشارت نتائج دراسة كل من نيومان وكوندريت ولايني وبول (Newman, Kundert, Lane, & Bull, 1988) والتي هدفت لمعرفة أثر اختلاف ترتيب الفقرات على درجات اختبار الاختيار من متعدد حسب الأهمية الإحصائية والصعوبة المعرفية إلى أن الطلبة الذين تلقوا الفقرات بترتيب زيادة المعرفة، بصرف النظر عن ترتيب الصعوبة الإحصائية، كانت علاماتهم أعلى على الفقرات الصعبة، أما الطلبة الذين تلقوا الفقرات متفاوتة المعرفة والصعوبة الإحصائية، فسجلوا أعلى درجات على الفقرات متوسطة الصعوبة.

كما قام بلاش (Blach, 1989) بدراسة هدفت للإجابة عن السؤال: هل يحصل المفحوصين على درجات عالية عندما تكون الفقرات لاختبار تحصيلي في اختبار الاختيار من متعدد مرتبة بشكل متسلسل حسب صعوبتها أعلى من الدرجات التي يحصلون عليها عندما تكون فقرات الاختبار الاختيار من متعدد موزعة عشوائياً؟ للتحقق من الصدق التقاطعي عبر المجموعات، وتحديد فيما إذا كان

قلق الاختبار يتفاعل مع ترتيب الفقرات، فقد أجريت ثلاث تجارب على عينة مكونة من (253) من طلبة علم النفس، حيث تم تطبيق ثلاث صور للاختبار أحدهما مرتب ترتيباً عشوائياً، وثانيهما مرتبة حسب صعوبتها، وثالثهما استخدم فيها التصميم داخل الأفراد باستخدام تصميم المجموعات المتوازنة (Counterbalanced Design)، حيث وازن بين تسلسل الفقرات والتوزيع العشوائي لها، وتم الكشف عن أثر تفاعل القلق في تجربة واحدة، ولم يتم إيجاد دليل على الأداء المتفوق في الاختبار المرتبة فقراته في أي من صور الاختبار الثلاثة.

وفي دراسة قام بها آمودت ومكشان (Aammodt & McShane, 1992) بعنوان ما وراء التحليل للتحقق من أثر اختلاف خصائص فقرات الاختبار على درجات الاختبار وزمن الاختبار (عدد مرّات إكمال الاختبار). تكوّنت عينة المفحوصين من (32) طالباً، وزّعوا إلى ثلاث مجموعات (الأولى: (8) طلاب تلقوا (14) فقرة، حققت التساوي في عدد البدائل، الثانية: (20) طالباً تلقوا (26) فقرة مرتبة حسب الصعوبة، والثالثة: (4) طلاب تلقوا (16) فقرة مرتبة حسب المحتوى. وأشارت النتائج لاختلاف بسيط بين علامة الاختبار (المرتبة من السهل إلى الصعب) مع الاختبار (الفقرات الموزعة عشوائياً)، حيث تدرّج الفقرات من (السهل إلى الصعب) كانت أسهل من الفقرات التي تم ترتيبها بطريقة عشوائية، وكذلك من الاختبار الذي بدأ بفقرات صعبة، وأن الاختبار الذي كانت فقراته منظمة تبعاً للمحتوى كانت أسهل من الاختبار الذي تم اختيار فقراته بطريقة عشوائية، ولوحظ انخفاض في مستوى القلق لدى المختبرين في الترتيب (السهل - الصعب) مقارنة بالتوزيع العشوائي، وأن بدء الاختبار بفقرات صعبة يخفف علامة الاختبار، وكان التباين المتوقع من خطأ العينة أكثر من التباين الملحوظ، والحدود العليا والدنيا في فترة الثقة أقل من وسط حجم الأثر، وجميع قيم حجم الأثر صغيرة وفق تصنيف كوهن.

وأجرى كل من نيلي وسبرنجستون ومكان (Neely, Springston, & McCann, 1994) دراسة هدفت لمعرفة فيما إذا كان ترتيب فقرات الاختبار من متعدد يجعل الاختبار أكثر تحدياً للطلبة؟ حيث تكوّنت عينة الدراسة من (391) طالباً وطالبة من طلبة علم النفس في جامعة (Cape Breton) في كندا، وقُسمت العينة إلى ثلاث مجموعات، أجريت عليهم ثلاث تجارب مختلفة، التجربة الأولى: وزّعت (50) فقرة موزعة توزيعاً عشوائياً على (104) من المختبرين، التجربة الثانية: تم إجراؤها في يومين دراسيين متتاليين، في اليوم الأول وزّعت (50) فقرة متسلسلة أو عشوائية الترتيب على (149) من المختبرين، وفي اليوم الثاني أخذ كل طالب الاختبار البديل الذي لم يطبق عليه في اليوم الأول، وفي

التجربة الثالثة وزّعت (100) فقرة متسلسلة أو عشوائية الترتيب على (138) من المختبرين. أشارت نتائج الدراسة أن لا تأثير رئيسي لشكل توزيع فقرات الاختبار وأن وسط الأداء العشوائي أعلى من وسط الأداء على التوزيع المتسلسل، أما التفاعل بين ترتيب الفقرات والقلق، فقد كان دالاً إحصائياً حيث كان أعلى قلق مع الترتيب المتسلسل وأعلى من القلق في التوزيع العشوائي.

وأجرى دافيد (David, 2001) دراسة هدفت إلى معرفة أثر ترتيب فقرات الاختبار من متعدد من حيث الصعوبة وموقع الإجابات الصحيحة على كراسة الإجابة وجنس المفحوص وأداء الطلبة على الاختبار. تكونت عينة الدراسة المكونة من (484) طالباً وطالبة في الصف الرابع الأساسي، ولتحقيق أهداف الدراسة تم إعداد أربعة نماذج من الاختبارات التحصيلية، بحيث تم ترتيب الفقرات في النموذج الأول من السهل إلى الأصعب، والنموذج الثاني وزّعت فيه الفقرات بشكل طبيعي من حيث مستوى الصعوبة، أما النموذج الثالث فكان موقع الإجابات الصحيحة في الجهة اليمين في نموذج الاختبار، والنموذج الرابع فكان موقع الإجابات الصحيحة في الجهة اليسار من نموذج الاختبار. وأشارت النتائج إلى التأثير الرئيس في درجات الاختبار الدال إحصائياً ولصالح الذكور، أما بالنسبة لترتيب الفقرات فكان دالاً إحصائياً ولصالح النموذج الأول.

كما قام كامل (2009) بدراسة هدفت إلى بحث تأثير إعادة ترتيب بعض فقرات اختبار ستانفورد - بينيه (الصورة الرابعة) على تغيير مستوى أداء الأطفال: دراسة للفروق تبعاً لبعض متغيرات الشخصية والديموجرافية، حيث أجريت الدراسة في القاهرة على عيّنة من الأطفال، تكونت من (200) طفل مقسّمة إلى مجموعتين من (94) ذكور و(106) إناث، واستخدم مقياس ستانفورد - بينيه الصورة الرابعة. حيث تم إجراء اختبار قبلي (المقياس بصورته الرابعة) وتطبيق اختبار بعدي (المقياس بعد إعادة ترتيب فقراته)، حيث استخدم برنامج SPSS لتحليل البيانات. أشارت نتائج الدراسة إلى وجود فروق دالة إحصائياً بين درجات عيّنة الدراسة على مقياس ستانفورد - بينيه (الصورة الرابعة) قبل وبعد إعادة ترتيب بعض الفقرات الفرعية للمقياس، وذلك لصالح التصحيح بعد إعادة الترتيب، ووجود فرق ذي دلالة إحصائية بين درجات عيّنة الدراسة على مقياس ستانفورد - بينيه (الصورة الرابعة) قبل وبعد إعادة ترتيب بعض الفقرات الفرعية للمقياس، تبعاً لمتغير الجنس وكان الفرق لصالح الإناث.

وأجرى وينستون ورويديجر (Weinstein & Roediger, 2012) دراسة حول أثر ترتيب الفقرات على تقييم الأداء على الاختبار، وتكونت عينة الدراسة من (50) طالباً وطالبة، (27)

ذكور و(23) إناث من طلبة جامعة واشنطن من سن (18-22)، حيث قسّم الطلاب إلى مجموعتين: تكوّنت كل مجموعة من (25) طالباً وطالبة، وتكوّن الاختبار بصورته النهائية من (100) فقرة بالمعرفة العامة، وتم ترتيب الفقرات بنموذجين الأول: من الفقرات السهلة إلى الفقرات الصعبة، والثاني: من الفقرات الصعبة إلى الفقرات السهلة، حيث أخذت كل مجموعة بشكل عشوائي أحد النموذجين، واستخدم التصميم عبر المجموعات لترتيب الفقرات من (السهل - للصعب، الصعب - للسهل). ثم أجاب كل طالب عن فقرات الاختبار، حيث توقف الطلاب (10) مرّات خلال الاختبار، وطلب منهم تقييم أدائهم على آخر(10) فقرات، وتصنيف انطباعهم عن الاختبار وتوقع أدائهم النهائي، وأظهرت النتائج تفاوتاً أكثر لتقييمات الأداء على كل مجموعة من الفقرات مرتبة من الصعب إلى السهل على الرغم من أن الأداء على الطريقتين لم يختلف، مما يبيّن أن الطلبة أخفقوا في تقييم أدائهم، كما أن صعوبة الفقرة تغيّرت عبر اختلاف توزيع الفقرات.

وفي دراسة ميرز ومورفي وجودمان وتورهان (Meyers, Murphy, Goodman, & Turhan, 2012) والتي هدفت إلى معرفة أثر تغيير موقع الفقرات على معالم الفقرة، ونتائج المعادلة المشتركة تحت النموذج الثلاثي المعلمة. أجريت في ولايتين أمريكيتين (جنوب غربي الولايات المتحدة، وجنوب شرقي الولايات المتحدة) على طلبة المدارس في المستوى الأساسي والثانوي، وفي كلتا الحالتين خضع الطلبة لنفس المعيار في (القراءة، الحساب، الكتابة، الدراسات الاجتماعية، والعلوم) وتكوّنت أداة الاختبار من نموذجين (ورقة وقلم، وباستخدام الحاسوب) وبدون وقت معتمداً على الصف والموضوع، وتكوّن الاختباران من (45-70) فقرة، اختيار من متعدد وبنيت التقديرات حسب نظرية استجابة الفقرة، وتكوّنت عينة الدراسة من ثلاثة مجموعات من الطلاب مختلفة الحجم، (صغيرة، متوسطة، وكبيرة) تم اختيارها؛ بحيث تنتج هذه المجموعات أخطاء تحاكي كمية الأخطاء المقدّرة، التي يمكن أن تحدث عند معايرة فقرات الاختبار باختلاف أحجام العينات. وتم استخدام الإحصائي D^2 كدالة لتغيّر موقع الفقرة، واستخدم منحني خصائص الفقرة لإيجاد الصعوبة والتمييز، ولتحديد أي من الفقرات ستحفظ في النهاية. أشارت النتائج إلى أن قيمة الإحصائي D^2 تتأثر بشكل كبير بتغيّر موقع الفقرة، تبعاً لكل نموذج (ورقة وقلم، وباستخدام الحاسوب) من نماذج المجموعات الثلاثة، ولم يتغيّر تغيّراً ذا دلالة إحصائية بتغيّر حجم العينة.

إن المتصفح لنتائج الدراسات السابقة يجد بأنه لم يتطرّق أي منها إلى البحث عن أثر موقع

الفقرات الصعبة، حيث كانت تعتمد توزيع الفقرات حسب إحصائيات مختلفة بنفس النموذج مثل: صعوبة الفقرة ومؤشر الصدق ومؤشر الثبات والصعوبة المعرفية وغيرها، إلا أن الدراسة الحالية تميّزت عن الدراسات السابقة، بحيث اعتمدت مؤشراً إحصائياً موحداً؛ لإعادة ترتيب الفقرات وهو معامل صعوبة الفقرة، حيث وزّعت الفقرات الصعبة بحيث تكون متسلسلة في بداية الاختبار وفي وسط الاختبار وفي نهايته. وأخيراً تم توزيعها بشكل عشوائي، وانفردت أيضاً بأنها بحثت تأثير الخصائص السيكومترية للفقرة (الصعوبة والتمييز)، والخصائص السيكومترية للاختبار (الصدق والثبات)، والأداء على الاختبار بتغيّر موقع الفقرات الصعبة.

2- مشكلة الدراسة وأسئلتها

إن قياس التحصيل بأدوات على درجة عالية من الصدق والثبات أمر بالغ الأهمية، ومن خلال استعراض الأدب التربوي ذي الصلة بموضوع الدراسة وتحديد بناء الاختبارات التحصيلية، تضاربت الآراء حول ترتيب الفقرات، ونتائج هذا الترتيب على الخصائص السيكومترية للاختبار والخصائص السيكومترية للفقرات، واختلفت طرق ترتيب الاختبار، وتناول أكثرها الترتيب (من السهل إلى الصعب) والترتيب (من الصعب إلى السهل)، دون أن تتطرق هذه الدراسات والمقالات إلى وضع الفقرات الصعبة في بداية الاختبار ووسطه ونهايته وترتيبها عشوائياً، حيث انفردت هذه الدراسة بهذا النمط من ترتيب نماذج الاختبار حسب ترتيب الفقرات الصعبة.

وهناك رأي لبعض المتخصصين في القياس أنه من الأفضل ترتيب فقرات الاختبار من السهلة نسبياً إلى الفقرات الأصعب، لأن هذا الترتيب يعطي المفحوص ثقة وتشجيعاً عاليين، وتحديداً ذوي القدرات الضعيفة نسبياً. وكذلك يقلل من أرجحية تبديد الكثير من وقتهم في الفقرات الأعلى من قدرتهم، ومن ثم إهمال الفقرات الأسهل التي يمكنهم الإجابة عنها بصورة صحيحة (Anastasi, 2000). ولم تقدم الدراسات السابقة أو المتخصصون دليلاً عملياً على أهمية ذلك وانعكاساتها على الخصائص السيكومترية للاختبار المستخدم، ومن هنا تنشأ الحاجة إلى مثل هذه الدراسات.

من خلال الاطلاع على الدراسات السابقة لوحظ بأن الدراسات، تناولت أثر ترتيب الفقرات على الخصائص السيكومترية للاختبار والفقرات، (وفي حدود علم الباحثان) لم يجد دراسة مباشرة تتعلق بأثر موقع الفقرات الصعبة في اختبارات الاختيار من متعدد على خصائصه السيكومترية وخصائص فقراته، وهذا يبرر القيام بهذه الدراسة، وبالتحديد فإن مشكلة الدراسة تتمحور في الإجابة عن الأسئلة الآتية:

- 2-1- هل هناك فروق ذات دلالة إحصائية بين متوسطات معاملات الصعوبة للفقرات تعزى لاختلاف موقع الفقرات الصعبة (بداية الاختبار، وسط الاختبار، نهاية الاختبار، موزعة عشوائياً)؟
- 2-2- هل هناك فروق ذات دلالة إحصائية بين متوسطات معاملات التمييز للفقرات تعزى لاختلاف موقع الفقرات الصعبة (بداية الاختبار، وسط الاختبار، نهاية الاختبار، موزعة عشوائياً)؟
- 2-3- هل هناك فروق ذات دلالة إحصائية بين معاملات ثبات كرونباخ ألفا تعزى لاختلاف موقع الفقرات الصعبة (بداية الاختبار، وسط الاختبار، نهاية الاختبار، موزعة عشوائياً)؟
- 2-4- هل هناك فروق ذات دلالة إحصائية بين معاملات الصدق تعزى لاختلاف موقع الفقرات الصعبة (بداية الاختبار، وسط الاختبار، نهاية الاختبار، موزعة عشوائياً)؟
- 2-5- هل يختلف أداء الطلبة على اختبار الاختيار من متعدد باختلاف موقع الفقرات الصعبة في الاختبار؟

3- أهمية الدراسة:

تتمثل أهمية الدراسة الحالية في كونها من الدراسات العربية القليلة والنادرة عريباً - في حدود علم الباحثين- التي تبحث في قضية ترتيب الفقرات الصعبة في الاختبار، وانعكاس ذلك على خصائص الفقرات والاختبار، وهي بذلك تأتي منسجمة مع الاهتمام المتزايد والدعوات المستمرة بدراسة العوامل التي يمكن أن تؤثر على خصائص الفقرات والاختبار بشكل عام، خصوصاً فيما يتعلق بالاختبارات التحصيلية. كذلك تكمن أهمية هذه الدراسة بأنها ستقدم مؤشراً عملياً ودليلاً نظرياً لزيادة دقة خصائص الاختبار اعتماداً على ترتيب الفقرات الصعبة عند إخراج الاختبار بصورته النهائية. إذ أن ما تتوصل إليه الدراسة الحالية من تفضيل لترتيب الفقرات الصعبة في الاختبار سيرشد مطوري الاختبارات التحصيلية إلى أنسب الطرق في الحصول على دقة أعلى للاختبارات وأداء الطلبة عليه، ومن المؤمل كذلك إسهام هذه الدراسة بفتح المجال أمام الباحثين لإجراء المزيد من الدراسات في هذا المجال في ضوء ما ستتوصل إليه الدراسة من نتائج.

4- مصطلحات الدراسة:

- اختبار الاختيار من متعدد: اختبار تتكون كل فقرة من فقراته من أربعة ولها بدائل أربعة، يعطى الفرد علامة (1) عند اختيار البديل الصحيح والعلامة (0) عند اختيار البديل الخاطئ، وله أربعة نماذج ولا تختلف في المدة الزمنية المحددة للاختبار.

- معامل صعوبة الفقرة: نسبة الطلبة الذين أجابوا عن الفقرة إجابة صحيحة.

- معامل تمييز الفقرة: نسبة من أجاب عن الفقرة إجابة صحيحة من الفئة العليا مطروحاً منه نسبة من أجاب عن الفقرة إجابة صحيحة من الفئة الدنيا، ويقاس إجرائياً بقيمة معامل الارتباط بين درجات المفحوصين على الفقرة ودرجاتهم الكلية على الاختبار (Corrected Item Total Correlation).

- معامل ثبات الاتساق الداخلي للاختبار: وهو معامل يتم تقديره ليعبر عن درجة التجانس الداخلي بين فقرات الاختبار، وتم تقديره في هذه الدراسة باستخدام معامل كرونباخ ألفا.

- معامل الصدق المرتبط بمحك: حساب الارتباط بين درجات المقياس ومحك خارجي مستقل، وتم تقديره في هذه الدراسة من خلال حساب معامل ارتباط بيرسون بين الدرجات على اختبار الحاسوب بنماذج الأربعة كل على حدا، المعد من قبل الباحثان وعلامة الطالب النهائية في مادة الحاسوب في المدرسة.

5- حدود الدراسة:

تقتصر الدراسة الحالية على:

- 1-5- طلبة الصف الأول الثانوي في المدارس الثانوية التابعة لمديرية التربية والتعليم لقصبة إريد.
- 2-5- الاختبار التحصيلي المعد من قبل الباحثين لقياس المهارات والمعارف المتعلقة بالأنظمة العددية ولغة بيسك من مبحث الحاسوب للصف الأول الثانوي.
- 3-5- المفاهيم والمصطلحات التي تناولها الباحثان في الدراسة.

6- الطريقة والإجراءات:

لتحقيق الهدف من الدراسة استخدم الباحثان المنهج شبه التجريبي (Quasi Experimental Design) حيث كان المتغير المستقل موقع الفقرات الصعبة بمستوياته الأربعة (بداية الاختبار، وسط الاختبار، نهاية الاختبار، عشوائي) في حين كانت المتغيرات التابعة هي: صعوبة الفقرة، تمييز الفقرة، ثبات الاختبار، معامل الصدق وأداء الطلبة على الاختبار.

7- مجتمع الدراسة وعينتها:

تكون مجتمع الدراسة من جميع طلبة الصف الأول الثانوي في مدارس مديرية التربية والتعليم لمنطقة قصبة إربد، والبالغ عددهم (8332) بواقع (4012) طالباً و(4320) طالبة، موزعين على (61) مدرسة، منها (27) مدرسة للذكور و(34) مدرسة للإناث تبعاً للتقرير الإحصائي السنوي لمديرية التربية والتعليم في منطقة قصبة إربد للعام الدراسي 2013/2012 م . أما عينة الدراسة فقد تكونت من (519) طالبا وطالبة، منهم (252) طالبا و (267) طالبة موزعين على (6) مدارس منها ثلاث مدارس للذكور وثلاث مدارس للإناث، تم اختيارها بالطريقة المتيسرة (Available Sample) نظراً لاستعداد الإدارة المدرسية للتعاون، وتقديم التسهيلات اللازمة أثناء تطبيق الاختبار.

8- أداة الدراسة:

تم بناء اختبار تحصيلي في مبحث الحاسوب للصف الأول الثانوي؛ بهدف قياس تحصيل الطلبة في وحدتين: الأنظمة العددية (الوحدة الثالثة) ولغة بيسك (الوحدة الخامسة)، وذلك لتزامن الانتهاء من شرح الودحتين مع موعد تطبيق الاختبارات، وقد اقتضى بناء فقرات الاختبار ووضعه في صيغته النهائية الاسترشاد بالأسس العامة المتبعة في بناء اختبارات التحصيل الصفية التي وضعها جرونلند ولين (Gronlund & Linn, 1990) وفيما يأتي الإجراءات التي تم اتباعها:

أولاً: تحديد الغرض من الاختبار، وهو قياس تحصيل طلبة الصف الأول الثانوي في موضوعي الأنظمة العددية من الوحدة الثالثة، ولغة بيسك من الوحدة الخامسة في كتاب الحاسوب المقرر للعام الدراسي 2013/2012.

ثانياً: تحديد الموضوعات الداخلة في الاختبار، وذلك من خلال تحليل المحتوى لموضوعي الأنظمة العددية ولغة بيسك في الكتاب، بالاستعانة بدليل المعلم لمادة الحاسوب للصف الأول الثانوي؛ حيث تم صياغة النتائج التعليمية أهداف المحتوى) موضوع الاختبار، وبعد ذلك قام الباحثان بإعداد لائحة (جدول مواصفات) للاختبار، تمّ فيه ربط مستويات النتائج التعليمية بمحتوى المادة الدراسية موضوع الاختبار. حيث قام الباحثان بتوزيع الفقرات حسب الأهمية النسبية لكل موضوع تبعاً لحجم المادة التعليمية في وحدة الأنظمة العددية ووحدة لغة بيسك موضوع الاختبار.

ثالثاً: قام الباحثان بصياغة فقرات الاختبار من نوع الاختيار من متعدد، لكل فقرة منها أربعة بدائل إحداها الإجابة الصحيحة لقياس كل هدف من أهداف المحتوى، وقد يتطلب قياس بعضها أكثر من فقرة واحدة، وقد روعيت فيها الأسس الفنية في كتابة هذا النوع من الفقرات، ومطابقتها للهدف الذي يقيسه من حيث المحتوى والمستوى المعرفي، كما روعي ألا تعتمد إجابة إحدى هذه الفقرات على إجابة الفقرات الأخرى، وقد بلغ عدد الفقرات (60) فقرة بصورته الاولى.

رابعاً: قام الباحثان بالتحقق من صدق المحتوى، بعرض الصورة الأولى للاختبار، وجدول المواصفات، وتحليل المحتوى، والنتائج التعليمية (أهداف المحتوى)، على مجموعة من المتخصصين (المحكمين) من أصحاب الخبرة والكفاية، ثلاثة منهم مشرفون تربويون في مادة الحاسوب يحملون درجة الماجستير، واثنان منهم يحملون درجة الدكتوراة في القياس والتقويم، وسبعة منهم معلمون يدرّسون مادة الحاسوب، من أجل أخذ وجهات نظرهم في صدق الفقرة الاختبارية في قياس الهدف المحدد، واقتراح ما يروونه من تعديل مناسب. وللتحقق من ذلك بالنسبة للاختبار فقد طلب من كل محكم على حدة أن يبدي رأيه في الأمور الآتية: درجة الاتفاق بين الأهداف السلوكية (النتائج التعليمية) والهدف الأساسي للاختبار وتمثيلها له، ومدى قياس الفقرات للهدف السلوكي المتعلق بها، ومدى اتساق الاختبار ككل مع الهدف الأساسي منه، وأن يقترح ما يراه مناسباً لتعديل الفقرات أو الأهداف أو الفقرات والأهداف معاً. وبعد ذلك - وبناءً على آراء المحكمين - تم تعديل بعض الفقرات التي أجمع عليها المحكمون مثل (فقرة غامضة، وجود

أكثر من جواب صحيح بين البدائل، تغيير صيغة المتن، وجود بدائل ضعيفة). وتم الأخذ بالاقترحات والملاحظات التي أجمع عليها المحكمون، والتي ساعدت في إعادة النظر في بعض الأهداف ومراجعة عدد من الفقرات، وحذف عدد منها أو تبديلها بحيث يكون لها مدلول واضح، بالإضافة إلى تصويب الأخطاء اللغوية والنحوية والإملائية والمطبعية فيها. كما تم تغيير عدد من البدائل لبعض الفقرات، والتي أجمع المحكمون على أنها بدائل ضعيفة، وبلغ عدد الفقرات المحذوفة (8) فقرات، والتي كانت مرقمة وفق الترتيب الآتي في الاختبار (1,9,14,31,33,46,51,59)، وأعيد صياغة أو تبديل الفقرات (10,28,36,39,43,58). وبعد ذلك تم كتابة فقرات الاختبار، ثم قام الباحثان بتجميع الفقرات بعد تعديلها ومراجعتها، وبلغ عددها في صورتها الأولية (52) فقرة من نوع الاختبار من متعدد، وتم إعداد ورقة تعليمات للاختبار ورقة الإجابة النموذجية.

خامساً: التجريب الأولي للاختبار: تم تطبيق الاختبار بصورته الأولية على عينة استطلاعية مكونة من (57) طالبة من خارج عينة الدراسة موزعين على شعبتين؛ منهم (25) طالباً و(32) طالبة، تم اختيارهما بشكل عشوائي من شعب الصف الأول الثانوي؛ وكان الغرض من التجريب هو: التعرف إلى مستوى صعوبة الفقرات، وقدرتها التمييزية، ومن أجل تحديد الزمن الذي يستغرقه الطلبة في الإجابة عن فقرات الاختبار، وجمع أي ملحوظات حول فقرات الاختبار، وكذلك الفقرات التي تحتاج إلى تعديل أو حذف في ضوء ذلك، والتأكد من وضوح الصياغة اللغوية للفقرات، والكشف عن الفقرات الغامضة أو العبارات التي كثر التساؤل عنها، وبالتالي تعديل أو حذف الفقرات في ضوء ذلك. وتم تصحيح أوراق الاختبار يدوياً وإجراء عملية التحليل لبيانات التجريب الأولي باستخدام برنامج التحليل الإحصائي (SPSS) لحساب معامل الصعوبة لكل فقرة من خلال إيجاد نسبة الطلبة الذين أجابوا عن الفقرة إجابة صحيحة من بين الطلبة الذين حاولوا الإجابة عن هذه الفقرة. كما تم حساب معاملات التمييز للفقرات، والتي تمثل معاملات الارتباط بوينت بايسيرال (Point biserial) بين الفقرة والاختبار (Corrected Item Total Correlation) وبناءً على نتائج التحليل فقد كانت فقرات الاختبار جميعها ذات معاملات صعوبة موجبة، في حين وجدت فقرتان معاملي التمييز لهما سالبين. وبالتالي تم حذفهما ليصبح عدد فقرات الاختبار بصورته النهائية بعد

التجريب الأولي (50) فقرة، والإبقاء على الفقرات ذات الصعوبة الموجبة، والتي قيمها أقل من 0.20؛ لتشكيل نماذج الاختبار المتباينة في موقع فقراتها الصعبة.

سادساً: لغايات تشكيل نماذج الاختبار تم اختيار 10 فقرات تراوحت صعوبتها بين (0.12 - 0.20) والتي تعد فقرات صعبة والمجموعة (3,11,15,20,25,30,34,38,44,49) في الصورة

النهائية للاختبار فقد تم تشكيل نماذج الاختبار على النحو الآتي:

- النموذج الأول: كانت الفقرات العشرة الأولى هي الفقرات الصعبة.
- النموذج الثاني: كانت الفقرات العشرة الوسطى هي الفقرات الصعبة.
- النموذج الثالث: كانت الفقرات العشرة الأخيرة هي الفقرات الصعبة.
- النموذج الرابع: كانت الفقرات العشرة الصعبة موزعة عشوائياً.

سابعاً: تم تقدير ثبات الاتساق الداخلي للاختبار بصورته النهائية (50) فقرة باستخدام معادلة كرونباخ ألفا (Cronbach α)، وبلغت قيمته (0.84)، وقد عدّت هذه القيمة مناسبة لتحقيق أغراض هذه الدراسة.

9- جمع وتحليل البيانات:

- لتحقيق الهدف من الدراسة، قام الباحثان بتطبيق نماذج الاختبار عشوائياً على أفراد عينة الدراسة والبالغ عددهم 519 طالبا وطالبة، تغيب منهم (5) طلاب وطالبات عن الاختبار بواقع نموذج واحد لكل طالب. وبعد تطبيق نماذج الاختبار على أفراد عينة الدراسة، تم فرز كل نموذج على حده وتصحيحه، ثم أدخلت البيانات إلى ذاكرة الحاسوب، واستخدم البرنامج الإحصائي (SPSS) لإجراء التحليلات الإحصائية اللازمة للإجابة عن أسئلة الدراسة، والمتمثلة بحساب معاملات الصعوبة، ومعاملات الصعوبة المحوّلة، وتحليل التباين الأحادي، واختبار شففيه للمقارنات البعدية للسؤال الأول، و حساب معاملات التمييز (معامل ارتباط بوينت بايسيربال)، وقيم Z الفشرية المقابلة لها، وتحليل التباين الأحادي واختبار شففيه للمقارنات البعدية، وتقدير الثبات للاتساق الداخلي باستخدام معادلة كرونباخ ألفا (Cronbach α)، وتم حساب معاملات الارتباط بين الدرجات على الاختبار، والدرجة الكلية في مبحث

أثر موقع الفقرات الصعبة في اختبار اختيار من متعدد على خصائصه السيكومترية..... د. بني عطا. الشريفين

الحاسوب، واستخدام الإحصائي (M) للكشف عن دلالة الفروق في معاملات الثبات لنماذج الاختبار الأربعة، ولا يشترط تساوي المجموعات وأعداد الفقرات لكل نموذج (Hakstain and Whalen, 1976)، ويتبع هذا الإحصائي توزيع كاي تربيع (X^2) بدرجات حرية تساوي (عدد المعاملات - 1)، وتحسب قيمة M من الصيغة الرياضية الآتية:

$$M = \frac{J-1}{18J} \left[\sum_{k=1}^4 B_k - \frac{[\sum_{k=1}^4 (1 - \alpha_k)^{-1/3}]^2}{\sum_{k=1}^4 B_k (1 - \alpha_k)^{-2/3}} \right]$$

وتحسب قيمة B_k من القانون:

$$B_k = \frac{(9n_k - 11)^2}{(n_k - 1)}$$

حيث تمثل الرموز:

K: رقم النموذج، n_k : عدد أفراد المجموعة في النموذج K، J: عدد الفقرات، α_k : معامل ثبات كرونباخ ألفا للنموذج K.

. والإحصائي (V) (Hays, 1981. P. 467) للفروق بين عدّة معاملات ارتباط مستقلة والذي يتبع توزيع كاي تربيع بدرجات حرية تساوي (د=3). وتحسب قيمة V من القانون:

$$V = \sum_{j=1}^4 (n_j - 3)(z_j - u)^2$$

وتحسب قيمة u من القانون:

$$u = \frac{\sum_{j=1}^4 (n_j - 3)z_j}{\sum_{j=1}^4 (n_j - 3)}$$

حيث تمثل الرموز: J: رقم النموذج، N_j : عدد أفراد المجموعة في النموذج J، Z_j : العلامة الفشرية المقابلة لمعامل الارتباط J.

10- نتائج الدراسة:

10-1- النتائج المتعلقة بأثر موقع الفقرات الصعبة على الخصائص السيكومترية للفقرات (الصعوبة، التمييز).

10-1-1- صعوبة الفقرة

لمعرفة أثر توزيع الفقرات الصعبة في كل نموذج من نماذج الاختبار الأربعة، فقد تم حساب معاملات الصعوبة للفقرات في كل نموذج من النماذج الأربعة. وذلك بإيجاد نسبة الطلبة الذين أجابوا عن الفقرة إجابة صحيحة من بين المفحوصين الذين حاولوا الإجابة عن هذه الفقرة، كما تم حساب المتوسطات الحسابية لهذه المعاملات، والجدول 1 يبين معامل الصعوبة لكل فقرة من فقرات النماذج الأربعة بالإضافة إلى معاملات الصعوبة للفقرات محولة إلى قيم دلتا (Δ) المناظرة.

جدول 1. معاملات الصعوبة للفقرات وقيم دلتا Δ

لنماذج الأربعة والمتوسطات الحسابية لهذه المعاملات

رقم الفقرة	النموذج الأول		النموذج الثاني		النموذج الثالث		النموذج الرابع	
	معامل صعوبة الفقرة		معامل صعوبة الفقرة		معامل صعوبة الفقرة		معامل صعوبة الفقرة	
	Δ	P	Δ	P	Δ	P	Δ	P
1	18.4	0.09	15.84	0.24	15.72	0.25	14.84	0.32
2	18.6	0.08	15.84	0.24	15.52	0.26	14.92	0.31
3	17.6	0.13	15.76	0.25	15.64	0.26	14.56	0.35
4	17.9	0.11	14.52	0.35	14.32	0.37	15.44	0.27
5	18.6	0.08	14.96	0.31	15.00	0.31	16.72	0.18
6	18.9	0.07	14.8	0.33	14.72	0.33	14.00	0.40
7	17.3	0.14	15.08	0.30	15.00	0.31	14.60	0.34
8	18.4	0.09	14.60	0.34	14.32	0.37	14.20	0.38
9	17.3	0.14	14.96	0.31	14.72	0.33	14.88	0.32
10	16.9	0.17	14.08	0.39	13.92	0.41	16.32	0.20
11	15.4	0.27	14.16	0.39	13.92	0.41	13.84	0.42

14.32	0.37	14.88	0.32	14.68	0.34	15.8	0.24	12
13.64	0.44	14.08	0.40	14.08	0.39	15.8	0.24	13
13.40	0.46	13.92	0.41	14.08	0.39	14.6	0.35	14
14.60	0.34	13.04	0.50	13.16	0.48	16.1	0.28	15
16.20	0.21	13.12	0.49	13.40	0.46	14.8	0.33	16
13.32	0.47	12.96	0.50	12.92	0.51	15.2	0.29	17
13.32	0.47	13.20	0.48	12.92	0.51	14.4	0.37	18
12.72	0.53	13.76	0.43	14.08	0.39	15.8	0.25	19
13.20	0.48	12.64	0.53	13.00	0.50	14.1	0.39	20
16.72	0.18	12.44	0.56	13.64	0.12	14.1	0.39	21
17.08	0.15	13.04	0.50	17.80	0.11	14.9	0.32	22
12.60	0.54	13.68	0.43	17.04	0.16	14.1	0.39	23
12.68	0.53	12.72	0.53	17.04	0.16	14.2	0.38	24
13.56	0.45	13.60	0.44	18.36	0.09	13.1	0.49	25
12.36	0.56	13.48	0.45	18.16	0.10	13.5	0.45	26
12.16	0.58	12.64	0.53	16.80	0.17	13.1	0.49	27
12.68	0.53	12.88	0.51	18.16	0.10	13.2	0.48	28
13.20	0.48	14.00	0.40	16.40	0.20	14.0	0.40	29
12.68	0.53	13.76	0.43	15.76	0.25	13.0	0.50	30
15.52	0.26	12.72	0.53	12.96	0.54	13.1	0.49	31
12.88	0.51	13.60	0.44	13.16	0.48	13.1	0.49	32
12.88	0.51	13.52	0.45	13.84	0.42	14.1	0.39	33
12.32	0.57	13.60	0.44	12.92	0.51	13.6	0.44	34
12.72	0.53	13.20	0.48	13.56	0.44	13.6	0.44	35
13.64	0.44	13.04	0.50	13.48	0.45	13.6	0.44	36
17.08	0.15	13.28	0.47	13.00	0.50	13.6	0.44	37
13.12	0.49	13.84	0.42	13.08	0.49	13.0	0.50	38
12.08	0.59	14.56	0.35	14.24	0.38	14.2	0.38	39
12.96	0.50	14.64	0.34	13.68	0.43	13.8	0.42	40
12.88	0.51	16.56	0.19	12.76	0.52	13.2	0.48	41
13.12	0.49	17.64	0.12	13.68	0.43	14.0	0.40	42
15.84	0.24	16.56	0.19	13.56	0.44	13.6	0.44	43

12.88	0.51	16.56	0.19	13.56	0.44	13.6	0.44	44
12.72	0.53	17.80	0.12	13.00	0.50	13.7	0.43	45
12.88	0.51	17.80	0.12	13.00	0.50	14.0	0.40	46
15.00	0.31	16.32	0.20	13.40	0.46	13.3	0.47	47
13.12	0.49	17.64	0.12	13.68	0.43	13.8	0.42	48
13.84	0.42	16.24	0.21	14.44	0.36	14.6	0.35	49
13.64	0.44	15.72	0.25	14.84	0.32	15.1	0.30	50
13.92	0.42	14.43	0.37	14.52	0.36	14.87	0.34	الوسط الحسابي
1.28	0.12	1.49	0.13	1.56	0.13	1.78	0.14	الانحراف المعياري

يتضح من نتائج الجدول 1 أن قيم معاملات الصعوبة في النموذج الأول كانت هي الأدنى؛ حيث تراوحت قيم معاملات الصعوبة للفقرات الصعبة العشرة الأولى بين (0.07 - 0.17) والتي الأرقام (1 إلى 10)، أما بقية فقرات النموذج الأول فقد تراوحت معاملات صعوبتها بين (0.24 - 0.5)، وكان الوسط الحسابي لمعاملات الصعوبة للنموذج الأول (0.34).

وفي النموذج الثاني؛ تراوحت قيم معاملات الصعوبة للفقرات العشرة الصعبة (0.09 - 0.25)، وتحمل الأرقام (21 إلى 30)، وتراوحت قيم معاملات صعوبة بقية الفقرات بين (0.24 - 0.54). وكان الوسط الحسابي لمعاملات الصعوبة لهذا النموذج (0.36).

وفي النموذج الثالث حيث موقع الفقرات العشرة الصعبة في نهاية الاختبار، وتحمل الأرقام (41 إلى 50)، تراوحت قيم الفقرات الصعبة بين (0.12 - 0.25)، أما بقية فقرات الاختبار فتراوحت معاملات صعوبتها بين (0.25 - 0.56)، وكان الوسط الحسابي لمعاملات الصعوبة (0.37).

وفي النموذج الرابع تراوحت قيم معاملات الصعوبة للفقرات الصعبة (الموزعة بشكل عشوائي) والتي تحمل الأرقام (4، 5، 10، 16، 21، 22، 31، 37، 43، 47) بين (0.15 - 0.31) وقيم معاملات الصعوبة لبقية فقرات النموذج العشوائي كانت بين (0.31 - 0.60)، وكان الوسط الحسابي لمعاملات الصعوبة (0.42).

أثر موقع الفقرات الصعبة في اختبار اختيار من متعدد على خصائصه السيكومترية..... د. بني عطا. الشرفين

ويتضح كذلك من النتائج الواردة في الجدول نفسه أن أعلى متوسط ملاحظ لمعاملات الصعوبة كان للنموذج الرابع والذي تم توزيع الفقرات الصعبة فيه بشكل عشوائي، وأدنى متوسط لمعاملات الصعوبة كان للنموذج الأول.

وبالنظر إلى النتائج الواردة في الجدول 1 يظهر بأن هناك تبايناً ملحوظاً في متوسطات معاملات الصعوبة لنماذج الاختبار.

ومن أجل الكشف عن دلالة الفروق بين هذه المتوسطات والإجابة عن سؤال الدراسة الأول، تم تحويل معاملات الصعوبة للفقرات إلى قيم دلّتا (Δ) المناظرة، والتي هي علامات معيارية محولة بوسط مقداره (13)، وانحراف معياري مقداره (4)، وذلك لأن قيم دلّتا (Δ) المناظرة، تقع على مقياس فثوي، وحسب المعادلة التالية ($Z = 13 + 4\Delta$) كما في جدول 1، وتم استخدام تحليل التباين الأحادي (One Way Anova)؛ لمعرفة مدى الاختلاف بين متوسطات معاملات الصعوبة بين النماذج المختلفة، والجدول 2 يبين نتائج التحليل.

جدول 2. نتائج تحليل التباين الأحادي

لمتوسطات معاملات الصعوبة المحولة في النماذج الأربعة

الدلالة الإحصائية	قيمة الإحصائي F	متوسط مجموع المربعات	درجات الحرية	مجموع المربعات	مصدر التباين
0.025	3.192	7.79	3	23.370	بين المجموعات (النماذج)
		2.44	196	478.317	داخل المجموعات (الخطأ)
			199	501.687	الكلّي

يتضح من الجدول 2 وجود فروق ذات دلالة إحصائية بين متوسطات معاملات الصعوبة في النماذج الأربعة عند مستوى ($\alpha = 0.05$). ولتحديد مواقع الفروق ذات الدلالة الإحصائية، فقد تم

استخدام اختبار شففيه (Scheffe) للمقارنات البعدية، ويبين جدول 3 نتائج المقارنات الثنائية بين متوسطات معاملات الصعوبة للنماذج الأربعة وفق اختبار شففيه Scheffe

جدول 3. نتائج المقارنات الثنائية بين متوسطات معاملات الصعوبة المحولة للفقرات

في النماذج الأربعة تبعاً لاختبار شففيه

النموذج	الأول 14.87	الثاني 14.52	الثالث 14.42	الرابع 13.92
الأول 14.87	-			
الثاني 14.52	0.35	-		
الثالث 14.42	0.45	0.10	-	
الرابع 13.92	0.95*	0.60	0.50	-

* دالة إحصائياً عند مستوى الدلالة ($\alpha = 0.05$)

يتضح من جدول 3 أن هناك فرقاً ذا دلالة إحصائية بين متوسطات معاملات الصعوبة المحولة للفقرات، تعزى لاختلاف موقع الفقرات الصعبة بين النموذجين: الأول والرابع لصالح النموذج الرابع، بينما لم يظهر الجدول فروقاً ذات دلالة إحصائية بين متوسطات معاملات الصعوبة المحولة للفقرات، تعزى لاختلاف موقع الفقرات الصعبة بين بقية النماذج.

10-1-2- تمييز الفقرة:

تم إيجاد معاملات التمييز للفقرات في كل نموذج من النماذج الأربعة، وذلك بإيجاد معامل الارتباط بين نتائج المفحوصين على هذه الفقرة، ونتائجهم على الاختبار الكلي (Corrected Item Total Correlation). كما تم حساب المتوسطات الحسابية لهذه المعاملات، و جدول 4 يبين معاملات التمييز لكل فقرة من فقرات النماذج الأربعة، بالإضافة إلى معاملات التمييز للفقرات محولة إلى درجات فيشير الزائية (Zr).

جدول 4. معاملات التمييز للفقرات وقيم Z الفشرية

في النماذج الأربعة والمتوسطات الحسابية

النموذج الأول		النموذج الثاني		النموذج الثالث		النموذج الرابع		رقم الفقرة
معامل تمييز الفقرة		معامل تمييز الفقرة		معامل تمييز الفقرة		معامل تمييز الفقرة		
Zr	rpbis	Zr	rpbis	Zr	rpbis	Zr	rpbis	
0.61	0.55	0.45	0.42	0.39	0.37	0.69	0.59	1
0.42	0.40	0.33	0.31	0.26	0.35	0.30	0.29	2
0.81	0.70	0.58	0.52	0.56	0.51	0.35	0.34	3
1.06	0.79	0.37	0.35	0.30	0.29	0.33	0.31	4
0.54	0.50	0.47	0.44	0.47	0.43	0.44	0.42	5
0.44	0.41	0.48	0.45	0.45	0.42	0.41	0.39	6
0.45	0.42	0.36	0.35	0.28	0.27	0.28	0.27	7
0.48	0.44	0.41	0.39	0.43	0.40	0.30	0.29	8
0.34	0.33	0.38	0.36	0.22	0.22	0.88	0.70	9
0.54	0.49	0.66	0.58	0.63	0.55	0.68	0.59	10
0.45	0.42	0.44	0.42	0.41	0.39	0.60	0.53	11
0.38	0.37	0.37	0.35	0.42	0.40	0.27	0.26	12
0.80	0.67	0.41	0.38	0.25	0.24	0.52	0.48	13
0.45	0.42	0.38	0.36	0.35	0.34	0.27	0.26	14
0.44	0.41	0.33	0.32	0.32	0.31	0.38	0.37	15
0.60	0.54	0.35	0.34	0.29	0.28	0.44	0.41	16
0.34	0.33	0.36	0.35	0.35	0.34	0.21	0.21	17
0.52	0.48	0.38	0.36	0.31	0.30	0.42	0.39	18
0.36	0.35	0.36	0.34	0.23	0.22	0.24	0.24	19
0.33	0.32	0.39	0.38	0.36	0.35	0.61	0.55	20
0.58	0.52	0.35	0.34	0.82	0.67	0.45	0.42	21
0.62	0.55	0.26	0.26	0.54	0.49	0.37	0.35	22
0.39	0.37	0.22	0.22	0.45	0.43	0.30	0.29	23
0.34	0.33	0.27	0.26	0.50	0.46	0.32	0.31	24
0.39	0.37	0.22	0.22	0.46	0.43	0.30	0.30	25
0.45	0.42	0.23	0.23	0.50	0.46	0.27	0.27	26
0.42	0.40	0.35	0.34	0.38	0.33	0.32	0.31	27
0.33	0.32	0.27	0.26	0.31	0.30	0.32	0.31	28
0.37	0.35	0.38	0.37	0.77	0.65	0.22	0.22	29
0.31	0.30	0.35	0.34	0.45	0.42	0.33	0.32	30
0.53	0.48	0.22	0.22	0.34	0.33	0.35	0.34	31

0.37	0.35	0.21	0.21	0.21	0.22	0.21	0.21	32
0.35	0.34	0.23	0.23	0.20	0.20	0.36	0.35	33
0.39	0.37	0.32	0.31	0.23	0.23	0.32	0.31	34
0.37	0.36	0.34	0.33	0.22	0.22	0.23	0.23	35
0.42	0.40	0.27	0.26	0.22	0.21	0.22	0.21	36
0.52	0.48	0.28	0.27	0.34	0.33	0.51	0.47	37
0.49	0.45	0.29	0.29	0.24	0.24	0.25	0.24	38
0.41	0.39	0.27	0.26	0.34	0.33	0.35	0.33	39
0.34	0.33	0.65	0.57	0.42	0.40	0.30	0.29	40
0.39	0.34	0.81	0.67	0.30	0.29	0.28	0.27	41
0.38	0.36	0.50	0.46	0.24	0.24	0.26	0.25	42
0.85	0.69	0.53	0.49	0.30	0.29	0.23	0.22	43
0.39	0.37	0.25	0.53	0.33	0.32	0.29	0.28	44
0.38	0.36	0.53	0.48	0.38	0.36	0.50	0.46	45
0.32	0.31	0.48	0.45	0.23	0.22	0.38	0.36	46
0.60	0.54	0.37	0.35	0.40	0.38	0.30	0.30	47
0.38	0.36	0.35	0.34	0.29	0.29	0.42	0.40	48
0.47	0.43	0.78	0.65	0.52	0.48	0.28	0.27	49
0.84	0.68	0.44	0.42	0.63	0.56	0.66	0.58	50
0.48	0.43	0.39	0.37	0.38	0.36	0.37	0.35	الوسط الحسابي
0.16	0.11	0.14	0.11	0.14	0.12	0.14	0.11	الانحراف المعياري

Zr : علامة فيشر المقابلة لمعامل تمييز الفقرة في النموذج

يتضح من نتائج الجدول 4 أن قيم معاملات التمييز لفقرات النموذج الأول، تراوحت بين (0.21 - 0.70) بوسط قدره (0.35)، وتراوحت معاملات التمييز لفقرات النموذج الثاني ما بين (0.2 - 0.67) بوسط قدره (0.36)، وتراوحت معاملات التمييز لفقرات النموذج الثالث بين (0.21 - 0.67) بوسط قدره (0.37)، بينما تراوحت معاملات التمييز للنموذج الرابع بين (0.30 - 0.79) بوسط قدره (0.43). ويتضح من نفس الجدول أن أعلى متوسط معامل تمييز رقمياً، كان للنموذج الرابع ذي التوزيع العشوائي للفقرات الصعبة.

وللإجابة عن سؤال الدراسة الثاني، ومن أجل الكشف عن دلالة الفروق بين متوسطات معاملات التمييز بين النماذج المختلفة، تم تحويل قيم معاملات التمييز للفقرات إلى درجات فيشر الزائفة (Zr)، حيث يصبح شكل التوزيع طبيعياً لقيم معاملات التمييز، وتم استخدام تحليل التباين الأحادي

أثر موقع الفقرات الصعبة في اختبار اختيار من متعدد على خصائصه السيكومترية..... د. بني عطا. الشرفين

(One Way Anova)؛ للكشف عن الفروق في متوسطات معاملات التمييز بين النماذج الأربعة وجدول 5 يبين نتائج هذا التحليل.

جدول 5. نتائج تحليل التباين الأحادي لمتوسطات معاملات التمييز في النماذج الأربعة

الدلالة الإحصائية	قيمة الإحصائي F	متوسط مجموع المربعات	درجات الحرية	مجموع المربعات	مصدر التباين
0.010	3.844	0.140	3	0.421	بين المجموعات (النماذج)
		0.036	510	18.606	داخل المجموعات (الخطأ)
			513	19.027	الكلية

يتضح من جدول 5 أن هناك أثراً لموقع الفقرات الصعبة على معاملات التمييز للفقرات في النماذج الأربعة. حيث تشير قيمة الإحصائي "F" المحسوبة بدرجات حرية (3) إلى وجود فروق ذات دلالة إحصائية عند مستوى ($\alpha = 0.05$) بين متوسطات معاملات التمييز للفقرات. ولتحديد مواقع الفروق ذات الدلالة الإحصائية، فقد تم استخدام اختبار شففيه للمقارنات البعدية، ويبين جدول 6 نتائج المقارنات الثنائية بين متوسطات معاملات التمييز لفقرات النماذج الأربعة وفق اختبار شففيه.

جدول 6. نتائج المقارنات الثنائية بين متوسطات معاملات التمييز للفقرات

في النماذج الأربعة حسب اختبار شففيه

النموذج	الأول 0.37	الثاني 0.38	الثالث 0.39	الرابع 0.48
الأول 0.37	–			
الثاني 0.38	0.01	–		
الثالث 0.39	0.02	0.01	–	
الرابع 0.48	0.11*	0.10	0.09	–

* دالة إحصائية عند مستوى الدلالة ($\alpha = 0.05$)

تظهر النتائج الواردة في الجدول 6 أن هناك فرقاً ذا دلالة إحصائية بين متوسطي معاملي التمييز بين النموذجين، الأول: الفقرات العشرة الأولى هي الفقرات الصعبة والرابع (الفقرات الصعبة موزعة

عشوائياً) ولصالح النموذج الرابع عند مستوى الدلالة ($\alpha = 0.05$)، بينما لم تظهر النتائج فروقاً ذات دلالة إحصائية بين متوسطات معاملات التمييز بين بقية نماذج الاختبار.

ثانياً: النتائج المتعلقة بأثر موقع الفقرات الصعبة على الخصائص السيكومترية للاختبار (الثبات والصدق).

تم تقدير ثبات الاتساق الداخلي لكل نموذج من نماذج الاختبار الأربعة باستخدام معادلة كرونباخ ألفا (Cronbach's Alpha). وتم حساب معامل الصدق المرتبط بمحك، وذلك بحساب معامل ارتباط بيرسون بين درجات الطلاب على كل نموذج ودرجاتهم النهائية في مادة الحاسوب المقررة في الفصل الدراسي الثاني من العام الدراسي 2012/2013، وجدول 7 يبين معاملات الثبات ومعاملات الصدق المرتبط بمحك وتباين الأداء على نماذج الاختبار الأربعة.

جدول 7. معاملات الثبات ومعاملات الصدق المرتبط بمحك وتباين الأداء على نماذج الاختبار

الأربعة

نموذج الاختبار	معامل ثبات كرونباخ α	معامل الصدق المرتبط بمحك	عدد أفراد العينة
النموذج الأول	0.876	0.81	126
النموذج الثاني	0.881	0.83	122
النموذج الثالث	0.893	0.86	129
النموذج الرابع	0.925	0.89	137

يلاحظ من نتائج جدول 7 أن النموذج الرابع الذي توزعت الفقرات الصعبة فيه بشكل عشوائي، يتمتع بثبات أعلى من بقية النماذج الأخرى، يليه النموذج الثالث والذي كانت الفقرات الصعبة في نهايته، ثم النموذج الثاني والذي كان موقع الفقرات الصعبة في وسطه، وأدناها النموذج الأول حيث كان موقع الفقرات الصعبة في بدايته.

وللإجابة عن سؤال الدراسة الثالث تم اختبار الفروق بين معاملات الثبات لنماذج الاختبار الأربعة عن طريق الإحصائي (M) المقترح من قبل هاكستين وولين (Hakstain & Whalen, 1976) الذي يتبع توزيع كاي تربيع بدرجات حرية تساوي (عدد المعاملات-1). وقد كشف التحليل الإحصائي الخاص بقيم كرونباخ ألفا عن وجود فروق دالة إحصائية على مستوى الدلالة ($\alpha = 0.05$) بين معاملات ثبات كرونباخ ألفا، حيث كانت قيمة الإحصائي (M) المحسوبة (3.58)، وهي أقل قيمة

من كاي تربيع (χ^2) الحرجة (7.82) بدرجات حرية (د=3) عند مستوى الدلالة الإحصائية ($\alpha=0.05$).

ولتقدير الصدق المرتبط بمحك للاختبار، تم حسب معامل الارتباط بيرسون بين أداء الطلبة على الاختبار الذي أعده الباحثان، والعلامة النهائية للطلّاب في مادة الحاسوب (محك) للنماذج الأربعة، ويلاحظ من الجدول 8 أن قيم معاملات الارتباط بين أداء الطلبة على اختبار الحاسوب والعلامة النهائية في مادة الحاسوب، كانت جميعها موجبة، وهذا يعني أن نسبة التباين المشترك بين الاختبارين كانت عالية، حيث كانت أقل قيمة للنموذج الأول وأعلى قيمة للنموذج الرابع.

وللإجابة عن سؤال الدراسة الرابع تم اختبار دلالة الفروق بين قيم معاملات الصدق المرتبط بمحك باستخدام الإحصائي (V) لاختبار دلالة الفروق بين عدة معاملات ارتباط مستقلة، والذي يتبع توزيع كاي تربيع بدرجات حرية تساوي (د=3) عند مستوى الدلالة ($\alpha=0.05$)، وجدول 9 يبين نتائج التحليل المتعلقة بمعاملات صدق المحك.

جدول 9. نتائج التحليل المتعلقة بمعاملات صدق المحك

النموذج	معامل الصدق	قيمة V	χ^2	درجات الحرية
الأول	0.81	8.21	7.82	3
الثاني	0.83			
الثالث	0.86			
الرابع	0.89			

يلاحظ من الجدول 9 أن قيمة (V) أكبر من قيمة χ^2 ، وبالتالي توجد فروق ذات دلالة إحصائية على مستوى الدلالة ($\alpha=0.05$) بين معاملات الصدق المرتبط بمحك تعزى لموقع الفقرات الصعبة. حيث كان أعلى معامل صدق للنموذج الرابع؛ أي أن النموذج الرابع كان الأكثر دقة في قياس محتوى الاختبار.

وللكشف عن دلالة الفروق بين معاملات الصدق المرتبط بمحك، تم استخدام الاختبار Z، للعلامات الفشرية (Feldt, 1980)، ويبين جدول 10 نتائج المقارنات الثنائية بين معاملات الصدق المرتبط بمحك.

جدول 10 نتائج المقارنات الثنائية بين معاملات صدق المحك

النموذج	معامل الارتباط	عدد الأفراد لكل نموذج	Z الفشرية	قيمة Z	
				المحسوبة	الدرجة
الأول	0.81	126	1.127	0.48	1.96
الثاني	0.83	122	1.188		
الأول	0.81	126	1.127	1.31	1.96
الثالث	0.86	129	1.293		
الأول	0.81	126	1.127	2.36*	1.96
الرابع	0.89	137	1.422		
الثاني	0.83	122	1.188	0.82	1.96
الثالث	0.86	129	1.293		
الثاني	0.83	122	1.188	1.86	1.96
الرابع	0.89	137	1.422		
الثالث	0.86	129	1.472	1.04	1.96
الرابع	0.89	137	1.422		

يلاحظ من جدول 10 وجود فروق ذات دلالة إحصائية في معاملات صدق المحك بين النموذجين الأول والرابع ولصالح النموذج الرابع، بينما لا توجد فروق ذات دلالة إحصائية بين النماذج الأخرى.

ثالثاً: النتائج المتعلقة بأثر موقع الفقرات الصعبة في أداء طلبة الصف الأول الثانوي على اختبار الحاسوب.

وفيما يتعلق بأداء أفراد عينة الدراسة على الاختبار والموزعة في أربعة نماذج، وفقاً لاختلاف موقع الفقرات الصعبة، فقد تم حساب الأوساط الحسابية والانحرافات المعيارية لعلامات الطلبة على النماذج الأربعة للاختبار والجدول 11 يوضح ذلك.

جدول (11) الأوساط الحسابية والانحرافات المعيارية لعلامات الأفراد في النماذج الأربعة

النموذج	عدد أفراد العينة	الوسط الحسابي	الانحراف المعياري
الأول	126	16.96	8.70
الثاني	122	17.94	8.71
الثالث	129	18.57	9.37
الرابع	137	20.8	11.05

يتضح من النتائج الواردة في الجدول 11 بأن أعلى وسط هو (20.8) للنموذج الرابع، وأقل وسط هو (16.96) للنموذج الأول، وأكبر انحراف معياري هو (11.05) للنموذج الرابع، وأقل انحراف معياري هو (8.70) للنموذج الأول، إذ إن التباين في الأداء يزداد كلما كانت الفقرات الصعبة موزعة توزيعاً عشوائياً.

ولإجابة عن سؤال الدراسة الخامس فقد استخدم الباحثان تحليل التباين الأحادي، لمعرفة فيما إذا كانت هناك فروق ذات دلالة إحصائية، ويبين جدول 12 نتائج تحليل التباين الأحادي لأداء الطلبة على اختبار الحاسوب وفقاً لاختلاف موقع الفقرات الصعبة.

جدول (12) نتائج تحليل التباين الأحادي لاستجابات أفراد عينة الدراسة على النماذج الأربعة

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط مجموع المربعات	قيمة الإحصائي F	الدلالة الاحصائية
بين المجموعات (النماذج)	1051.794	3	350.598	3.844	.010
داخل المجموعات (الخطأ)	46515.228	510	91.206		
الكلي	47567.021	513			

يظهر من جدول 12 أن هناك تأثيراً لنموذج الاختبار المستخدم (موقع الفقرات الصعبة) في أداء المفحوصين. حيث تشير قيمة الإحصائي F إلى وجود فروق ذات دلالة إحصائية عند مستوى $\alpha = 0.05$ بين متوسطات أداء المفحوصين، ولتحديد مواقع الفروق ذات الدلالة الإحصائية، فقد تم استخدام اختبار شففيه للمقارنات البعدية، ويبين جدول 13 نتائج المقارنات البعدية بين متوسطات النماذج الأربعة وفق اختبار شففيه.

جدول 13. المقارنات البعدية بين متوسطات النماذج الأربعة وفق اختبار شففيه

النموذج	الأول 16.96	الثاني 17.94	الثالث 18.57	الرابع 20.80
الأول 16.96	-			
الثاني 17.94	.98231	-		
الثالث 18.57	1.61333	.63102	-	
الرابع 20.80	3.83530 [*]	2.85300	2.22198	-

تشير نتائج التحليل المبينة في جدول 13 إلى وجود فرق واحد ذي دلالة إحصائية بين النموذج الأول والنموذج الرابع ولصالح النموذج الرابع، حيث يلاحظ أن أداء الطلبة في النموذج الرابع أفضل من أداء الطلبة في النموذج الأول، وأما بقية النماذج فلا توجد بينها فروق ذات دلالة إحصائية.

11- مناقشة النتائج والتوصيات

11-1 مناقشة النتائج المتعلقة بأثر موقع الفقرات الصعبة على الخصائص السيكمترية للفقرة (الصعوبة والتمييز).

11-1-1-1-1 صعوبة الفقرة: أظهرت نتائج تحليل التباين الأحادي للكشف عن الفروق بين متوسطات معاملات الصعوبة للفقرات للنماذج الأربعة، وجود فروق جوهرية بين معاملات الصعوبة تعزى لموقع الفقرات؛ حيث تبين أنه عند تغيير موقع الفقرات الصعبة من بداية الاختبار إلى توزيعها بشكل عشوائي، ارتفعت قيم معاملات الصعوبة، وهي نسبة من أجاب عن الفقرة إجابة صحيحة للفقرات، لتصبح أسهل وتقرب من 0.50، وأن قيم معاملات دللتا المناظرة لها قلت. وقد أشارت نتائج اختبار شففيه للمقارنات البعدية إلى وجود فرق ذي دلالة إحصائية بين وسط قيم معاملات الصعوبة المحولة في النموذج الأول، ووسط قيم معاملات الصعوبة المحولة في النموذج الرابع، ولصالح النموذج الرابع الذي تتوزع فيه الفقرات الصعبة توزيعاً عشوائياً؛ أي أن أداء الطلبة في الاختبار كان أفضل في النموذج الموزعة فقراته الصعبة توزيعاً عشوائياً، وانسجمت هذه النتيجة مع دراسة نيومان وكونديرت ولايني وبول (Newman, Kundert, Lane, & Bull, 1988)، والتي أشارت إلى أن علامات الطلبة كانت أعلى في الفقرات الصعبة على الفقرات غير المرتبة إحصائياً (عشوائية) ومرتبة حسب الصعوبة المعرفية، واتفقت مع ما أشار إليه النبهان (2004) بأن من

العوامل التي تؤثر في صعوبة الفقرة موقعها بين الفقرات، واتفقت أيضاً مع دراسة وينستون ورويدجر (Weinstein, Roediger, 2012) حيث أشارا إلى أن صعوبة الفقرة تغيرت عبر اختلاف توزيع الفقرات.

واختلفت مع دراسة آمودت ومكشان (Aammodt & McShane, 1992) والتي أشارت نتائجها إلى اختلاف بسيط بين علامة الاختبار المرتبة فقراته (السهل - الصعب) مع الترتيب العشوائي. ويعزو الباحثان النتيجة بأنه عندما كانت الفقرات الصعبة كلها في بداية الاختبار ربما انعكس ذلك على أداء الطالب عند الاستجابة عليها من حيث التعامل معها بشكل فاعل، وربما أدى ذلك إلى شعور الطالب بالإحباط وعدم الرغبة بإنهاء الاختبار أو إنهائه بشكل غير فاعل، وذلك بسبب إخفاقه في كثير من الأحيان في الإجابة عن الفقرات الصعبة، لذا تكمن أهمية البدء بإعطاء الفقرات المألوفة أو السهلة للمفحوص من أجل رفع ثقته بنفسه، وضمان تحفيزه وإثارة دافعيته، وتخفيض مستوى القلق لديه لإنهاء الاختبار. أمّا عندما كانت الفقرات الصعبة موزعة بشكل عشوائي على الاختبار ككل، وكما هو موضح في النموذج الرابع، فقد يتم استثارة دافعية المفحوص بسبب إجابته عن بعض الفقرات الصعبة بين الحين والآخر، وهذا قد يكون متوقعاً من قبل المفحوص أو الطلبة، حيث لديهم قناعة أو معرفة بأن الاختبار لن يكون بنفس المستوى من الصعوبة، لذلك عندما يجيبون عن بعض الفقرات، ويواجهون بعض الفقرات الصعبة، لن يصلوا إلى مرحلة الإحباط التي قد يصلون إليها في حال كانت الفقرات الصعبة متتابعة.

11-1-2-تمييز الفقرة: أظهرت نتائج تحليل التباين الأحادي للكشف عن الفروق بين متوسطات معاملات التمييز للفقرات بين نماذج الاختبارات الأربعة (الفقرات الصعبة في بداية الاختبار، الفقرات الصعبة في وسط الاختبار، الفقرات الصعبة في نهاية الاختبار، الفقرات الصعبة موزعة عشوائياً) وجود فروق جوهرية بين متوسطات معاملات التمييز تعزى لاختلاف موقع الفقرات الصعبة. وكشفت نتائج اختبار شففيه للمقارنات البعدية عن وجود فرق ذي دلالة إحصائية بين النموذج الأول (الفقرات الصعبة في بداية الاختبار) والنموذج الرابع (الفقرات الصعبة موزعة عشوائياً) ولصالح النموذج الرابع، وجاءت هذه النتيجة منسجمة مع صعوبة الفقرات. واتفقت هذه النتيجة مع دراسة ميرز ومورفي وجودمان وتورهان (Meyers, Murphy, Goodman, & Turhan, 2012) والتي هدفت إلى معرفة أثر تغيير موقع الفقرات الصعبة على معالم الفقرة،

ونائج المعادلة المشتركة تحت النموذج الثلاثي المعلمة. وخلصت الدراسة إلى أن تمييز الفقرة يتغير بتغير موقع الفقرة في الاختبار، وأشار النبهان (2004) بأن تمييز الفقرة يساعد في تحديد قدرتها على التمييز بين الطلبة ذوي التحصيل المتدني والطلبة ذوي التحصيل العالي.

ويمكن تفسير هذه النتيجة على أساس أن تباين الأداء على فقرات النموذج الرابع كان أعلى من تباين الأداء على فقرات النموذج الأول، مما يدل على أن الفقرات الصعبة الموزعة عشوائياً تجعل فقرات الاختبار لها القدرة على التمييز بين المجموعات المتباينة من الطلبة، بينما كان تباين الأداء على فقرات نموذجي الاختبار الثاني والثالث قريباً نوعاً ما، كما أنه من الممكن النظر إلى أن الفقرات في النموذج الأول قد تكون سبباً في ارتفاع مستوى القلق لدى الطلبة ذوي التحصيل المرتفع، وذلك بسبب انعكاس هذا الترتيب على الحالة النفسية للطلاب، ومن المعروف بأنه مع ارتفاع القلق يتأثر التحصيل، لذلك فقد يكون مستوى معاناة الطلبة ذوي التحصيل المرتفع أو شعورهم بالقلق أعلى من غيرهم من الطلبة؛ وذلك بسبب غياب التعزيز الفوري، الذي قد يكون وقّره النموذج الرابع بسبب قدرة الطالب على الإجابة عن بعض الفقرات في بداية الاختبار، والتي قد تسمى في بعض الأحيان بالفقرات التشجيعية، وهذا قد يؤدي إلى إثارة دافعيتهم في الاستمرار بالإجابة عن بقية فقرات الاختبار. أما النموذج الأول فقد يكون لعب دوراً سلبياً من حيث شعور الطلبة بأنه قد لا تكون هنالك نتائج إيجابية لدراساتهم، حيث إن جميع الطلبة - على اختلاف مستوياتهم - قد يشعرون بالعجز عن الإجابة، وهذا قد يكون كفيلاً لخلق الإحباط لدى ذوي التحصيل المرتفع، وعاملاً مثيراً لعدم إكمالهم الاختبار بدافعية مرتفعة بسبب شعورهم بأن نتائجهم ستكون متقاربة مع غيرهم من الطلبة ذوي التحصيل المتدني، والذين لا يكونون قد بذلوا جهداً في الاستعداد للاختبار، مما يؤثر في طريقة إكمالهم الاختبار.

11-2- مناقشة النتائج المتعلقة بأثر توزيع الفقرات الصعبة على الخصائص السيكومترية للاختبار

(الصدق والثبات).

فيما يتعلق بالإجابة عن السؤال الخاص بأثر توزيع الفقرات الصعبة لاختبار الحاسوب في الاتساق الداخلي، تم تقدير قيمة معامل ثبات كرونباخ ألفا، والتي تقوم على فكرة حساب معاملات الارتباط بين الدرجات على الفقرات والدرجة الكلية في الاختبار، وقد تبين أن قيم معامل الثبات تتمتع بدرجة مرتفعة من الاتساق الداخلي؛ حيث أظهرت نتائج التحليل الإحصائي أن ثبات

الاتساق الداخلي للنموذج الأول (0.876) وللنموذج الثاني (0.881) وللنموذج الثالث (0.893) وكان أعلاها للنموذج الرابع (0.925) والذي تم توزيع الفقرات الصعبة فيه بشكل عشوائي.

ويمكن تفسير هذه النتيجة في ضوء ما أشار إليه عودة (2010) حول العوامل المؤثرة في ثبات الاختبار، والذي أشار إلى أن أبرز هذه العوامل هو تجانس المجموعة أو (تجانس الطلبة)، ويرى الباحثان بأن التجانس قد حصل بين الطلبة إلى حد ما في النموذج الأول، والذي رتبت فيه الفقرات الصعبة في البداية، مما قد يكون شجع الطلبة على التخمين العشوائي، وهذا يقود إلى أن تكون علاماتهم متقاربة إلى حد تبدو فيه مجموعة الطلبة كمجموعة متجانسة؛ أي أن الاختبار ذو قدرة تمييزية ضعيفة مقارنة مع النموذج الرابع. ومن المعروف أنه إذا كانت المجموعة متجانسة فإن ثبات الاختبار سيتأثر. أما النموذج الرابع والذي كانت فقراته موزعة بشكل عشوائي، فقد كان له قدرة تمييزية عالية، وفق ما أشارت إليه نتائج الدراسة سابقاً بمعنى أن المجموعة غير متجانسة، وبذلك فإن فرصة تأثر ثبات الاختبار تقل في هذا النموذج عن النماذج الأخرى.

وفيما يتعلق بالإجابة عن السؤال الخاص بالعلاقة الارتباطية بين أداء الطلبة على الاختبار الذي أعده الباحثان، وعلامة الطالب النهائية في مادة الحاسوب للفصل الثاني للنماذج الأربعة. فقد تم تقدير معاملات الارتباط للنماذج الأربعة فكانت معاملات الارتباط بين أداء الطلبة على الاختبار الذي أعده الباحثان وعلامة الحاسوب النهائية عالية وموجبة، وقد يعود السبب في ذلك إلى قوة العلاقة بين المتنبي وهو في هذه الدراسة (الاختبار الذي أعده الباحثان) والحك المستخدم (علامة الطالب النهائية في مادة الحاسوب للفصل الثاني)، حيث إن المتنبي والحك يقيسان نفس المحتوى ونفس السمة عند جميع الأفراد، وهي التحصيل في الحاسوب.

ومن المعلوم أن قيمة معامل الصدق المرتبط بمحك تعتمد على مجموعة من العوامل التي تتعلق بالحك والمتنبي، فارتفاع قيمة الثبات سواء أكان للمحك أم للمتنبي أم كليهما يزيد من قيمة معامل الارتباط بينهما، فقد كان معامل الارتباط بين أداء الطلبة على الاختبار وعلامة الحاسوب النهائية أعلى ما يمكن للنموذج الرابع والذي كان يتمتع بدرجة أعلى من الثبات مقارنة مع النماذج الأخرى. ويمكن أن يعزى الارتفاع في قيم معاملات الصدق المرتبط بمحك إلى جدية الطلبة أثناء تطبيق الاختبار، وتدوين أسمائهم وأسماء مدارسهم حتى تسنى للباحث معرفة علاماتهم النهائية في مبحث الحاسوب (أي على

الحك). ويعني ذلك أن النموذج الرابع: (كانت الفقرات العشرة المتدنية الصعوبة موزعة عشوائياً) أبرز قدرة تنبؤية أكبر بعلامة الطالب في الحاسوب مقارنة بنماذج الاختبار الأخرى.

11-3- مناقشة النتائج المتعلقة بأثر موقع الفقرات الصعبة في أداء طلبة الصف الأول الثانوي على اختبار الحاسوب.

أما بخصوص الإجابة عن السؤال الرابع الخاص بأثر توزيع الفقرات الصعبة لاختبار الحاسوب في أداء الطلبة، تم حساب الأوساط الحسابية والانحرافات المعيارية لعلامات الطلبة على النماذج الأربعة. وأشارت النتائج إلى أن الوسط الحسابي للنموذج الرابع كان الأعلى من بين النماذج الأربعة، وكذلك أكبر انحراف معياري. وأشارت نتائج تحليل التباين الأحادي للكشف عن الفروق بين المتوسطات بأنه كان دالاً إحصائياً، ولمعرفة لصالح أي نموذج كانت الدلالة الإحصائية؛ تم استخدام اختبار شفهي للمقارنات البعدية للكشف عن الفروق إلى وجود فرق واحد ذي دلالة إحصائية بين النموذج الأول والنموذج الرابع لصالح النموذج الرابع؛ أي أن وسط درجات الاختبار قد ارتفع عند توزيع الفقرات بشكل عشوائي، واختلفت هذه النتيجة مع دراسة بلاش (Blach, 1989) والتي حاولت معرفة هل يحصل المفحوصين على درجات عالية عندما تكون الفقرات لاختبار تحصيلي في اختبار الاختيار من متعدد مرتبة بشكل متسلسل حسب صعوبتها أعلى من الدرجات التي يحصلون عليها عندما تكون فقرات الاختبار الاختيار من متعدد موزعة عشوائياً، حيث أشارت نتائج الدراسة بأنه لا يوجد دليل على الأداء المتفوق في أي من أنماط الترتيب: حسب الصعوبة، عشوائي، الموازنة بين العشوائي وتسلسل حسب الصعوبة.

واتفقت هذه النتيجة مع دراسة نيلى وسبرنجستون وميكان (Neely, Springston, & McCann, 1994) والتي هدفت لمعرفة هل ترتيب فقرات الاختيار من متعدد تجعل الاختبار أكثر تحدياً للطلبة؟ وأشارت نتيجة هذه الدراسة أن أعلى قلق كان في الترتيب المتسلسل، وكان أعلى من القلق في التوزيع العشوائي. وقد تفسر هذه النتيجة في ضوء الخصائص النفسية للمفحوصين حيث أشار عدد من الطلبة بعد الالتقاء بهم وتحديداً الذين استجابوا على النموذج الأول إلى أن مستوى القلق لديهم ارتفع، وأصبحوا أقل تركيزاً في عملية الإجابة؛ لأنهم كانوا عندما يفشلون في بعض الأحيان في الإجابة على أي فقرة ثم ينتقلون إلى الفقرة التالية، كان لديهم اعتقاد بأن الفقرات التالية ستكون أسهل، وعندها يفاجؤون بأن الفقرة التالية كذلك صعبة فإنهم شعروا بالإحباط، لأنهم لن يستطيعوا الحصول على الدرجات التي يتوقعون، وهذا قد يكون سبباً في انخفاض دافعتهم؛ وبالتالي حصولهم على درجات منخفضة رغم أن

فقرات الاختبار فيما بعد كانت ضمن مستوى الصعوبة المتفق عليه إحصائياً، إلا أنه قد يكون وصول الطلبة إلى هذه الفقرات متأخراً وغير مجدٍ بالنسبة لهم؛ حيث إن هناك العديد من الفقرات المتتالية التي واجهتهم بصعوبة مرتفعة، وبالتالي غير متأكدين من إجاباتهم عليها.

أما في النموذج الرابع فإن الطلبة قد يكونون عند استجابتهم قد تجاوزوا الفقرة الصعبة عند عدم القدرة على الإجابة عنها إلى الفقرة التي تليها، والتي قد تكون أسهل من سابقتها، الأمر الذي قد يكون فكرة لدى الطلبة بأن فقرات الاختبار ليست بنفس مستوى الفقرة الصعبة. وبالتالي انخفاض مستوى القلق عند توزيع الفقرات بشكل عشوائي واستمرار الدافعية، حيث المجال أكبر للحصول على إجابات صحيحة. بمعنى أن الطلبة يحصلون على تعزيز متقطع، وهذا النوع من التعزيز يزيد من استمرارية محاولة الطلبة في الإجابة المتتالية للفقرات، نتيجة زيادة ثقة الطلبة بأنفسهم، الأمر الذي قد يكون له أثر إيجابي في تحسن مستوى التركيز لديهم، وبالتالي تحسن أداء الطلبة، وارتفاع درجاتهم وزيادة الوسط الحسابي في النموذج العشوائي.

12- الخلاصة والتوصيات:

لقد بينت النتائج الخاصة أثر موقع الفقرات الصعبة على الخصائص السيكمومترية للفقرة (الصعوبة والتمييز) إلى وجود فرق ذي دلالة إحصائية ($\alpha = 0.05$) بين صعوبة الفقرات في النموذج الأول والنموذج الرابع، وعدم وجود فروق ذات دلالة إحصائية بين بقية النماذج.

وكذلك أشارت النتائج إلى وجود فرق ذي دلالة إحصائية ($\alpha = 0.05$) بين تمييز الفقرات في النموذج الأول والنموذج الرابع ولصالح النموذج الرابع، وعدم وجود فروق ذات دلالة إحصائية بين بقية النماذج.

كما أشارت النتائج الخاصة بأثر موقع الفقرات على الخصائص السيكمومترية للاختبار (الصدق والثبات) فكان النموذج الرابع له أعلى قيمة معامل ثبات، بينما كانت أدنى قيمة للنموذج الأول.

وأشارت النتائج إلى وجود فروق جوهرية بين معاملات صدق نموذجي الاختبار، بين النموذج الأول الذي تم وضع الفقرات الصعبة في بدايته، والنموذج الرابع الذي تم توزيع الفقرات الصعبة فيه بشكل عشوائي.

ويعني ذلك أن أفضل موقع للفقرات الصعبة في الاختبار أن توزع عشوائياً وألا تكون متتالية، سواءً في بداية الاختبار أم وسطه أم نهايته، حيث إن الخصائص السيكمومترية للاختبار تتأثر عند توزيع الفقرات الصعبة بشكل عشوائي، ومن ناحية أخرى فإن توزيع الفقرات الصعبة بشكل عشوائي يوفر الوقت والجهد الذي يبذله المعلمون في بناء هذا النمط من الفقرات.

وهذا الاستنتاج يقدم الدعم لتوزيع الفقرات الصعبة بشكل عشوائي، حيث إن وجود الفقرات الصعبة في اختبار بشكل متتالي قد يسبب كثيراً من التشتت والإرباك، مقارنة بتوزيع تلك الفقرات عشوائياً عند الإستجابة عليها. وتعارضت هذه النتيجة مع وجهة النظر التي تدعم وضع فقرات سهلة في بداية الاختبار لتشجيع الطلبة على الاستمرار في الإجابة (عودة، 2010). ومن ناحية أخرى فإن توزيع فقرات الاختبار بشكل عشوائي، يوفر الوقت والجهد الذي يبذله المعلمون في بناء اختبار تكون فيه الفقرات الصعبة مرتبة. كما أن بناء فقرات الاختبار من متعدد بتوزيع عشوائي للفقرات الصعبة، يسمح بتغطية أكبر قدر ممكن من محتوى المادة الدراسية المراد معرفة مدى إتقان التلاميذ لها في زمن اختباري قصير.

ويعني هذا أن نموذج الاختبار الذي توزع فيه الفقرات الصعبة بشكل عشوائي أبرز قدرة تنبؤية بعلامة الطالب في الحاسوب مقارنة ببقية النماذج. وبناءً على ذلك يمكن لنموذج الاختبار الذي توزعت الفقرات الصعبة فيه بشكل عشوائي أن يكون الأفضل من حيث الخصائص السيكمومترية للاختبار (الصدق والثبات)، ومن حيث الخصائص السيكمومترية للفقرة (الصعوبة والتمييز).

وهذا الاستنتاج يعزز دور توزيع الفقرات الصعبة بشكل عشوائي في اختبار الاختيار من متعدد في قياس نواتج التعلم في المستويات العقلية العليا من المجال المعرفي، واعتماداً على ذلك فإن الباحثان يوصيان بما يأتي:

- للعاملين في مجال التربية والتعليم والذين يستخدمون الاختبارات: بناء اختبارات الاختيار من متعدد لغرض قياس تحصيل الطلبة في المواد الدراسية سواء في الاختبارات الفصلية أم النهائية، توزيع الفقرات الصعبة (معامل صعوبة متدني) بشكل عشوائي بين فقرات الاختبار.
- للباحثين إجراء دراسة مماثلة تعتمد عدد أقل من البدائل في اختبار اختيار من متعدد، وتضم كمّاً أكبر من الفقرات في إطار مسابقات دراسية أخرى بهدف تدعيم نتائج البحث الحالي، بما يكفل إمكانية تعميمها على نطاق واسع.
- إعادة الدراسة بتطبيق نظرية الاستجابة للفقرة.

المراجع

المراجع العربية

- الشريفين, علي عبد الله. (2013). أثر موقع الفقرات الصعبة في اختبار اختيار من متعدد على خصائصه السيكمومترية وأداء الطلبة عليه. رسالة ماجستير غير منشورة, الأردن, اريد: جامعة اليرموك.
- العمري، حسان. (2011). أثر ترتيب الفقرات في المقاييس النفسية على التجانس الداخلي للمقياس وخصائص فقراته. مجلة جامعة دمشق، 27 (4+3)، 111 - 129.
- عودة، أحمد. (2010). القياس والتقويم في العملية التدريسية. إريد: دار الأمل للنشر والتوزيع.
- كامل، تغريد. (2009). إعادة ترتيب بعض فقرات اختبار ستانفورد - بينيه (الصورة الرابعة) علي تغيير مستوي أداء الأطفال دراسة للفروق تبعاً لبعض متغيرات الشخصية والديموجرافية. رسالة ماجستير، القاهرة: جامعة عين شمس.
- النبهان، موسى. (2004). أساسيات القياس في العلوم السلوكية. عمان: دار الشروق للنشر والتوزيع.

المراجع الأجنبية

- Aamodt, G., & McShance, T.(1992).A **Meta-Analytic investigation of the ffect of various test item characteristics on test score and test score and test completion times** . *Puplic Personnel Mangement*. 21(12), 151-160.
- Aiken, L .92003).***Psychological Testing and Assessment*** (11th ed.).Boston:Pearson Education Group.
- Anastasi, A. (2000). *Psychological Testing*. New York. Prentice – Hell and latent class models in the social sciences (pp. 89-98). Munster, Germany: Waxmann
- Ben-Simon, A., Bedescu, D., & Nevo B. (1997). A comparativestudy of measures of partial Knowledge in multiple choice test applied. *Psychological Measurement*, 22 (1), 65-88.
- Blach, W. (1989). Item order affects performance on multiple-choice exams. *Teaching of Psychology*, 16 (2), 75-77
- Blunch, N. J. (1984). Positional Bias in Multiple-Choice Questions. *Journal of Marketing Research*, 21, 216-220.
- Brown, G. (1983). *Principles of Educational and Psychological Testing*. New York: Rinehart and Winsern.

- David, K. (2001). Item Order Response Location and Examination Sex and Handeness and Performance on Multiple-Choice Test. *The Journal of Educational Research*, 73(4), 225-229.
- Eble, R.L, and Frisbie, D.A, (1986) Essential of Educational Measurement . Fourth edition, New Jersey: Prentice-Hall.
- Gronlund, N., & Linn, R. (1990). *Measurement and Evaluation Teaching* (6th ed.). New York: Macmillan Publishing Company.
- Hakstain, A., & Whalen, E. (1976). A K-Sample significance Test for Independent Alpha Coefficients. *Psychometrika*, 41(2), 219-231.
- Hays, W. (1981), *Statistics*. (3rd ed.), p467.
- McMillan, J. H. (2004). *Classroom Assessment: Principles and for effective Instruction*. 3rd ed. United states of America.
- Meyers, J., Murphy, S., Goodman, J., & Turhan, A. (2012). The Impact of Item Position Change on Item Parameters and Common Equating Results under the 3PL Model. *Annual Meeting of the National Council on Measurement in Education*, Vancouver: Canada, 1 – 44.
- Neely. D., Springston, F., & McCann. S. (1994). Does item

order affect performance on multiple-choice exams?.

Teaching of Psychology, 21(1), 44 - 46 .

- Newman, L., Kundert, K., Lane, S., & Bull, K. (1988).
Effect of Varying Item Order on Multiple-Choice Test
Scores: Importance of Statistical and Cognitive Difficulty.
Applied Measurement in Education, 1(1), 89-97.
- Nitko, A. J. (2001). *Educational Assessment of Students* (3rd
Edition). New Jersey: Prentice-Hall, Inc.
- Weinstein, Y., & Roediger, L. (2012). The effect of
question order on evaluations of test performance: how does
the bias evolve?. *Psychonomic Society*, 40(1), 727 - 735.
- Weiten, W. (1984). Violation of selected item construction
principles in educational

<< وصل هذا البحث إلى المجلة بتاريخ 2015/8/9، وصدرت الموافقة على نشره بتاريخ 2015/12/30 >>