



**تأثير اختلاف ترتيب مفردات الاختبار التحصيلي مرجعي المحك
على معاملات الصعوبة والتمييز لمفردات الاختبار والأداء
الأكاديمي وفق نظرية الاستجابة للمفردة لدى طلاب كلية التربية**

إعداد

د. محمد مصطفى عليوة

مدرس علم النفس التربوي
كلية التربية - جامعة الزقازيق

د. هانم أحمد سالم

مدرس علم النفس التربوي
كلية التربية - جامعة الزقازيق

تأثير اختلاف ترتيب مفردات الاختبار التحصيلي مرجعي المحك على معاملات الصعوبة والتمييز لمفردات الاختبار والأداء الأكاديمي وفق نظرية الاستجابة للمفردة لدى طلاب كلية التربية

إعداد

د. محمد مصطفى عليوة

مدرس علم النفس التربوي
كلية التربية - جامعة الزقازيق

د. هانم أحمد سالم

مدرس علم النفس التربوي
كلية التربية - جامعة الزقازيق

مستخلص البحث باللغة العربية

هدف البحث إلى معرفة تأثير اختلاف توزيع مفردات الاختبار المرجعي المحك والمعد في ضوء المحتوى المعرفي إما بشكل (مرتب بتسلسل)، أو (مرتب عشوائياً) على الأداء الأكاديمي ومعاملتي الصعوبة والتمييز للاختبار ومفرداته؛ حيث تم حساب دقة تقديرات قدرات الطلاب، والخصائص السيكومترية للمفردات والاختبار وفق النموذج ثنائي البارامتر (المعلم) في نظرية الاستجابة للمفردة. ولتحقيق هدف البحث تم إعداد اختبار تحصيلي مرجعي المحك في مقرر "الفروق الفردية والقياس النفسي" لدى طلبة الجامعة من الدبلوم العام في التربية نظام العام الواحد (تفرغ) بكلية التربية-جامعة الزقازيق للعام الجامعي ٢٠١٧/٢٠١٨م، وتم تصحيح الاختبار آلياً عبر جهاز الإسكانر؛ حيث تم الكشف عن تأثير اختلاف توزيع المفردات في نوعين من الاختبارات الموضوعية: (الاختبار من متعدد (MCQ)، و(الصواب والخطأ (T/F)). وتكون الاختبار في صورته المبدئية من (٦٠) مفردة من نوع الاختيار من متعدد ذي الأربعة بدائل (أ، ب، ج، د)، و(٤٠) مفردة من نوع الصواب والخطأ (X، ✓)، وُزعت في نموذجين (أ، ب)، حيث توزعت المفردات بطريقة مرتبة بتسلسل طبقاً للمحتوى العلمي المعرفي في النموذج (أ) وكان عدد طلاب هذا النموذج (٤٥١)، وبطريقة مرتبة عشوائياً في النموذج (ب) وعدد طلاب هذا النموذج (٤٠٧)، وبذلك يكون لدينا عدد متقارب للطلاب في كل نموذج. وتكونت عينة البحث الاستطلاعية من (٥٣٩) طالباً وطالبة، أستخدمت العينة الاستطلاعية في حساب صدق وثبات الاختبار، وكذلك حساب درجة القطع. وتكونت عينة البحث النهائية التي تم اختيارها بالطريقة العشوائية من (٨٥٨) طالباً وطالبة بعد حساب درجة القطع. ثم استخدمت المعالجات الإحصائية المناسبة للتأكد من مدى مطابقة بيانات العينة النهائية لافتراضات نظرية الاستجابة للمفردة والمتمثلة في أحادية البعد، والاستقلال الموضوعي، جودة مطابقة البيانات، وتم حساب الخصائص السيكومترية للاختبار، وتقديرات القدرة للأفراد، ومعالم المفردات والخطأ المعياري لكل منها. تم حساب هذه البارامترات وفقاً للنموذج ثنائي المعلم أي في معاملات الصعوبة والتمييز والخطأ المعياري في تقدير كل منها، كذلك تم بحث تأثير نوع النموذج على الأداء الأكاديمي، وقد أظهرت أهم نتائج البحث النهائية ما يلي: ١- لا توجد فروق ذات دلالة إحصائية عند مستوى (α=0.05) وذلك في تقديرات معاملات الصعوبة والتمييز للمفردات وأخطاءهما المعيارية ترجع للمحتوى: (متسلسل معرفياً، أو عشوائياً) وفق النموذج ثنائي البارامتر، ٢- توجد فروق في معامل الثبات الإمبريقي للاختبار وفق النموذج ثنائي البارامتر، وكذلك، ٣- توجد فروق ذات دلالة إحصائية بين نموذجي الاختبار (أب) في الأداء التحصيلي الأكاديمي لصالح نموذج (أ) المرتب بتسلسل معرفي.

مقدمة البحث:

تُعد الاختبارات التحصيلية وبالأخص الاختبارات مرجعية المحك من أهم الوسائل التقويمية التي يستخدمها المعلم أياً كانت مراحل التعليم، وخاصة التعليم الجامعي، فإذا أحسن المعلم بنائها واستخدامها تصبح قادرة على قياس الأهداف المرجوة كإمام الطالب بالمعلومات والمهارات في المقررات الدراسية، وتلك الاختبارات لها اعتبارات عملية يجب مراعاتها مثل سهولة التطبيق وسهولة التصحيح وتفسير الدرجة، بالإضافة إلى الخصائص التجريبية مثل الصدق والثبات، والخصائص العامة مثل الشمول والموضوعية.

ويشير (صلاح مراد، وأمين سليمان، ٢٠٠٢، ص ١٨) إلى أن الاختبارات التحصيلية إذا أحسن المعلم بناءها واستخدامها تكون عوناً له وتلاميذه وصابغى القرار على السواء حيث يستخدمها المعلم في التعرف على المستوى التحصيلي الذي وصل إليه تلاميذه من تحسن أو تأخر. وبالتالي مراقبة العملية التعليمية، وبالنسبة للطالب فإن معرفته نتائج الاختبارات يعمل على تعزيز سلوكه ورفع مستوى الطموح لدى الطالب الناجح ووجود التنافس بين الطلاب ومساعدة الطالب الضعيف على تحسين أوضاعه وتجاوز أخطائه، وبالنسبة لصابغى القرار فالاختبارات تزوده بقراءة واضحة تقريباً عن سير العملية التعليمية من خلال نتائجها وما يمكن اتخاذه من قرارات بناء على هذه النتائج.

وتتعدد أشكال المفردات التي يمكن استخدامها في تقويم تحصيل الطلاب منها ما يحتاج إنتاج الإجابة منها الاختبارات المقالية التي تحتوي على عدد قليل من الأسئلة مثل المقال الطويل أو المقال القصير ويقوم الطالب باسترجاع وسرد المعلومات الموجودة في ذاكرته، وهناك اختبارات تحتاج إلى انتقاء الإجابة منها اختبارات الاختيار من متعدد، والصواب والخطأ، حيث إن اختبارات الاختيار من متعدد عبارة عن جزئيين عبارة تقديمية (جذع السؤال) وتحمل في طياتها قضية محددة تحتاج إلى تفسيرها أو إكمالها ليكتمل معناها ثم يلي هذه العبارة التقديمية مجموعة من البدائل أو المشتتات ويطلب من الطالب اختيار أصح البدائل منها، بينما أسئلة الصواب والخطأ تتكون من مجموعة من العبارات ويطلب من الطالب أن يقرأ كل عبارة ويضع علامة (✓) أمام العبارة الصحيحة أو علامة (×) أمام العبارة الخاطئة وهذه الأسئلة تقيس مدى استيعاب الطلاب للحقائق العلمية والجغرافية والتاريخية واللغوية والنظريات والمفاهيم العلمية.

وقام (Carson & Ostrosky, 1992) بإعداد أربعة اختبارات يتكون كل منها من نموذجين في مقرر مبادئ الإدارة وتم ترتيب الفقرات في النموذج الأول تسلسلياً تبعاً للمحتوى،

والنموذج الثاني عشوائياً وأظهرت النتائج وجود فروق دالة إحصائية في أداء الطلاب لصالح النموذج الأول.

وأعد Neely, Springston & McCann (1994) ثلاثة اختبارات يتكون كل منها من نموذجين، وتم ترتيب المفردات في النموذج الأول من السهل إلى الصعب، والنموذج الثاني عشوائياً وبعد تطبيق الاختبارات على (٢٥٣) طالباً جامعياً، وأظهرت النتائج أنه لا توجد فروق دالة إحصائية في تحصيل الطلاب ترجع إلى ترتيب المفردات.

وتوصل Barnett et al (2007) إلى وجود تأثير لترتيب المفردات في الاختبار وطريقة عرضها على استجابات الافراد. وبحث Sue(2009) تأثير ترتيب مفردات الاختبار على أداء الطلاب في مقرر مبادئ الإدارة وتم تقديم نموذجين للطلاب مع ترتيب المفردات في النموذج الأول تسلسلياً تبعاً للمحتوى، والنموذج الثاني عشوائياً، وأسفرت النتائج أنه لا توجد فروق دالة إحصائية في درجات الطلاب ترجع إلى ترتيب المفردات. وأشارت إلى إمكانية المعلم إعداد عدة نماذج من الاختبار بإعادة ترتيب المفردات دون حدوث تغير في الدرجات أو الوقوع في عملية التحيز.

بينما طبق Doerer & Calhoun(2009) ثلاث نماذج اختبار من نوع الاختيار من متعدد على (١٢٩٦) طالباً في مقرر الاقتصاد وتم ترتيب المفردات طبقاً للمحتوى (تصاعدي أي من السهل إلى الصعب، وتنازلي من الصعب إلى السهل، وعشوائي) على الترتيب، وتوصلت النتائج إلى وجود فروق دالة إحصائية في تحصيل الطلاب ترجع إلى ترتيب المفردات لصالح النموذج الأول التصاعدي.

في حين توصل Pettijohn, & Sacco(2007) بعد تطبيق نفس الثلاث نماذج السابقة (تصاعدي بمعنى من السهل إلى الصعب، وتنازلي من الصعب إلى السهل، وعشوائي) على الترتيب في مقرر مبادئ علم النفس وأظهرت النتائج أنه لا توجد فروق دالة إحصائية في درجات الطلاب ترجع إلى نموذج الاختبار وترتيب مفرداته.

وبالتالي فإن الطريقة التي يتم بها ترتيب مفردات في الاختبار قد تؤثر على الأداء التحصيلي الأكاديمي للطلاب سواء كان ترتيب المفردات من السهل إلى الصعب أو من الصعب إلى السهل أو ترتيبها حلزوني أو متسلسل أو عشوائي. ولذا يهتم البحث الحالي بتحليل مفردات الاختبار التحصيلي ومعرفة معاملات الصعوبة ومعامل التمييز وتحديد مدى فاعلية البدائل

وبالتالي يتم تحديد مواطن القوة والقصور لدى الطلاب، بالإضافة إلى بحث تأثير ترتيب المفردات طبقاً للمحتوى المعرفي في ضوء نموذجين (متسلسل، عشوائي) على الأداء التحصيلي الأكاديمي للطلاب والخصائص السيكومترية للاختبار ومفرداته.

حيث إنه من خلال معايشة الباحثين للطلاب وُجدت شكوى من ترتيب مفردات الاختبار وأيهما أفضل في الأداء الأكاديمي نتيجة تغير ترتيب المفردات بشكل متسلسل للمحتوى التدريسي أم ترتيبها عشوائياً. كما أنه يوجد تناقض في نتائج البحوث السابقة بخصوص هذا الشأن، بالإضافة إلى التعرف على الخصائص السيكومترية للاختبار والمفردات. ولذا تم بناء اختبار تحصيلي مرجعي المحك في مادة الفروق الفردية والقياس النفسي لطلاب الدبلوم العامة في التربية (التفرغ) وتم ترتيب المفردات بطريقتين طبقاً للمحتوى المعرفي (متسلسل، عشوائي). ولم يستخدم الباحثان النظرية التقليدية لقصورها في تقدير الخصائص السيكومترية للاختبار ومفرداته ولهذا يتم استخدام نظرية الاستجابة للمفردة في تحليل إجابات الطلاب على نموذجي الاختبار لأن هذه النظرية تقيس قدرات الطلاب متحررة من صعوبة المفردة وكذلك تقديرات صعوبة المفردة تكون متحررة من قدرات الطلاب.

مشكلة البحث:

تحدد مشكلة البحث في التساؤلات الآتية:

س١: هل تختلف معاملات الصعوبة للاختبار ومفرداته باختلاف ترتيب المفردات طبقاً لتسلسل المحتوى معرفياً أو عدم تسلسله (عشوائيته) في ضوء النموذج ثنائي المَعْلَم لدى طلاب الدبلومة العام في التربية ؟

س٢: تختلف معاملات التمييز للاختبار ومفرداته باختلاف ترتيب المفردات طبقاً لتسلسل المحتوى معرفياً أو عدم تسلسله (عشوائيته) في ضوء النموذج ثنائي المَعْلَم لدى طلاب الدبلومة العام في التربية ؟

س٣: هل يختلف الأداء الأكاديمي لطلاب الدبلومة العام في التربية باختلاف نوع النموذج (متسلسل معرفياً أم مرتب عشوائياً)؟

أهداف البحث:

يهدف البحث إلى معرفة:

- ١- أفضل ترتيب مفردات للاختبار طبقا للتسلسل المعرفي أو العشوائية والذي من شأنه أنه قد يؤثر على الأداء الأكاديمي للطلاب ومعاملات الصعوبة والتميز للاختبار ومفرداته.
- ٢- أفضل الطرق التي تساعد المعلم على ترتيب مفردات الاختبار التحصيلي ليحصل طلابه على أنسب الدرجات، لتحقيق العدالة والإنصاف في التقويم.
- ٣- الطرق الملائمة لترتيب المفردات لتحقيق الخصائص السيكومترية الجيدة للاختبار ككل ومفرداته.
- ٤- رسم المسار الصحيح للقائمين على عمليات التقويم من المعلمين والأساتذة والخبراء فيم يتعلق بجودة إعداد الاختبارات.

أهمية البحث:

يهتم البحث الحالي بـ:

- ١- قياس مستوى تحصيل الطلاب ومدى تحقق الأهداف السلوكية من خلال قياس الأداء الأكاديمي ومدى اتقان الطلاب للمحتوى المعرفي عن طريق تحديد درجة القطع.
- ٢- تحسين الأداء المستقبلي للطلاب من خلال ترتيب مفردات الاختبار.
- ٣- اعتبار البحث الحالي أحد الخطوات المهمة في مسار التطوير للمناهج والمقررات الدراسية فيم يتعلق بتقويم الامتحانات.
- ٤- تحقيق جانب العدالة من خلال قياس مستوى تحصيل الطلاب بدقة واستقلالية عن قدراتهم ومدى تحقق الأهداف السلوكية من خلال اتقان الطلاب للمحتوى المعرفي عن طريق تحديد درجة القطع اتقان (٦٠%) من المحتوى.
- ٥- تحسين طرق وأساليب تقويم الامتحانات للطلاب بالاعتماد على طريقة حديثة في القياس.
- ٦- الرفع من ثبات وصدق الاختبارات التحصيلية عن طريق ترتيب مفرداتها طبقا للمحتوى المعرفي.

حدود البحث:

تتشكل هذه الحدود في:

- الحدود المكانية: كلية التربية جامعة الزقازيق.
- الحدود الزمنية: العام الدراسي (٢٠١٧/٢٠١٨م) الفصل الدراسي الثاني.
- الحدود البشرية: طلاب الدبلومة العام في التربية.
- الحدود المعرفية: محتوى مقرر الفروق الفردية والقياس النفسي.

مصطلحات البحث:

- **الاختبار مرجعي المحك:** هو مجموعة من المفردات التي تقيس فئة من الأهداف ترتبط بعدد من الكفايات المتضمنة الموجودة في سلوك محدد، وتفسر الدرجة في هذا النوع من الاختبارات في ضوء درجة القطع التي تميز بين المتقين وغير المتقين لتلك الكفايات (أبو المجد الشوربجي، عزت حسن، ٢٠٠٥، ص ٥٦١)، أما التعريف الاجرائي للاختبار التحصيلي المرجع إلى محك هو الاختبار المتعلق بمحتوي مادة الفروق الفردية والقياس النفسي ويرتبط بمجموعة من الأهداف محددة تحديداً دقيقاً وفقاً لدرجة القطع Cut-score بطريقة أنجوف المعدلة.
- **اختبار الاختيار من متعدد:** التعريف الاجرائي هو مجموعة المفردات أو الأسئلة التي تقيس تحصيل الطلاب في مقرر الفروق الفردية والقياس النفسي ويتضمن عبارة تقديمية كراس (جذع) يليها أربعة بدائل وعلى الطالب اختيار البديل الصحيح.
- **اختبار الصواب والخطأ:** التعريف الاجرائي يتضمن مجموعة من المفردات أو العبارات في مقرر الفروق الفردية والقياس النفسي وعلى الطالب أن يقرأ كل عبارة ويضع علامة (✓) أمام العبارة الصحيحة أو علامة (×) أمام العبارة الخاطئة.
- **النموذج المتسلسل معرفياً في ترتيب المفردات:** التعريف الاجرائي هو ترتيب المفردات وتسلسلها في ورقة الاختبار حسب تسلسل منطقي لمحتوى المادة الدراسية التي تم تدريسها.
- **النموذج العشوائي في ترتيب المفردات:** التعريف الاجرائي هو ترتيب المفردات وتسلسلها في ورقة الاختبار بشكل عشوائي بعيداً عن تسلسلها في المادة الدراسية التي تم تدريسها.
- **الأداء الأكاديمي:** التعريف الاجرائي يقصد به مجموع الدرجات التي يحصل عليها الطالب في اختبار مقرر الفروق الفردية والقياس النفسي.
- **معامل الصعوبة:** يُقصد به كما وضح (Lee,2007,8) أنه النقطة التي على متصل السمنة الكامنة ويتوقع عندها احتمالية إجابة الطالب عن المفردة إجابة صحيحة دون تخمين يساوي (٠,٥) من خلال المعادلة $(1+c)/2$. لأن التخمين (c) صفر.
- **معامل التمييز:** يُقصد به كما عرفه (Reeve,2002,3) بأنه ميل المماس لمنحنى خاصية المفردة عند النقطة التي تحدد صعوبتها أي عن النقطة التي يكون فيها احتمالية إجابة الطالب عن المفردة إجابة صحيحة دون تخمين يساوي (٠,٥).

▪ النموذج ثنائي البارامتر (المعلم) Two - Parameter: وهو أحد نماذج نظرية الاستجابة للمفردة ويعتمد على بارامترتي صعوبة المفردة، وتمييزها ومن أمثلتها نموذج لورد. ويسمح هذا النموذج للمنحنيات المميزة للمفردة بالتقاطع (صلاح الدين علام، ٢٠٠٥، ص ٢٠٠).

الإطار النظري:

يعتبر التحصيل الدراسي محور اهتمام كثير من المربين والعلماء على مستوى العالم، من حيث اهتمامهم بمدى اكتساب الطلاب للمعلومات وتعلمهم محتوى دراسي معين، وتستخدم الاختبارات التحصيلية لقياس التحصيل الدراسي وتظهر في عدة أنواع مختلفة.

وصنف (أحمد عودة، ٢٠٠٥، ص ١٢٠) الاختبارات التحصيلية وفقا لمعايير مختلفة أحدهما تُصنف طبقاً لطريقة الإجابة فهناك الاختبارات ذات الإجابة المستتجة (إنتاج الإجابة) اختبارات المقال، والاختبارات ذات الإجابة المنتقاة (التعرف على الإجابة)، ويتوقف ذلك على مستوى الهدف المراد قياسه، والعمر الزمني للمتعلم، والغرض من الاختبار وعدد الطلاب وظروف التطبيق ومهارة المعلم في صياغة الأسئلة.

وبالتالي فإن الاختبارات التحريرية وسيلة تقويم لا غني عنها في العملية التعليمية، حيث إنها تساعد على التقييم الفعلي لأداء الطالب وذلك من خلال نوعين من الاختبارات وهي الاختبارات المقالية والاختبارات الموضوعية، إلا أن في العصر الحديث بدأ الاهتمام أكثر بالاختبارات الموضوعية، وهناك قواعد عامة في بناء الأسئلة والمفردات الاختبارية.

ويذكر (صلاح مراد، وأمين سليمان، ٢٠٠٢، ص ١٥٢) أن الهدف من التعرف على الأنواع المختلفة لمفردات الاختبار هو إيجاد الألفة لدى الطالب والمعلم ووضعي الاختبارات بأنواع المفردات هذه، وأن يكون مصمم مفردات الاختبار لديه المهارة في صياغتها والتدريب عليها والتعرف على المواصفات الفنية في كل نوع من أنواع المفردات على حده حتى يتمكن من عمل بنك أسئلة جيد الصياغة يستخدمه فيما بعد في إعداد الاختبار حسب الهدف من الاختبار لدى المعلم أو المؤسسات التعليمية أو متخذي القرار التربوي.

ويوضح (صلاح الدين علام، ٢٠٠٩، ص ص ٩٣-٩٦) أنه يجب على المعلم قبل أن يبدأ في بناء الأسئلة والمفردات الاختبارية المناسبة لقياس الأهداف التعليمية سواء أكانت مقالية أو موضوعية التي حددها في جدول التقويم الصفي مراعاة بعض القواعد الأساسية ومنها: ملائمة الأسئلة للأهداف التعليمية المراد قياسها. إعداد مخطط التقويم عند كتابة مفردات

الاختبار والاسترشاد به. ملائمة الأسئلة لمستوى الطلاب المختبرين. ملائمة مستوى صعوبة الأسئلة للغرض من التقويم. تجنب غموض الكلمات في الأسئلة. إعداد مفاتيح التصحيح الذي يساعد المعلم على تقدير الدرجات بموضوعية بناء على عدد المفردات ومدى تمثيل عينة المفردات للأهداف التعليمية.

وتوجد أنواع عديدة للاختبارات الموضوعية منها أسئلة الاختيار من متعدد، وأسئلة الصواب والخطأ، والمزاوجة، والتكملة، والاستجابة الحرة، وإعادة الترتيب. ويعد الاختيار من متعدد من أكثر أشكال الفقرات شيوعاً واستخداماً في الاختبارات التحصيلية، وذلك بسبب كفاءته في قياس التحصيل الدراسي بشكل أكثر ثباتاً وصدقاً. وسوف يتم التركيز على أسئلة الاختيار من متعدد، حيث إن هذا النوع من الأسئلة أكثر الأنواع مرونة واستخداماً. وتشمل مفردات الاختيار من متعدد على عبارة تقديمية تتميز بالوضوح في صياغتها ويفهمها جميع الطلاب بنفس المعنى مع مراعاة العمر الزمني والعقلي المعرفي للطلاب ثم تتبعها قائمة من الإجابات المقترحة (إجابتين أو أكثر) ويختار الطالب الإجابة الصحيحة من بينها، وتسمى هذه الإجابات المقترحة " البدائل " أو " الاستجابات ". وتقوم العبارة التقديمية بعرض المهمة الواجب أدائها أو السؤال المراد إجابته، أما البدائل فإنها تشتمل على إجابة واحدة صحيحة أو أفضل إجابة، وبقيّة البدائل تسمى المشتتات أو المموهات، ووظيفة هذه المشتتات تقديم إجابات أو حلول تبدو مقبولة ظاهرياً للطالب غير المستنكر جيداً وتحاول إبعاد تفكير الطلاب عن الإجابة الصحيحة. وتوجد أنواع عديدة لمفردات الاختيار من متعدد كما يوضحها (Gronlund & Linn, 1990) ومنها:

- ١- مفردات تتطلب إجابة صحيحة واحدة: وفيها يختار الطالب إجابة صحيحة واحدة من بين البدائل المعطاة وهذا النوع هو الشائع في الاختبارات التحصيلية وأبسطها.
- ٢- مفردات تتطلب اختيار أفضل البدائل: وفيها تحتتمل بعض المشكلات أحيانا حلولاً أو إجابات مختلفة ولكن توجد من بينها إجابة أفضل يمكن أن يختارها الطالب ومثال على ذلك كأن يقدم الطالب أفضل استنتاج من بيانات معطاة أو أفضل أسلوب لأداء عمل معين، أو أفضل تطبيق لقاعدة أو مبدأ ما. وهذا النوع أكثر صعوبة في إجابته من النوع الأول.

٣- مفردات تتطلب أكثر من إجابة صحيحة واحدة: هناك نوع من المفردات يتم بناؤه بحيث تكون الإجابة عليه أكثر من إجابة صحيحة ولكن لا يفضل استخدام هذا النوع من المفردات لأن المفردات تكون أقرب في الشبه إلى مفردات صواب أو خطأ.

٤- مفردات تتطلب إكمال فراغات باستخدام بدائل معطاة ويستخدم هذا النوع بكثرة في اللغات والحساب.

٥- مفردات تبدأ بعبارة تقديمية على صيغة سؤال وفيها يتم صياغة العبارة التقديمية في صورة سؤال مباشر إذا تتطلب الأمر ذلك.

٦- مفردات تبدأ بعبارة تمهيدية منفية حيث إنه توجد بعض المفردات يستخدم معها عبارات منفية، وينبغي إن يُلفت نظر الطالب إلى المطلوب منه بوضع خط تحت الكلمة المنفية في العبارة التقديمية.

ويقدم ("صلاح الدين" علام، ٢٠٠٩، ص ص ١٠٤-١٠٦) مجموعة من المقترحات يجب مراعاتها عند بناء أنواع مفردات الاختيار من متعدد، حيث إنه لا بد من مراعاة المستوى المعرفي للهدف التعليمي المراد قياسه، بالإضافة إلى مراعاة مقترحات عند كتابة مفردات الاختيار من متعددة والتي تتلخص في:

١- مقترحات لكتابة العبارة التقديمية ويتلخص ذلك في أن المشكلة أو القضية الرئيسة يجب أن تكون واضحة ومحددة ومكتملة ومركزة وخالية من الكلمات الزائدة وتجنب الغموض في الصياغة.

٢- مقترحات لكتابة بدائل الإجابات، حيث يجب أن تكون البدائل جميعها متجانسة في محتواها، كما تكون جميع المشتتات صالحة إلى حد ما كإجابات محتملة للمفردة، ولا يوجد تداخل بين البدائل، تجنب أن تكون البدائل آراء ذاتية، تجنب استخدام "جميع ما سبق" أو "ليس مما سبق". أن تكون البدائل كافية وعلى نفس الدرجة من الطول أو القصر، استخدام الأخطاء العشوائية عند صياغة البدائل، صياغة البدائل بلغة بسيطة وسهل فهمها.

٣- مقترحات لكتابة الإجابة الصحيحة حيث يجب أن يكون للمفردة إجابة صحيحة واحدة، ويجب تجنب الترابط اللغوي بين العبارة التقديمية والإجابة الصحيحة، وتجنب جعل

الإجابة الصحيحة مختلفة عن المشتتات، يجب تغيير وضع الإجابة الصحيحة بين البدائل بطريقة عشوائية قدر الإمكان.

ويشير (أحمد عودة، ٢٠٠٥) إلى أن أسئلة الاختيار من متعدد تصنف في ثلاث مستويات معرفية وهي الحفظ، والتفسير أو التعليل، وحل المشكلات وهذا يتعلق بمهارات المعلم في كتابة السؤال ومهارات الطالب في القراءة. ويجب أن تكون المموهات (البدائل) جذابة للطلاب، ومن الأفضل ألا يقل نسبة عدد الطلاب الذين اختاروا المموه عن (٥٠٪) من الطلاب ككل، ويتم حساب جاذبية المموه من إيجاد نسبة الذين اختاروا المموه من الفئة العليا مطروحاً منه نسبة الذين اختاروا المموه من الفئة الدنيا. والمموه هو إجابة خاطئة فمن البديهي أن يكون عدد الطلبة الذين يختارون أي منها في الفئة العليا أقل ممن في الفئة الدنيا. ويلخص المعايير التي يجب مراعاتها للحصول على مموهات جذابة للطلاب ومنها ألا تكون البدائل قصيرة جداً قدر الإمكان، أن تكون المموهات من الأخطاء الشائعة ومألوفة لدى الطلاب، ألا تختلف البدائل في أطوالها، أن تكون المموهات متجانسة في المحتوى مع بعضها البعض، أن يكون عدد الإجابات الصحيحة موزعة بالتساوي وعشوائياً على كل البدائل ألا يحتوي الاختبار على فقرات تعتمد في إجابتها على فقرات أخرى بنفس الاختبار، تجنب بعض الألفاظ الخاصة مثل دائماً، كل، أبداً.

وهناك شكل آخر من مفردات الاختيار من متعدد وهو أسئلة الصواب والخطأ والتي تعتبر أبسط أنواعه حيث يتطلب وجود بديلين (نعم أو لا) أو (صواب أم خطأ) وتستخدم أسئلة الصواب والخطأ في قياس قدرة الطلاب على تمييز التعريفات الصحيحة والحقائق البسيطة والمفاهيم العلمية.

ويشير (على خطاب، ٢٠٠١، ص ٢٦٧) إلى أن مفردات الصواب والخطأ هي اختبارات شائعة الاستخدام وتتضمن عدد من العبارات التقريرية (المفردات) ويوضع أمام كل عبارة منها بديلين للإجابة (صح أم خطأ) أو كلمة (نعم أم لا) والطالب يحكم على صحة أو خطأ كل عبارة بوضع علامة أمام نعم أو لا.

ويقدم (صلاح الدين علام، ٢٠٠٩، ص ١٠٨) مجموعة من المقترحات لبناء مفردات الصواب والخطأ منها: أن تشمل المفردة فكرة أساسية واحدة تُصاغ بطريقة محددة واضحة لا تحتمل أكثر من إجابة. تجنب العبارات المنفية قدر الإمكان. عدم نقل العبارات مباشرة من الكتب المدرسية. تكون جميع العبارات متساوية في الطول

تقريباً. استخدام عدد متساوي تقريباً من العبارات الصحيحة والعبارات الخاطئة. -
تجنب استخدام كلمات مثل "دائماً"، "أبداً"، "مطلقاً"، "كل".

ويرى (أحمد عودة، ٢٠٠٥) أن الاختبار وسيلة جيدة للتعلم، لأن الطالب يعتبر الاختبار موقفاً تعليمياً جديداً ويسترجع فيه المعلومات السابقة لهذا لا بد من ضرورة ترتيب فقرات الاختبار وفقاً لعدة طرق منها:

١- ترتيب المفردات حسب نوعها: فإذا احتوى الاختبار على أكثر من نوع من أنواع المفردات، فمن الأفضل ترتيب المفردات طبقاً للنوع الواحد وهذا يسهل على المعلم توضيح التعليمات والتصحيح وكذلك المحافظة على التهيؤ العقلي للطلاب أثناء الاختبار.

٢- ترتيب المفردات حسب صعوبتها حيث يتم ترتيب المفردات من السهل إلى الصعب أو العكس.

٣- ترتيب المفردات حسب المحتوى وهو ترتيب المفردات في تسلسل طبقاً لتسلسل المادة الدراسية التي تم تقديمها للطلاب أو العكس أو الترتيب العشوائي.

٤- ترتيب المفردات حسب المستوى العقلي حيث يتم ترتيب المفردات ذات المستوى الواحد من الصعوبة لتظهر كمجموعة واحدة.

ويعتمد الباحثان في البحث الحالي على نموذجين للاختبار التحصيلي من نوع الاختيار من متعدد والصواب والخطأ مرتبة مفرداتهما طبقاً للمحتوى المعرفي إلى النموذج المتسلسل والنموذج العشوائي ومحاولة دراسة أثر ذلك على الأداء التحصيلي الأكاديمي للطلاب. ويتم ذلك من خلال إعداد اختبار تحصيلي المُرَجَّع إلى محك (مرجعية المحك).

ويوضح (عادل خضر، ٢٠٠٧، ص ص ٥٩-٦٠) أن الاختبارات التحصيلية مرجعية المحك هي تلك الاختبارات التي تهتم بقياس أداء الطلاب في ضوء مستوى التمكن المطلوب من الأداء ومدى تمكن الطالب من مجموعة الأهداف التعليمية السلوكية المرتبطة بعينة من السلوك وتصمم تلك الاختبارات بناء على مجموعة من نواتج التعلم المحددة تحديداً جيداً ضمن المحتوى التعليمي للطلاب.

ويضيف (على خطاب، ٢٠٠١، ص ٢٩٤) أن مفردات الاختبارات مرجعية المحك تتميز بانخفاض معامل تمييزها لأن الهدف من تلك الاختبارات ليس تحديد الفروق الفردية بين الطلاب ولكن تحديد مدى وصول المتعلم إلى مستوى الاتقان للمعرفة أو المهارة المطلوبة. ولكي يكون الاختبار التحصيلي جيداً لا بد من توافر مجموعة من الخصائص السيكومترية مثل الصدق والثبات ومعاملات الصعوبة والتمييز للمفردات وهذا في ضوء نظرية الاستجابة للمفردة. وهنا لا بد من التفرقة البسيطة بين نظرية الاستجابة للمفردة والنظرية التقليدية للقياس.

١- **نظرية القياس التقليدية:** يطلق على هذه النظرية نظرية الدرجة الحقيقية والتي يستخدمها الباحث بهدف الحصول على الدرجة الحقيقية للمفحوص في مجال سلوكي معين ويُعبر عنها من خلال المعادلة: الدرجة الحقيقية = درجة المفحوص في أداة القياس \pm درجة الخطأ (أبو المجد الشوربجي، وعزت حسن، ٢٠١٢، ص ص ٨٣-٨٤).

وتذكر (سامية شحاتة، ٢٠١٢، ص ٣٣) أن النظرية التقليدية تقوم على أربع مسلمات وهي أن: أداء الفرد يمكن قياسه، أداء الفرد هو دالة لخصائصه، مبدأ الفروق الفردية، القياس الظاهري الكلي يتكون من قياس حقيقي وآخر يرجع إلى الخطأ.

ولهذا توجد عدة مشكلات في تقدير الطلاب طبقاً لنظرية القياس التقليدية منها تأثير درجة الطالب بمفردات الاختبار وتأثر خصائص المفردة بقدرة الطلاب المستجيبين وتأثر الثبات بموقف الاختبار. وظهرت النظرية الحديثة لقياس درجة الطالب في معزل من تأثير قدرته أو خصائص المفردة.

٢- **نظرية الاستجابة للمفردة** تعد هذه النظرية من النظريات السيكومترية المعاصرة التي اتضحت فائدتها في التغلب على كثير من مشكلات القياس التقليدية، حيث إنها تحاول نمذجة العلاقة بين مستوى السمة المقاسة واستجابة الطالب على مفردات الاختبار (صلاح الدين علام، ٢٠٠٥، ص ٥٣).

ويذكر (أبو المجد الشوربجي، وعزت حسن، ٢٠١٢، ص ١٢٦) أن نظرية السمات الكامنة (نظرية الاستجابة للمفردة) تقوم على مجموعة من المسلمات منها: الخطأ المعياري

للقياس يختلف باختلاف الدرجات أو أنماط الاستجابة، يمكن الحصول على تقديرات غير متحيزة لخصائص مفردات الاختبار.

ولخص (أمين سليمان، ورجاء أبو علام، ٢٠١١، ص ٨٤) أهم الفروق بين نظريتي القياس التقليدية والاستجابة للمفردة كما في الجدول التالي:

جدول (١)

الفروق بين نظريتي القياس التقليدية والاستجابة للمفردة

النظرية الحديثة/ نظرية الاستجابة للمفردة IRT	النظرية التقليدية/ الكلاسيكية CTT
القدرة/ السمة الكامنة لدى المشاركين تظهر في ارتفاع عدد الأفراد الذين يجيبون على مفردة ما إجابة صحيحة	سهولة المفردة تظهر في ارتفاع عدد الطلاب الذين يجيبون عليها إجابة صحيحة والعكس صحيح.
تفصل بين خصائص الاختبار والخصائص الكامنة للمفحوصين، أي أن الدرجة المرتفعة في الاختبار تعكس قدرة مرتفعة، والعكس صحيح.	لا تفصل بين خصائص الاختبار والخصائص الكامنة لدى المفحوصين
ثبات الاختبار لا يرتبط بعدد مفردات الاختبار، ولكن بجودة مفرداته، فقد يكون الاختبار قصير ويتمتع بدرجة ثبات عالية بسبب جودة مفرداته.	ثبات الاختبار مرتبط بعدد مفردات الاختبار، فكلما زاد عدد المفردات زاد ثبات الاختبار والعكس صحيح
لا تفترض هذه النظرية والنماذج التي تندرج تحتها التوزيع الاعتمالي لدرجات الطلاب.	توزيع الدرجات يأخذ شكل التوزيع الاعتمالي وهذا التوزيع مرتبط بعينة الطلاب التي طُبق عليها الاختبار من حيث حجم العينة، أو نوعية الطلاب.
يعتمد تفسير نتائج الاختبار على مميزات مشتركة بين قدرة الطلاب، وصعوبة المفردات، وبالتالي فإن تفسير الدرجة يرجع إلى المفردة وليس إلى الجماعة المرجعية.	يعتمد تفسير نتائج الاختبار على موقع درجة الطالب بالنسبة لمتوسط أداء الجماعة. والدرجة لا تشير إلى مستوى القدرة الواقعية للطالب وإنما تشير إلى ترتيبه.
بالإمكان مقارنة نتائج اختبار لدى مجموعات مختلفة من المفحوصين، وأيضاً لدى أشكال مختلفة من الاختبارات بسبب عدم تغيير بارامترات القدرة والمفردة.	نتائج الاختبار غير قابلة للمقارنة من عينة إلى أخرى، ولا من صيغة اختبار إلى صيغة اختبار أخرى.
تسمح بالتنبؤ بكيفية أداء المفحوصين على مفردات الاختبار	لا تسمح بالتنبؤ بكيفية أداء المفحوصين على مفردات الاختبار.
الخطأ المعياري يكون لكل طالب بشكل منفرد ومستقل عن أفراد العينة ولكل مستوى معين من القدرة.	تفترض هذه النظرية أن الخطأ المعياري ثابت بالنسبة للجماعة التي طُبق عليها الاختبار والخطأ المعياري ينطبق على كل الدرجات في المجتمع المعطى.

ويتبنى الباحثان مفهوم "نماذج نظرية الاستجابة للمفردة" التي هي نظرية السمات الكامنة وتساعد على تعميم وتطوير الاختبارات التحصيلية مرجعية المحك والتي تعتبر ذات أهمية كبيرة في تطوير النظم التعليمية؛ نظراً لأنها تتطلب عناية فائقة ودقة بالغة في تحديد النطاقات السلوكية للمعارف والمهارات التي تتضمنها الكفايات التعليمية، كما تتطلب استخراج أساليب دقيقة ومناسبة في تصميمها وانتقاء مفرداتها وكذلك في تحليل وتفسير نتائجها، وذلك للتعرف على مقدار صعوبة مفرداتها وقدرتها على التمييز بين الأفراد المتمكنين وغير المتمكنين من الأهداف التعليمية التي وضعت لقياسها.

وتقوم نظرية الاستجابة للمفردة على عدة افتراضات أهمها:

أحادية البعد، المنحني المميزة للمفردة، الاستقلال الموضوعي (استقلال القياس)، التحرر من السرعة، وفيما يلي عرضاً مفصلاً لتلك الافتراضات.

1- أحادية البعد:

يذكر (Cunningham, 1986,P.158) أن أحادية البعد تعني أن كل مفردات الاختبار لها نفس الصفة أو السمة الكامنة ويتم ذلك من التحليل العاملي ويتضح ذلك في اختبارات الذكاء واختبارات الاستعداد الأكاديمي، ويشكل وجود أحادية البعد مشكلة كبيرة في اختبارات التحصيل، حيث يتم انتقاء المفردات طبقاً لجدول المواصفات، ومن الأفضل أن تكون مرتبطة مع الدرجة الكلية. ويرى كل من (Lord & Stocking, 1990, P. 281)، (Yen & Edwardson, 1999, P. 236) أن هذا الافتراض مقبول، فمعظم الاختبارات تبني لقياس سمة واحدة، وعلي سبيل المثال اختبار القدرة اللفظية أو الاختبارات الخاصة بالورقة والقلم تتطلب أن يكون لدى الطالب قدرة على القراءة للإجابة على مفردات الاختبار. وقدم Owidha, 2007,pp. (28-31) بناء على رأى (Bejar) طريقة بديلة لطريقة التحليل العاملي للاختبار أحادي البعد، وتستند هذه الطريقة إلى معاملات الارتباط البينية لتقديرات القدرة الناتجة من الاختبارات ذات التركيبات المتعدد للمحتوى، وتبين أفضلية هذه الطرق عن طريقة التحليل العاملي.

٣- المنحنيات المميزة للمفردات:

تفترض نماذج الاستجابة للمفردة وجود دالة مميزة خاصة بكل مفردة يتخذ كل منها شكل منحني الترجيح اللوغاريتمي الاحتمال الذي يسمى "المنحني المميز للمفردة" وشكل هذا المنحني يوضح كيفية تغير مستوى السمة في علاقتها بالتغيرات في احتمالات إجابة معينة فإذا

كانت المفردة ثنائية الدرجة كما في مفردات الاختيار من متعدد، فإن المنحني المميز للمفردة يعبر عن "انحدار احتمال الإجابة الصحيحة على "مستوى السمة" ("صلاح الدين" علام، ٢٠٠٥، ص ٢٩).

ويذكر (Yen & Edwardson, 1999, P. 236) أن المنحنيات المميزة للمفردة هي دوال رياضية تصف شكل العلاقة بين احتمال نجاح الطالب في الإجابة على مفردة ما والقدرة التي تقيسها مجموعة المفردات التي يشتمل عليها الاختبار، ويتم ذلك من خلال دالة الاستجابة للمفردة. ويضيف (Baker, 2001, P.7) أن المنحنيات المميزة للمفردات تتخذ شكل المنحني التجميعي (التراكمي) الطبيعي، ويعتبر تمثيلاً جيداً للعلاقة بين احتمال الإجابة الصحيحة ومستويات القدرة موضوع القياس.

٣- الاستقلال الموضوعي:

وهو افتراض موازٍ لافتراض أحادية البعد، ويعرفه (Henson, 1999, P. 14) أن استجابات الطالب لمفردات الاختبار مستقلة استقلالاً إحصائياً، وهذا يعني ألا تؤثر استجابة الطالب لإحدى مفردات الاختبار على استجابته للمفردات الأخرى. ويذكر ("صلاح الدين" علام، ٢٠٠٥، ص ٦٣) أن الاستقلال الموضوعي هو احتمال الإجابة الصواب للطالب على مفردة اختبارية يكون مستقلاً عن ناتج إجابته على أي مفردة أخرى بالاختبار في حالة ضبط كل من القيمة التقديرية لقدرته، والقيمة التقديرية لصعوبة المفردة؛ لذلك فإن الاستقلال الموضوعي يشير إلى أن مفردات الاختبار تكون غير مرتبطة، وربما يبدو هذا متعارضاً مع أحد مبادئ نظرية القياس الكلاسيكية التي تؤكد أهمية الاتساق الداخلي للاختبارات، أي كلما كانت درجات المفردات مرتبطة زاد ثبات هذه الدرجات. ويذكر (Dawning, 2003, P. 741) أن افتراض الاستقلالية يعني ببساطة أن إحدى المفردات لا يكون لها تأثير على احتمالية وصول الطلاب إلى الإجابة الصحيحة في ضوء المفردات الأخرى.

وتشير (أمينة كاظم، ١٩٨٨، ص ٤٤١) إلى أن تتحقق استقلالية القياس هذه عندما مناسبة المفردات المكونة للاختبار والنموذج، وهذا يعتمد على صدق المفردات في قياس القدرة موضوع

القياس، وفي صدق تدرجها على مقياس هذه القدرة. بالإضافة إلى ملاءمة الأفراد الذين يجيبون على مفردات الاختبار للنموذج، وهذا يعتمد على صدق استجابات هؤلاء الأفراد.

ومع هذا يتبين أن الاستقلال الوصفي يُعد افتراضاً ربما يصعب تحقيقه في كثير من الاختبارات وبخاصة الاختبارات العقلية. نظراً لتداخل هذا الافتراض مع افتراض أحادية البعد، يمكن التحقق مع ملاءمة افتراض الاستقلال لأي مجموعة من البيانات الاختبارية باستخدام أساليب التحليل العاملي ("صلاح الدين" علام، ٢٠٠٥، ص ٦٤).

٤- التحرر من السرعة:

أحد الافتراضات التي تشترك فيها جميع نماذج السمات الكامنة الشائعة الاستخدام، ويقصد بها مدى اعتماد درجة الفرد في الاختبار على أدائه فيه، وتقاس سرعة الأداء في الاختبار عادة بنسبة الأفراد الذين لم يصلوا أو لم يجيبوا عن آخر مفردة في الاختبار (فاروق عبد الفتاح، ١٩٩٠، ص ٦١٤).

ويذكر (Hambleton & Swaminathan, 1989, P. 30) أن نظرية الاستجابة للمفردة تقترض أن عامل السرعة لا يلعب دوراً في الإجابة عن مفردات الاختبار بمعنى أن إخفاق الطلاب في إجابة مفردات الاختبار يرجع إلى انخفاض قدراته، وليس إلى تأثير عامل السرعة على استجاباتهم، ويمكن تقدير ما إذا كان عامل السرعة قد لعب دوراً في الإجابة، عن طريق معرفة عدد الطلاب الذين لم يتمكنوا من الإجابة عن جميع مفردات الاختبار الذي أجرى عليهم وعندما تكون السرعة أحد العوامل المؤثرة في الأداء على الاختبار، فإن هناك قدرتين تؤثران في هذا الأداء وهما القدرة التي يقيسها الاختبار، والقدرة الخاصة لسرعة الأداء، وهذا انتهاك لافتراض أحادية البعد.

ويرى الباحثان أنه يمكن تحقيق هذا الافتراض بإعطاء الطلاب الوقت الكافي للإجابة على مفردات الاختبار وذلك لضبط عامل السرعة وكذلك التغلب على عنصر التخمين على المفردات التي لم يستطيعوا الوصول إليها. كما يمكن تقدير الدور الذي يلعبه عامل السرعة في الإجابة عن طريق حصر عدد الطلاب الذين فشلوا في الانتهاء من إجابة جميع مفردات الاختبار الذي أجرى عليهم.

وهناك نماذج مختلفة لنظرية الاستجابة للمفردة والتي تعتمد بشكل أساسي على البيانات ثنائية الاستجابة (١، ٠) التي تعنى (نجاح، فشل) على الترتيب وتصنف هذه النماذج إلى ثلاثة أنواع هي: نموذج أحادي البارامتر، ونموذج ثنائي البارامتر، ونموذج ثلاثي البارامتر لتقدير بارامترات المفردات ويمكن توضيح ذلك فيما يلي:

١- النموذج أحادي البارامتر (نموذج راش) (One Parameter Logistic Model (Rash model)

هو أحد وأشهر النماذج الرياضية الاحتمالية المرتبطة بنظرية الاستجابة للمفردة، ويعتبر أن السمة المقاسة أحادية البعد وحقق هذا النموذج نتائج مرضية في مجال تحقيق موضوعية القياس السلوكي (McCamy,2002,p.4). ويذكر (Pei-Huachen,2005,p.10) أن نموذج "راش" يركز على مَعْلَم (بارامتر) واحد هو صعوبة المفردة وموقعها بالنسبة لصعوبة جميع المفردات، كما يفترض ذلك النموذج تساوى جميع مفردات الاختبار في التمييز بين مستويات القدرة المراد قياسها وأن الطالب لا يلجأ إلى التخمين.

٢- النموذج ثنائي البارامتر (2PL) Two Parameter Logistic Model

يفترض (Lee,2007,pp7-8) أن النموذج ثنائي البارامتر (المَعْلَم) يحتوى على مَعْلَمين هما صعوبة المفردة، وتمييز المفردة، كما أن الطالب لا يلجأ إلى التخمين مثل نموذج "بيرنبوم". وسوف يتبنى الباحثان النموذج ثنائي المَعْلَم لأنه أكثر ملائمة لعنوان البحث الحالي.

٣- النموذج ثلاثي البارامتر (3PL) Three Parameter Logistic Model

يشير (Lee,2007,p.10) إلى أن النموذج ثلاثي البارامتر (المَعْلَم) امتداد للنموذج اللوغاريتمى ثنائي البارامتر (المَعْلَم)، حيث يمكن إضافة بارامتر ثالث هو معامل التخمين إلى مَعْلَمين صعوبة المفردة، وتمييز المفردة. ويرى (صلاح الدين علام، ٢٠٠١، ص ٢١٣) أهمية إضافة بارامتر التخمين في الاعتبار حيث إن بعض مفردات الاختبار تسمح أحيانا للأفراد منخفضي القدرة من الإجابة الصحيحة عليها من خلال التخمين وبالأخص في الاختبارات ذات الاختيار من متعدد.

وبالتالي زاد الاهتمام بنظرية الاستجابة للمفردة في الآونة الأخيرة لأنه في ضوء نماذجها المختلفة التي تعتمد على قدرة الطالب ومَعْلَم المفردات (معامل التخمين - معامل التمييز - معامل الصعوبة) منها نموذج راش أحادي المَعْلَم أو النماذج ثنائية المَعْلَم مثل نموذج بيرنبوم والذي أضاف بارامتر التمييز إلى صعوبة المفردة أو ثلاثية المَعْلَم مثل نموذج هامبليتون

واستخدام البارامترات الثلاثة (معامل التخمين - معامل التمييز - معامل الصعوبة) يتم تدرج المفردات ثم بناء بنوك الأسئلة التي يسعى القائمون على نظم التقويم لامتحانات إلى تعميمها في الاختبارات التحصيلية بالمراحل الدراسية المختلفة في كافة قطاعات التعليم.

البحوث السابقة:

يشتمل هذا الجزء على البحوث ذات الصلة بترتيب المفردات في الاختبار وتأثير ذلك على الأداء الأكاديمي والخصائص السيكومترية للاختبار ومفرداته كما يلي:

كشف بحث سناء عباينة (٢٠١٧) عن تأثير ترتيب مفردات اختبار الاختيار من متعدد طبقاً لعدد خطوات الحل على الخصائص السيكومترية للاختبار ومفرداته وأداء الطلاب عليه وفقاً للنظرية التقليدية في القياس. وتم بناء اختبار تحصيلي لطلاب الصف التاسع الأساسي لمادة الرياضيات في وحدة الأسس النسبية. وتكونت العينة من (١٧٩٨) طالباً وطالبة من طلاب الصف التاسع الأساسي. كما تكون الاختبار من (٦٤) مفردة من نوع الاختبار من متعدد ولكل مفردة أربعة بدائل وتم توزيع المفردات في ثلاثة نماذج مختلفة (تصاعدي، تنازلي، عشوائي). وأظهرت النتائج أنه توجد فروق دالة إحصائية بين النماذج الثلاثة في معاملات التمييز للمفردة. ولكنه لا توجد فروق دالة إحصائية بين النماذج الثلاثة في معاملات الصدق والثبات للاختبار ككل، كما أنه لا توجد فروق دالة إحصائية بين النماذج الثلاثة في متوسط أداء الطلاب.

وهدف بحث خالد سليمان (٢٠١٥) إلى التعرف على أثر ترتيب فقرات (مفردات) اختبار تحصيلي على معاملات تمييز وصعوبة الفقرات، وعلى صدق وثبات الاختبار وفقاً لنظرية القياس التقليدية، ولتحقيق أهداف البحث تم إعداد اختبار مكون من (٣٠) مفردة في مقرر التقويم التربوي لطلبة البكالوريوس في جامعة تبوك، ورتبت مفرداته على النحو الآتي: اختبار مفرداته مرتبة حسب مستوى الهدف الذي تقيسه المفردة وفق تصنيف بلوم للأهداف (تذكر، استيعاب، تطبيق، ..)، واختبار مفرداته مرتبة حسب تسلسل المحتوى، واختبار مفرداته مرتبة تصاعدياً حسب قيم معاملات صعوبة المفردات. طبقت أنماط الاختبار الثلاثة على عينة مكونة من (٣٠٠) طالب من المسجلين في المقرر الدراسي بالفصل الأول. ودلت النتائج على ما يلي: لا توجد فروق دالة إحصائية في المتوسطات الحسابية لمعاملات صعوبة مفردات الاختبار تعزى لاختلاف نمط ترتيب مفرداته. وتوجد فروق دالة إحصائية بين المتوسطات الحسابية

لمعاملات تمييز المفردات تعزى لنموذج ترتيب المفردات لصالح الاختبار الذي رتبت مفرداته تصاعدياً حسب قيم الصعوبة، ولا توجد فروق دالة إحصائية بين معاملات ثبات وصدق الاختبار تعزى لنموذج ترتيب مفرداته.

ودرس بحث فريال محمود (٢٠١٥) أثر ترتيب مفردات الاختبار ومستوى القلق على أداء الطلاب، ولتحقيق أهداف البحث تكونت العينة من (٤٥٤) طالبة من جامعة أم القرى، وتم إعداد اختبار نهاية الفصل الدراسي في مادة مدخل إلى علم النفس من نوع الاختيار من متعدد مكون من (٦٠) مفردة وعمل ست نماذج مختلفة تم ترتيب مفرداتهم على حسب درجة الصعوبة (تصاعدي، تنازلي، عشوائي)، وعلى حسب المحتوى المعرفي (تسلسلي، عكسي، عشوائي). وباستخدام الأساليب الإحصائية أسفرت النتائج عن أنه لا توجد فروق دالة إحصائية بين النماذج الثلاثة حسب درجة الصعوبة في متوسط أداء الطالبات، ولكنه توجد فروق دالة إحصائية بين النماذج الثلاثة حسب المحتوى المعرفي في متوسط أداء الطالبات. وكان أفضل أداء للطالبات على النموذجين الرابع والخامس (تسلسلي، عكسي) وأصعب النماذج هو النموذج العشوائي.

واهتمت هناء الخرشة (٢٠١٥) بدراسة أثر ترتيب فقرات الاختبار على دقة معالم الفقرات وقدرات الافراد والخصائص السيكومترية للاختبار، وتكونت العينة من (١٦٠٠) طالباً وطالبة بواقع (٨٠٠) طالب لكل نموذج موزعة على (٣٥) مدرسة من مدارس المرحلة الثانوية وتم تطبيق اختبار تحصيلي من نوع الاختيار من متعدد في مادة الفيزياء على طلاب الصف العاشر في وحدتي الكهرباء الساكنة والكهرباء المتحركة وتم إعداد نموذجين من الاختبار نموذج (A) وفيه تم ترتيب الفقرات من السهل إلى الصعب، ونموذج (B) وفيه ترتيب الفقرات من الصعب إلى السهل وباستخدام النموذج اللوجستي ثلاثي المَعْلَم وتحليل البيانات وأظهرت النتائج أنه لا توجد فروق دالة إحصائية في الخصائص السيكومترية للاختبار ترجع إلى ترتيب فقرات الاختبار طبقاً للنموذج (A,B).

واهتم أيمن الحجاج (٢٠١٣) بدراسة الترتيب الأفضل لمفردات الاختبار طبقاً لمعاملات التمييز والصعوبة للمفردة للحصول على أفضل أداء للطلاب في الاختبارات التحصيلية وأفضل الخصائص السيكومترية للاختبار. وتم إعداد اختبار في مادة الكيمياء للصف العاشر مكون من (٦٠) مفردة من نوع الاختيار من متعدد من أربع بدائل وصمم خمسة نماذج بعد تطبيقه على

عينة استطلاعية وتكونت العينة النهائية من (٤٠٠) طالب من طلاب الصف التاسع وتم استخدام تحليل التباين الأحادي لفحص دلالة الفروق بين النماذج الخمسة وأسفرت النتائج عن وجود فروق دالة إحصائية بين النماذج الخمسة في أداء الطلاب.

وهدف بحث (2013) Duncan & Mummery إلى معرفة تأثير ترتيب المفردات على كفاءة النشاط الرياضي في المجتمعات الأسترالية وتكونت العينة من (١٢٤٣) من الطلاب الملحقين بدورة الاتصالات وتم تقديم استبيان عن النشاط الرياضي بطريقتين مختلفتين في ترتيب مفرداته وباستخدام معامل الارتباط أسفرت النتائج عن وجود علاقة بين ترتيب المفردة في استبيان النشاط البدني والنشاط الذي يقوم به الطالب.

وقام (2013) Kennedy & Butter بتطبيق نموذجين اختبار في مادة الرياضيات على (٥٩٤) طالبًا بجامعة براون الأمريكية وتم ترتيب المفردات من السهل إلى الصعب (النموذج الأول)، وترتيبها مرة أخرى من الصعب إلى السهل (النموذج الثاني). وقد أظهرت النتائج أن ترتيب المفردات لم يؤثر على أداء الطلاب في مادة الرياضيات.

وتناول بحث (2012) Chen تأثير ترتيب مفردات الاختبار حسب صعوبتها على القلق والأداء الأكاديمي وتكونت العينة من (٢٥٠) طالبًا وطالبة من طلاب الجامعة وطبق اختبار تحصيلي في المعرفة الثقافية ومقياس آخر للقلق وتم قسيم الطلاب إلى ثلاث مجموعات حسب مستوى القلق (مرتفع ومتوسط ومنخفض) وتم ترتيب مفردات الاختبار حسب الصعوبة (من الصعب إلى السهل) وباستخدام اختبار (ت) أسفرت النتائج عن وجود تأثير دال إحصائياً لترتيب المفردات على الأداء الأكاديمي لدى مرتفعي ومتوسطي القلق. ولكن لا يوجد تأثير دال إحصائياً لترتيب المفردات على الأداء الأكاديمي لدى منخفضي القلق.

هدف حمدي أبو جراد (٢٠١٢) إلى المقارنة بين الخصائص السيكومترية لاختبار الاختيار من متعدد واختبار التكميل في ضوء نظرية الاستجابة للمفردة ولتحقيق غرض الدراسة تم إعداد اختبارين في مادة العلوم (٣٠) مفردة للصف السادس الابتدائي أحدهما اختيار من متعدد والآخر من نوع التكميل وكلاهما مكون من (٣٠) وباستخدام نموذج راش تم تدريج ومعايرة مفردات كلا الاختبارين على عينة مكونة من (٦٧) طالبًا وطالبة والحصول على تقديرات الصعوبة مقدرة بوحدة اللوجيت وكشفت نتائج الدراسة عن حذف (٦) مفردات من اختبار الاختيار من متعدد وأشارت إلى أن معاملات ثبات تقديرات الصعوبة للمفردات والأفراد في اختبار التكميل أقل من معاملات ثبات اختبار الاختيار من متعدد.

وقد أجرى حسان العمري (٢٠١١) بحثاً استخدم فيه مقياس الثقة بالنفس، ثم إعادة ترتيب مفرداته بأربع طرائق، الثلاثة الأولى هي: ترتيب المفردات تنازلياً حسب معامل تمييز المفردة، ومعامل ثبات المفردة، ومعامل صدق المفردة، إضافة إلى الترتيب الرابع، وهو ترتيب المفردات عشوائياً. وتكونت العينة من (٢٦٤) طالباً من طلاب الجامعة وتم تقسيمهم إلى أربع مجموعات واستخدم الباحث تصميم القياسات المتكررة في عرض المقياس على عينة من الطلاب، وتلقى كل طالب الأنماط الأربعة للمقياس ولكن بترتيب مختلف من مجموعة لأخرى وباستخدام الأساليب الإحصائية أشارت نتائج البحث إلى تفوق أسلوب ترتيب المفردات اعتماداً على مؤشرات ثباتها، حيث أعطي هذا الترتيب أعلى معامل ثبات للمقياس بالإضافة إلى أن هذا الترتيب قد تقدم على الطرق الأخرى في تحسين متوسطات الخصائص السيكومترية لل فقرات نفسها.

ودرس رمضان رمضان (٢٠١٠) مدى تأثير ترتيب مفردات الاختبار حسب درجة صعوبتها، على تحصيل (١٣٣) طالباً وطالبة بالصف الثاني الثانوي من طلاب المرحلة الثانوية لمادة الرياضيات في المنوفية، وتم تحليل محتوى الدوال الحقيقية بهدف الحصول على المفاهيم الرياضية في اختبار يحتوي على (٢٣) مفردة من نمط الاختيار من متعدد، وتم حساب صدق وثبات الاختبار ومعاملات الصعوبة لكل مفردة، وقسم الاختبار إلى ثلاثة أقسام، اختبارات رتبت أسئلتها بالأشكال الثلاثة للاختبار (ترتيب تصاعدي، تنازلي، عشوائي). دلت نتائج البحث على أنه لا توجد فروق ذات دلالة إحصائية بين متوسطات درجات أفراد عينة البحث في الأشكال الثلاثة للاختبار، ولا توجد فروق ذات دلالة إحصائية بين متوسط درجات الطلاب الذكور أو متوسط درجات الطالبات الإناث، وأن أعلى المتوسطات الحسابية بالنسبة للأشكال الثلاثة للاختبار كان الترتيب الذي يبدأ بالأسئلة السهلة (الترتيب التصاعدي).

وأجرى عدنان اللحاني (٢٠١٠) دراسة بهدف التعرف على أهم طرق ترتيب مفردات الاختبار من حيث تسلسل محتوى المقرر الدراسي للمفردات ومستوى السهولة والصعوبة، والكشف عن نمط الترتيب الأفضل لمفردات الاختبار من حيث الصدق والثبات، ولتحقيق هدف الدراسة تم بناء اختبار تحصيلي متعدد البدائل في البابين الأول والثاني من مقرر الرياضيات للصف الثاني الثانوي قسم العلوم الطبيعية في وحدة المصفوفات والمحددات، وبناء على ذلك تم بناء خمسة نماذج للاختبار نفسه ويتكون كل منها من (٢٥) مفردة وتم ترتيب المفردات على النحو التالي: الأول تم ترتيب المفردات من الصعب إلى السهل، والثاني، تم ترتيب

المفردات من السهل إلى الصعب، والثالث، حسب تسلسل عرض المحتوى، والرابع تم ترتيب المفردات ترتيباً عشوائياً، والخامس تم ترتيب المفردات ترتيباً متبادلاً (دائرياً). وتم تطبيق النماذج الخمسة على (٦٠٠) طالب من طلاب الصف الثاني الثانوي بمدينة مكة المكرمة. دلت النتائج على أنه توجد فروق ذات دلالة إحصائية بين قيم معاملات الثبات تعزى لاختلاف ترتيب مفردات الاختبار بين النموذجين (الترتيب حسب تسلسل عرض المحتوى والترتيب من السهل إلى الصعب) لصالح نموذج الترتيب حسب تسلسل عرض المحتوى، كما أنه توجد فروق دالة إحصائية بين معاملات الثبات للنموذجين (الترتيب من السهل إلى الصعب، والترتيب من الصعب إلى السهل) لصالح نموذج الترتيب من الصعب إلى السهل ولا توجد فروق ذات دلالة إحصائية بين قيم معاملات صدق الاختبار بين النماذج الخمسة.

واهتم Doerner & Calhoun (2009) بدراسة أثر ترتيب المفردات على تحصيل الطلاب، وتم تطبيق ثلاث نماذج من اختبار نهائي يتكون من (٧٥) مفردة من نوع الاختيار من متعدد على (١٢٩٦) طالباً من الملتحقين في مقرر الاقتصاد. وتم ترتيب المفردات طبقاً للمحتوى المعرفي (تصاعدي، تنازلي، عشوائياً)، وأظهرت النتائج وجود فروق ذات دلالة إحصائية في درجات الطلاب في الاختبار ترجع لترتيب المفردات. فقد كانت أعلى الدرجات على النموذج الأول ثم الثاني ثم الثالث، مما يشير لاستفادة الطلاب الذين كان نموذج اختبارهم هو النموذج الأول أكثر من أقرانهم الذين أُجري على النموذجين الثاني والثالث.

التعليق على البحوث السابقة:

تناولت البحوث السابقة دراسة تأثير ترتيب فقرات الاختبار التحصيلي على الخصائص السيكومترية للاختبار وفقراته بالإضافة إلى التأثير على التحصيل الدراسي لدى عينات مختلفة كطلبة المرحلة الابتدائية مثل بحث حمدي أبوجرد (٢٠١٢)، أو طلبة المرحلة الإعدادية مثل بحث سناء عبابنة (٢٠١٧) أو طلبة المرحلة الثانوية مثل بحث رمضان رمضان (٢٠١٠) أو طلبة الجامعة مثل بحث Doerner & Calhoun (2009) ومنهم من استخدم النظرية التقليدية أو نظرية الاستجابة للمفردة على الاختبار التحصيلي مرجعي المعيار. وباستخدام الأساليب الإحصائية مثل اختبار (ت) وتحليل التباين فإنه يوجد الكثير من النتائج منها: ما توصل إليه كل من سناء عبابنة (٢٠١٧)، و Duncan & Mummery (2013) ورمضان رمضان (٢٠١٠)، و Doerner & Calhoun (2009) و Barnett et al (2007) إلى وجود تأثير دال إحصائياً

لنماذج ترتيب مفردات الاختبار على الأداء الأكاديمي والخصائص السيكمترية للاختبار، وتوصل إليه بحث كل من (Kennedy & Butter, 2013)، وخالد سليمان (٢٠١٥) من أنه لا يوجد تأثير دال إحصائياً لنماذج ترتيب فقرات الاختبار على الأداء الأكاديمي والخصائص السيكمترية للاختبار.

بعد الاطلاع على البحوث والدراسات السابقة إتضح أن هناك مجموعة من البحوث الأجنبية والعربية بحثت الترتيب الأفضل والأكثر فاعلية لمفردات الاختبارات، بالاعتماد على معامل صعوبتها أو تمييزها أو طبقاً للمحتوى من السهل إلى الصعب أو من الصعب إلى السهل أو عشوائياً، وذلك للحصول على أفضل أداء عند الطلاب في الاختبارات التحصيلية، بالإضافة إلى الحصول على أفضل الخصائص السيكمترية للاختبار ومفرداته، ولكن معظم هذه البحوث استخدمت توجه النظرية التقليدية في تحليل الاختبارات، والتي من أوجه النقد الموجهة لها اختلاف قيم معاملات الصعوبة والسهولة والتمييز لمفردات الاختبار باختلاف عينة المفحوصين، كذلك معظم البحوث السابقة استخدمت اختبارات تحصيلية مرجعية المعيار. كما لوحظ ندرة الدراسات العربية التي اهتمت بربط الأداء الأكاديمي بطريقة ترتيب مفردات الاختبار وفقاً لنظرية الاستجابة للمفردة، لذلك يأتي هذا البحث لدراسة تأثير ترتيب مفردات الاختبار على معاملات الصعوبة والتمييز للاختبار ومفرداته والأداء الأكاديمي وفقاً لنموذج نظرية الاستجابة للمفردة (النموذج ثنائي المعلم الذي يعتمد على بارامترى الصعوبة والتمييز. وبحث إشكالية الأثر هذا على الأداء التحصيلي للطلاب، وذلك لتحليل استجابات الطلاب على مفردات الاختبار، كما تم اعداد نمط اختبار تحصيلي مرجعي المحك لاعتباره من أهم الوسائل التقييمية التي يستخدمها المعلم، وخاصة في المرحلة الجامعية، فإذا أحسن المعلم بنائه واستخدامه يصبح قادراً على قياس الأهداف المرجوة مثل مدى إلمام الطلاب بالمعلومات والمهارات في المقررات الدراسية وتحقيق أكبر قدر من الموضوعية والعدالة. والوصول إلى طريقة مثلى لترتيب مفردات الاختبار ليحصل الطلاب على درجات مرضية.

فروض البحث:

في ضوء الإطار النظري والبحوث السابقة يمكن اشتقاق فروض البحث كما يلي:

- ١- توجد فروق دالة إحصائية بين ترتيب المفردات طبقاً لتسلسل المحتوى معرفياً أو عدم تسلسله (عشوائية الترتيب) في معاملات الصعوبة للاختبار ومفرداته طبقاً للنموذج ثنائي البارامتر (المعلم) لطلاب الدبلومة العام بكلية التربية.

٢- توجد فروق دالة إحصائية بين ترتيب المفردات طبقاً لتسلسل المحتوى معرفياً أو عدم تسلسله (عشوائية الترتيب) في معاملات التمييز للاختبار ومفرداته طبقاً للنموذج ثنائي البارامتر (المعلم) لطلاب الدبلومة العام بكلية التربية.

٣- لا يوجد تأثير دال إحصائياً لترتيب مفردات الاختبار طبقاً للمحتوى المعرفي (متسلسل أو عشوائي) في الأداء الأكاديمي لطلاب الدبلومة العام في التربية.

إجراءات البحث:

أولاً: منهج البحث:

تم استخدام المنهج الوصفي من نوع (التحليل السيكومتري المقارن) في اختيار عينة البحث الحالي، وبناء أداة البحث، والإجراءات المتبعة للتحقق من الخصائص السيكومترية لتلك الأداة، وكذلك التحقق من صحة فروض البحث الحالي.

ثانياً: عينة البحث:

١- طريقة اختيار العينة ومجتمع البحث ومبررات اختيارها:

تم اختيار عينة البحث الحالي بالطريقة العشوائية البسيطة من مجتمع البحث وهو طلاب وطالبات الدبلومة العام في التربية (نظام التفرغ للعام الواحد) بكلية التربية- جامعة الزقازيق بمحافظة الشرقية في الفصل الدراسي الثاني من العام الجامعي ٢٠١٧/٢٠١٨م.

٢- حجم العينة:

تكونت عينة البحث الاستطلاعية من (٥٣٩) طالباً وطالبة، وتم استخدام العينة الاستطلاعية في حساب الصدق والثبات للاختبار التحصيلي مرجعي المحك، وكذلك حساب درجة القطع. وتكونت عينة البحث الكلية قبل تطبيق درجة القطع من (١٤٥٠) طالباً وطالبة، (٧٣٧ من الذكور، ٧١٣ من الإناث)، وتخصصاتهم (٨٢٩ من طلاب التخصصت العلمية، ٦٢١ من طلاب التخصصات الأدبية)، وقد استخدمت المعالجات الإحصائية المناسبة للتأكد من مطابقة بيانات العينة النهائية لافتراضات نظرية الاستجابة للمفردة والمتمثلة في: أحادية البعد، والاستقلال الموضوعي، جودة مطابقة البيانات، وكذلك تم حساب الخصائص السيكومترية للاختبار، وتقديرات القدرة للأفراد، معالم المفردات والخطأ المعياري لكل منها، ودالة المعلومات للتحقق من فروض البحث الحالي، وتمثلت البيانات الديموغرافية للعينة النهائية كالآتي.

جدول (٢)

البيانات الديموغرافية لعينة البحث النهائية

المجموع	التخصص الأكاديمي		المتغيرات	
	أدبي	علمي	ذكر	النوع الاجتماعي
٧٣٧	٢٦٢	٤٧٥	ذكر	النوع الاجتماعي
٧١٣	٣٥٩	٣٥٤	أنثى	النوع الاجتماعي
١٤٥٠	٦٢١	٨٢٩	المجموع	

وبعد تطبيق درجة القطع على جميع الطلاب في العينة النهائية هذه ، وكما سيتم توضيحه تاليًا فإن قيمة درجة القطع (Cut-score) كانت (٦٠) درجة، أي أن الطالب الذي استطاع أن يجيب على (٦٠) مفردة من أصل (١٠٠) فما فوقها تم إدخاله وتضمينه في ملف التحليل، وبذلك يصبح عدد الطلاب في العينة النهائية (٨٥٨) وهذه العينة التي يتم اختبار الفروض عليها.

ثالثاً: خطوات البحث.

١- بناء الاختبار التحصيلي مرجعي المحك

(أ) الوصف: تمثلت أداة البحث في الاختبار التحصيلي مرجعي المحك في مقرر "الفروق الفردية والقياس النفسي"، وقد اشتمل على الفصول والوحدات الآتية: الفروق الفردية، الشخصية ومكوناتها، المفاهيم النفسية للذكاء، القياس النفسي والتربوي، أسس تصنيف الاختبارات، التقويم التربوي.

(ب) خطوات بناء الاختبار التحصيلي مرجعي المحك: تم إتباع مجموعة من الخطوات لبناء الاختبار وذلك بالاستعانة بعدة مراجع في هذا المجال وهي، (صلاح عبدالوهاب، ٢٠٠١)، ("صلاح الدين" علام، ٢٠٠١)، (إسماعيل الوليلي، ٢٠٠٢)، (محمد "فتح الله"، ٢٠٠٥)، (أبوالمجد الشوربجي، عزت حسن، ٢٠٠٥)، ("صلاح الدين" علام، ٢٠٠٥)، (عماد عبابنة، ٢٠٠٩)، (هند الحموري، أحمد الكلوت، ٢٠٠٩)، (Shawer, 2011)، (باسل أبو فودة، نجاتي بني يونس، ٢٠١٢)، (محمد عليوة، ٢٠١٣)، وقد اتفقت معظم الأبحاث السابقة على العناصر التالية مع ملاحظة أنه قد توجد خطوات تختلف في التسمية من بحث إلى آخر إلا أن أغلبها اتفق على النقاط التالية والتي تمثل خطوات بناء الاختبار. و الخطوات على النحو الآتي:

- ١- تحديد الموضوعات الرئيسية وما يندرج تحتها من موضوعات فرعية في الوحدة الدراسية والأهمية النسبية لكل موضوع فرعي في ضوء تصنيف ثلاثي لجوانب التعلم.
- ٢- تحديد المهارات الرئيسية للوحدة الدراسية و تحليلها إلى مكوناتها و تقييمها، هذا وتمثل الخطوتان (١)،(٢) النطاق السلوكي Behavioral Domain.
- ٣- صياغة الأهداف السلوكية في ضوء نواتج تحليل المهارات الرئيسية.
- ٤- تقييم الأهداف السلوكية.
- ٥- تحديد نوع بنود الاختبار.
- ٦- بناء البنود (المفردات) الاختبارية.
- ٧- تقييم البنود (المفردات) الاختبارية.
- ٨- تجريب البنود (المفردات) استطلاعياً
- ٩- تحديد درجة القطع (درجة الإلتقان).
- ١٠- تقدير الخصائص السيكومترية لمفردات الاختبار (الصدق والثبات والتمييز ومعاملات السهولة والصعوبة وفعالية المشتتات).

ونظراً لكبر وطول هذه الخطوات، فقد اختصرها الباحثان في النقاط السابقة ، وتم استخدام طريقة "أنجوف المعدلة" لحساب درجة القطع والتي كانت عند (٦٠%)، وهو ما يتوافق كذلك مع لائحة الكلية الداخلية (٢٠٠٩)، و المحددة لمستوي الاجتياز والالتقان عند (٦٠%). وقد تم استخدام معامل ليفنجستون لحساب الثبات، والصدق الوصفي عند حساب صدق الاختبار.

٣- التحقق من اعتدالية توزيع الدرجات للعينة النهائية:

تم اختبار إعتدالية التوزيع لنتائج العينة النهائية قبل التحقق من اختبار الفروض، حيث قام الباحثان باختبار اعتدالية التوزيع باستخدام عدة طرق كما ذكرها(عزت حسن، ٢٠١٦، ص ٢٢٠-٢٤٠) كما يلي:

(أ) معاملا التواء والتفلطح والخطأ المعياري لكل منهما:

تم استخدام حزمة البرامج الإحصائية للعلوم الاجتماعية الشهيرة الإصدار الأحدث منها

(Sps25)، وكانت النتائج على النحو الآتي:

جدول (٣)

معاملات الالتواء والتفلطح وقيم أخطائها المعيارية Descriptive Statistics لإحصاءات الوصفية

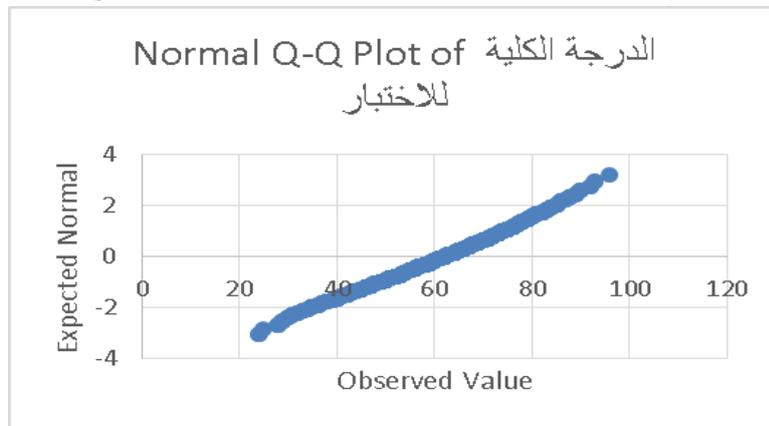
Kurtosis معامل التفلطح		Skewness معامل الالتواء		Std. Deviati on	Mean	Minim um	N	الدرجة الكلية للاختبار
الخطأ المعياري	قيمة المعامل	الخطأ المعياري	قيمة المعامل	الانحراف المعياري	متوسط الإحصاءات	القيمة الدنيا للإحصاءة	عدد الإحصاءات	
١٢٨.٠	-١٦٨.٠	٠٦٤.٠	٢٣٥-٠	١٢.٣٧٥	٦٩٩.٦١	٢٤	١٤٥٠	

يتضح من الجدول رقم (٣) أن درجات الطلاب تتوزع توزيعاً اعتدالياً في ضوء معامل الالتواء، وذلك لأن قيمة معامل الالتواء Skewness أصغر من ضعف الخطأ المعياري له بغض النظر عن الإشارة سالبة أو موجبة. بينما لا تتوزع توزيعاً اعتدالياً في ضوء معامل التفلطح، وذلك لأن قيمة معامل التفلطح (المفرطح) Kurtosis أكبر من ضعف الخطأ المعياري له. لذا ذهبنا إلى طرائق أخرى للتأكد من الاعتدالية أو عدمها.

(ب) استخدام اختبارات الرسوم البيانية:

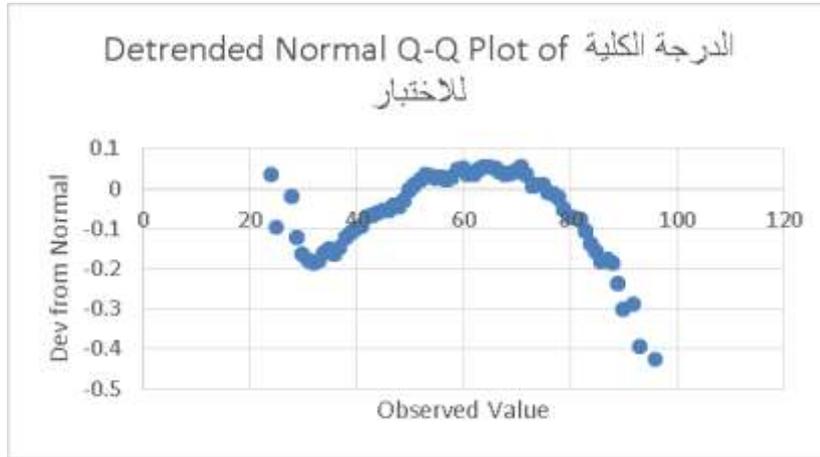
ومنها الرسم البياني (Normal Q-Q plots)، والرسم البياني (Detrended

Normal Q-Q plots)، اختبار رسم الصندوق Box- Plots كما يتضح من الأشكال التالية:



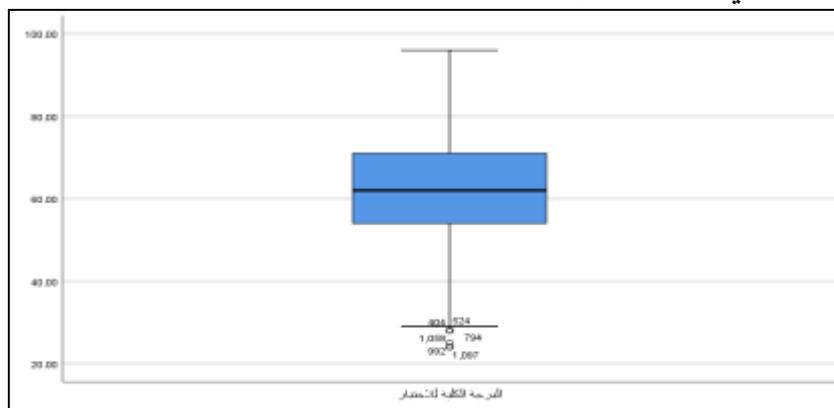
شكل (١): الرسم البياني (Normal Q-Q plots)

ويتضح من الشكل رقم (١) أن بيانات الاختبار إعتدالية، لأن معظم النقاط تقع على الخط المستقيم أو قريبة جداً منه. وهناك الرسم البياني (Detrended Normal Q-Q plots) كما في شكل رقم (٢) التالي



شكل (٢): الرسم البياني (Detrended Normal Q-Q plots).

ويُلاحظ من الشكل رقم (٢) أن القيم المشاهدة تأخذ التوزيع الأقرب للاعتدالية فهي تشبه شكل التوزيع الاعتدالي أو شكل حرف "يو" المقلوب (U-shape). وهناك اختبار رسم الصندوق Box- Plots كما في الشكل التالي :



شكل (٣): الرسم البياني (Box- Plots)

يتضح من الشكل رقم (٣) أن الوسيط يقع في منتصف الصندوق تماماً (الممثل بخط في وسط الصندوق) عند القيمة (٦٢)، مما يدل على إعتدالية التوزيع. مع وجود بعض القيم المتطرفة القريبة Outliers للحالات أرقام (٤٠٤، ٥٢٤، ٧٩٤، ٩٩٢، ١٠٥٨، ١٠٩٧). ونلاحظ من كل مما سبق، أن البيانات اعتدالية التوزيع، لذا تم استخدام الاختبارات والأساليب الإحصائية البارامترية عند التحقق من فروض البحث.

٣- التحقق من مدى تجانس الدرجات للعيينة النهائية:

تم استخدام اختبار (ت) للعينتين المستقلتين (t test -two independent samples) لبحث الفروق بين الذكور والإناث من الطلاب هل هي دالة أم لا، وكذلك حساب قيمة (ف) أو النسبة الفئوية لمعرفة التجانس بين درجات الاختبار للعيينة النهائية، وكانت النتائج الإحصائية كالآتي:

جدول (٤)

قيمة اختبار (ت) للفروق بين مجموعتي الطلاب والطالبات في الاختبار التحصيلي مرجعي المحك.

العيينة	العدد	المتوسط الحسابي	الانحراف المعياري	قيمة (ف)	قيمة (ت)
الطلاب	٧٣٧	٦١,٩١	١٢,٥٧	١,٩٦٦ غير دالة	٠,٦٦٥ غير دالة
الطالبات	٧١٣	٦١,٤٨	١٢,١٧	إحصائية	إحصائية

ويتضح من جدول رقم (٤) أن قيمة النسبة الفئوية (ف) غير دالة إحصائياً مما يدل على تجانس درجات الاختبار.

٤- التحقق من افتراضات نماذج نظرية الاستجابة للمفردة للعيينة النهائية :

(أ) أحادية البعد.

معظم نماذج الاستجابة للمفردة تفترض وجود سمة واحدة وراء أداء الطالب في الاختبار، ومن أجل ذلك قام الباحثان بإجراء تحليل عاملي استكشافي Exploratory Factor Analysis باستخدام طريقة تحليل المكونات الرئيسية Principal Component Analysis بواسطة برنامج (SPSS.25)، لكن قبل الشروع في استخدام طريقة التحليل العاملي تم التحقق من الشروط الواجب توافرها من قابلية مصفوفة الارتباط للتحليل العاملي، وهذه الشروط لخصها (أحمد تيفزة، ٢٠١٢، ص ٢٦-٥٠) كما يلي:

١- ينبغي أن تكون أغلب معاملات الارتباطات في المصفوفة تساوي أو أكبر من (٠,٣)، ودالة إحصائياً، وبفحص مصفوفة الارتباط نجد أن معظم معاملات الارتباط بين الفقرات وبعضها لا تزيد عن (٠,٣) أو تساويها، ولكن معظمها دال إحصائياً، وليس معنى ذلك عدم قابلية المصفوفة للتحليل العاملي، ولكنها قابلة للتحليل لأن هذه المعاملات دالة إحصائياً، بل على العكس فان عدم وجود ارتباطات مرتفعة يدل على تحقق شرط مهم من شروط نماذج الاستجابة للمفردة وهو شرط الاستقلالية (استقلالية الموضع/ المحل).

- ٢- يجب أن تكون القيمة المطلقة لمُحدد مصفوفة الارتباطات أكبر من (الصفر) أو ألا يساوي الصفر، وقد بلغ قيمة هذا المعامل $1,733 \times 10^{-7}$ ، وبالتالي قد تحقق هذا الشرط.
- ٣- ينبغي أن يكون اختبار برتليت Bartlett test of Sphericity دالاً (الدلالة أي أقل من ٠,٠٥)، وهذا قد تحقق أيضاً، وهو يدل على أن مصفوفة الارتباطات ليست مصفوفة الوحدة، بل تتوفر على الحد الأدنى من العلاقات، مستوى الدلالة أقل من (٠,٠٠٠١).
- ٤- ينبغي أن يكون اختبار Kaiser- Mayer-Olkin (KMO) كايزر- ماير- أولكين لكافة المصفوفة أعلى من (٠,٠٥) وفقاً لمحك كايزر، وقد بلغ قيمة هذا المعامل (٠,٩١) وقيمته مرتفعة مما يدل على تجانس العينة ككل، وهو مقياس عام لكفاءة التعيين. ويتضح ذلك من الجدول التالي.

جدول (٥)

اختبار كايزر- ماير أولكن، وبرتليت للتحقق من قابلية المصفوفة للتحليل العاملي.

المؤشر	القيمة	المؤشر	القيمة
مقياس كايزر- ماير أولكين	٠,٩١	درجات الحرية لاختبار برتليت	٤٩٥٠
قيمة مربع كاي التقريبية	٢٢٠٢٦,٧٩٦	الدلالة الإحصائية	٠,٠٠٠

يتضح من جدول رقم (٥) تحقق افتراضات قابلية المصفوفة للتحليل العاملي الاستكشافي، ومن ثم تم إجراء تحليل عاملي Factor Analysis، باستخدام طريقة تحليل المكونات الرئيسية Principal Component Analysis بواسطة برنامج (SPSS.25)، وذلك للتحقق من أن الاختبار يقيس بعداً واحداً كما هو مفترض، وذلك طبقاً للجدول التالي:

جدول (٦)

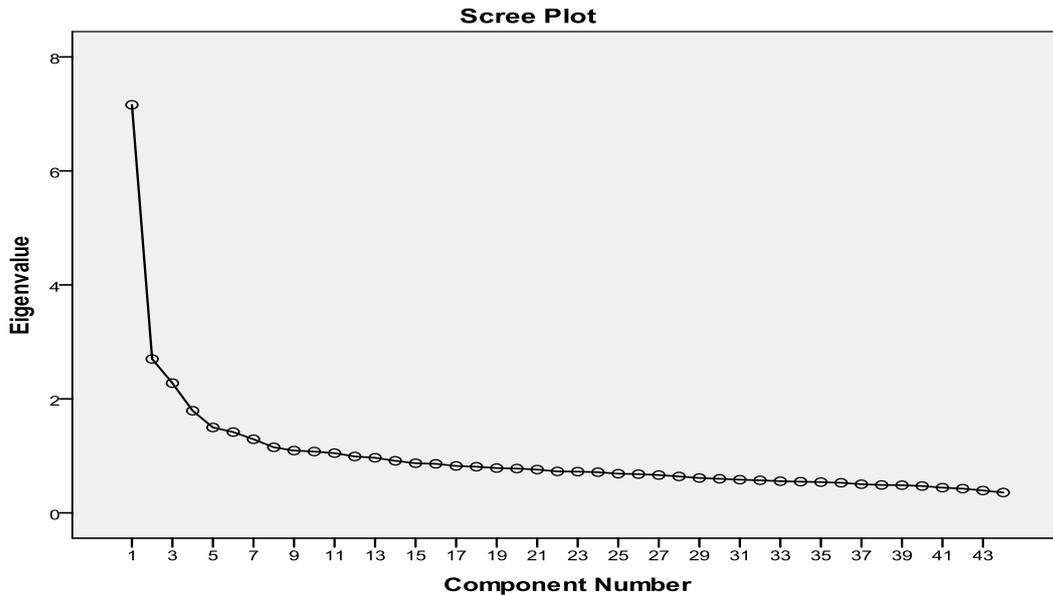
قيم الجذور الكامنة قبل وبعد التدوير، وتشبع العبارات على العوامل، ونسبة التباين التي يفسرها كل عامل والنسبة التراكمية.

قيم الجذور الكامنة بعد التدوير Rotation Sums of Squared Loadings			قيم الجذور الكامنة قبل التدوير Initial Eigen values			العامل Component
نسبة التباين التراكمية	حجم التباين المفسر	قيم الجذور الكامنة	نسبة التباين التراكمية Cumulative %	حجم التباين المفسر % of Variance	قيم الجذور الكامنة Total	
١٠,٦٦٤	١٠,٦٦٤	١٠,٦٦٤	١٠,٦٦٤	١٠,٦٦٤	١٠,٦٦٤	١
١٤,٠٦٨	٤,٤٠٤	٤,٤٠٤	١٤,٠٦٨	٤,٤٠٤	٤,٤٠٤	٢
			١٩,٠٦٤	٤,٩٩٦	٤,٩٩٦	٣
			٢٠,٤٠٩	١,٣٤٤	١,٣٤٤	٤
			٢١,٧١٧	١,٣٠٨	١,٣٠٨	٥
			٢٢,٩٨١	١,٢٦٤	١,٢٦٤	٦
			٢٤,٢٢٧	١,٢٤٦	١,٢٤٦	٧
			٢٥,٤٦٤	١,٢٣٧	١,٢٣٧	٨
			٢٦,٦٨٨	١,٢٢٤	١,٢٢٤	٩
			٢٧,٩٠١	١,٢١٣	١,٢١٣	١٠

ويتضح من الجدول رقم (٦) أن الاختبار تشبع بعامل واحد قبل التدوير (كما يظهر من قيم الجذور الكامنة الأولية)، وهذا العامل جذره الكامن = (١٠,٦٤٤)، وهي تساوي تقريباً مرتين قيمة الجذر الكامن للعامل الثاني الذي وصل إلى (١٠,٤٠٤) (نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني أكبر من ٢)، وهذا يشير إلى توافر افتراض أحادية البعد، حيث أشار العديد من خبراء القياس والتقويم ومنهم (ريكاس، ١٩٨٥) إلى أنه إذا فسر العامل الأول أكبر تباين، فإن ذلك يعتبر مؤشراً على أحادية البعد (في: محمد الشافعي، ٢٠١١، ص ٥٧٥).

وكذلك بعد التدوير حيث بلغ جذره الكامن (١٠,٦٤٤)، وهي تساوي تقريباً ضعف قيمة الجذر الكامن للعامل الثاني الذي وصل (٤,٤٠٤) وقد فسر (١٠,٦٤٤%) من التباين في درجات الاختبار ككل، مما يعني أن هذا العامل الأول هو المسيطر على تفسير التباين الكلي لدرجات الاختبار، وبالتالي يعتبر هذا الاختبار أحادي البعد أي أن هناك سمة كامنة واحدة يقيسها

الاختبار، وهي المسئولة عن تفسير ما يحدث من تباين في درجات الاختبار، كذلك تم التحقق أحادية البعد من الشكل رقم (٤) (منحنى المنحدر) والذي يعرف باختبار الهضبة أو الفرز (scree Plot) كالتالي.



شكل (٤): منحنى المنحدر للتحقق من أحادية البعد

ويتضح من الشكل رقم (٤) أن هناك انحداراً كبيراً بين قيم الجذر الكامن بين العامل الأول والعامل الثاني، بينما يقل هذا الانحدار بشكل واضح بين الجذور الكامنة للعوامل الأخرى، مما يشير إلى مؤشر آخر لتمتع الاختبار بأحادية البعد.

(ب) الاستقلال الموضوعي.

مما يؤكد تحقق هذا الافتراض أن مفردات الاختبار المستخدم تقيس سمة أحادية البعد، حيث أن تحقق أحادية البعد في البيانات كاف لتحقيق افتراض الاستقلال الموضوعي ودليل على تحققه.

ويُعد افتراض الاستقلال الموضوعي افتراضاً نظرياً ويعنى أن أداء الفرد لأحد البنود يجب ألا يؤثر سلباً أو إيجابياً على استجابته للمفردات الأخرى (أي أن الإجابة على أي مفردة اختبارية لا ينبغي أن تتأثر بالإجابة على المفردات الأخرى)، ويجب تحققه في البيانات لتلائم التحليلات المتعلقة بنماذج الإستجابة للمفردة. ويتم الكشف عن هذا الافتراض من إفتراضات نظرية الإستجابة للمفردة من خلال حساب معاملات الارتباطات الجزئية Partial Correlation بين كل زوج من أزواج المفردات عند مستويات الدرجة المختلفة، بحيث إذا جاءت قيم هذه

المعاملات قريبة بشكل كبير من الصفر فإن ذلك يعتبر مؤشراً على تحقق افتراض الاستقلال الموضوعي، ولقد استخدم برنامج spss.25 لحساب معاملات الارتباطات الجزئية بين المفردات مع ضبط الدرجة الكلية على الاختبار (إيهاب السيد، ٢٠١٢، ص ١٤٤). وهذا ما تحقق بالفعل.

ج) المنحنيات المميزة للمفردة .

يعتبر برنامج (Bilog-MgV3.0) الذي اعتمد عليه الباحثان في تحليل بيانات البحث الحالي، قد حقق هذا الافتراض المتعلق بوجود منحني مميز لكل مفردة اختبارية، حيث إن مخرجات التحليل باستخدام هذا البرنامج تشتمل على جانبين: أحدهما يمثل أشكالاً بيانية توضح المنحني المميز للمفردة ومدى ملاءمتها للنموذج المستخدم، والجانب الآخر يمثل أشكالاً بيانية توضح الدالة المميزة للمفردة (a,b) وقدرة الفرد (θ)، ودالة معلومات الاختبار، والخطأ المعياري المتعلق بالقياس، وبذلك فإن البيانات الحالية تحقق الافتراض المتعلق بوجود منحني مميز بكل مفردة من المفردات الاختبارية بعد استخدام البرنامج في الإجراءات والتحقق من افتراضات نماذج الاستجابة للمفردة في الاختبار. وتم فحص مطابقة الطلاب في ضوء نماذج لوجستية للنموذج الثنائي المعلم باستخدام برنامج Bilog-mg3. وتقدير قدرة الطلاب باستخدام طريقة بيز في التقدير. ولم يتم حذف أي من الطلاب.

د) التحقق من توافر افتراض السرعة.

وهذا الافتراض معناه ألا تلعب السرعة دوراً في أداء الطلاب على المفردات الاختبارية؛ أي أن الأسئلة المتروكة في نهاية الاختبار ينبغي أن تكون قد تركت بسبب أن قدرة المفحوص لم تمكنه من الإجابة عنها بشكل صائب، وليس بسبب أنه لم يكن هناك وقت كافٍ لمحاولة المفحوص أن يجيب عنها، عموماً فقد تم استبعاد الطلاب الذين لم يكملوا الإجابة على الاختبار بشكل عام، وقد راعى الباحثان في تطبيق اختبار البحث إعطاء الوقت الكافي للإجابة عن جميع بنود الاختبار وهو زمن (ساعتين)، ولم يتبين للباحثين أي ردود أفعال من الطلاب عن ضيق الوقت في الإجابة عن المفردات.

هـ) التحقق من حسن مطابقة بيانات الاختبار للنموذج .

تم استخدام برمجية (Bilog-MgV3.0) للتحقق من هذا الافتراض، حيث إن المفردة تكون غير مطابقة لنموذج التحليل، إذا كانت قيمة مؤشر المطابقة (Chi-square) دالة عند مستوى (٠.٠٥). ويتم ذلك أثناء التحقق من فروض البحث.

اشتراطات وإجراءات قبلية:

- ١- في العينة النهائية: عدد طلاب النموذج (أ) (٧٢٥) طالبًا وطالبة، وعدد طلاب النموذج (ب) (٧٢٥) طالبًا وطالبة، وبذلك يكون لدينا عدد متساوي لأفراد العينة في كل نموذج بالنسبة للعينة النهائية. وبعد تطبيق درجة القطع أصبح عدد الطلاب في العينة النهائية (٨٥٨) بنسبة (٥٩,٢) من العينة الكلية (١٤٥٠); منها (٤٥١) طالبًا وطالبة في النموذج (أ)، (٤٠٧) طالبًا وطالبة في النموذج (ب).
- ٢- في العينة الاستطلاعية: إجراءات الصدق والثبات لم ينتج عنها حذف أي من المفردات، فكما تم توضيحه من قبل أن الاختبار في صورته المبدئية تكون من (٦٠) مفردة من نوع الاختيار من متعدد ذي الأربعة بدائل (أ، ب، ج، د)، (٤٠) مفردة من نوع الصواب والخطأ (X, ✓)، وُزعت في نموذجين (أ، ب)، حيث توزعت المفردات جميعًا بطريقة مرتبة بالتسلسل طبقًا للمحتوي العلمي المعرفي في النموذج (أ)، وبطريقة مرتبة عشوائيًا في النموذج (ب) وأصبح الاختبار بصورته النهائية (١٠٠) مفردة.
- ٣- فيما يتعلق بملف البيانات الإحصائي: الخاص بالنموذج (ب) المرتب عشوائيًا، تم إعادة ترتيب وضع المفردات للتوافق مع نفس وضع المفردات في النموذج (أ) المتسلسل، فمثلاً المفردة رقم (٥) في النموذج (ب) تم إعادة وضعها في ملف البيانات الإحصائي لتصبح رقم (١) لكي تتوافق مع نفس المفردة (١) في النموذج (أ)، وهكذا لباقي المفردات، وبذلك هناك ترتيب موحد للمفردات جميعاً على سلم تدريجي واحد.
- ٤- فيما يتعلق بدرجة القطع: تم تطبيق درجة القطع على جميع الطلاب في العينة النهائية، وكما تم توضيحه سابقاً فإن قيمة درجة القطع (Cut-score) كانت (٦٠) درجة، أي أن الطالب الذي استطاع أن يجيب على (٦٠) مفردة من أصل (١٠٠) فما فوقها تم دخاله وتضمينه في ملف التحليل، وبذلك يصبح عدد الطلاب في العينة النهائية (٨٥٨) بنسبة (٥٩,٢) من العينة الكلية (١٤٥٠); منها (٤٥١) طالبًا وطالبة في النموذج (أ)، (٤٠٧) طالبًا وطالبة في النموذج (ب).
- ٥- ثبات وصدق الاختبار التحصيلي في النموذجين: من المعروف أن معامل الثبات الذي يتم حسابه من خلال الصيغ التقليدية يتأثر بمدى تغير درجات الاختبار، ولأن الاختبارات مرجعية المحك لا تصمم على أساس زيادة تباين درجات المفحوصين، فمن الواضح أنه لا

يمكن تجنب مشكلة تجانس (Homogeneity) درجات المفحوصين في مثل هذا النوع من الاختبارات وللتعامل مع تجانس درجات المفحوصين اقترح ليفنجستون (١٩٧٠) بديلاً لمعاملات الثبات التقليدية ، حيث استخدم علامة القطع (Cut- off Score) التي يمكن من خلالها التعرف على مستوى الإتقان كبديل عن الوسط الحسابي لدرجات المفحوصين، إذ اهتم بإعادة تعريف انحراف درجات الاختبارات مرجعية المحك وعليه يكون ثبات الاختبار مرجعي المحك المحسوب بهذه الطريقة هو على الأقل نفس قيمة ثبات الاختبار مرجعي المعيار، وتؤدي زيادة الفروق بين الوسط الحسابي للدرجات ودرجة القطع إلى زيادة معامل ثبات الاختبار، وبالتالي فإن تغيير علامة القطع سيؤدي إلى تغيير معامل الثبات (عماد عبابنة، ٢٠٠٩، ص ١٢٦).

وقد استخدم الباحثان المعادلة التالية في حساب معامل "ليفنجستون" (أحمد الرفاعي غنيم، ١٩٨٥، ص ٥٥).

$$r = \frac{E^2 + (M - J)^2}{E^2 + (M - J)^2}$$

حيث :

(٢ع) ترمز إلى تباين درجات مجموعة المفردات الكلية للمفحوصين .

(رأ) ترمز إلى معامل ثبات كيوذر ريتشاردسون ٢٠، أو ألفا إذا كانت البيانات ثنائية الاستجابة (٠,١).

(م) ترمز إلى المتوسط الحسابي لدرجات مجموع المفردات .

(ج) ترمز إلى درجة القطع .

وقد تم حساب قيمة معامل الثبات لنموذجي الاختبار عند درجة القطع (٦٠%) فكانت قيمة معامل ثبات النموذج (A) المرتب معرفياً هي (٠,٨٨١)، وقيمة معامل ثبات النموذج (B) المرتب عشوائياً هي (٠,٧٨٥) وتعد هذه القيم مرتفعة مما يدل على ثبات الاختبار. وبالإضافة لذلك تم حساب الثبات بالطرق التقليدية كما يلي:

1/5 ثبات وصدق الاختبار التحصيلي في النموذج (أ) المرتب بتسلسل معرفيا:

(أ) طريقة معامل ألفا كرونباخ، تم حساب ثبات كل جزء من جزئي الاختبار التحصيلي في النموذج (أ) بالإضافة إلى ثبات الدرجة الكلية للاختبار ككل، حيث إن قيمة معامل الثبات ألفا لاختبار الاختيار من متعدد في نموذج (أ) عند حذف المفردة تتراوح من (٠,٧٧٩) إلى (٠,٧٨٤) وهي قيم مرتفعة ولم يتم حذف أية مفردات ومعامل ثبات اختبار الاختيار من متعدد ككل قيمته (٠,٧٨٤) كما في ملحق رقم (١)، بالإضافة إلى حساب قيمة معامل الثبات ألفا لاختبار الصواب والخطأ نموذج (أ) عند حذف المفردة تتراوح من (٠,٨٠٣) إلى (٠,٨١٢) وهي قيم مرتفعة ولم يتم حذف أية مفردات ومعامل ثبات اختبار الصواب والخطأ في نموذج (أ) قيمته (٠,٨١٢) كما في ملحق رقم (٢)، كما أن قيمة معامل الثبات ألفا للاختبار التحصيلي ككل في نموذج (أ) عند حذف المفردة تتراوح من (٠,٦٧٣) إلى (٠,٦٧٨) وهي قيم مرتفعة ولم يتم حذف أية مفردات ومعامل ثبات الاختبار التحصيلي ككل نموذج (أ) قيمته (٠,٦٧٨) كما في ملحق رقم (٣).

(ب) طريقة التجزئة النصفية: تم حساب ثبات الأبعاد الفرعية والثبات الكلي للاختبار التحصيلي في مقرر الفروق الفردية والقياس النفسي بطريقة التجزئة النصفية لـ " سبيرمان / براون" فكانت معامل ثبات اختبار الاختيار من متعدد في النموذج (أ) (٠,٧٩٤)، ومعامل ثبات اختبار الصواب والخطأ في النموذج (أ) (٠,٨٢٥) ويتضح من السابق ثبات جزئي الاختبار التحصيلي وثبات المقياس ككل. وعموما فإن المقياس ككل على درجة عالية من الثبات.

(ج) الصدق الداخلي (صدق المفردات): إن عدد من المتخصصين يرغبون في تمييز أنواع الصدق في الاختبارات مرجعية المحك عنها في الاختبارات مرجعية المعيار، فيذكرون الصدق الوصفي (Descriptive Validity) كبديل عن صدق المحتوى إذ يرون أن هذه التسمية تعكس شمولية أكبر من صدق المحتوى كما أنهم يذكرون الصدق الوظيفي (Functional Validity) كبديل عن الصدق التجريبي الأمبريقي (Empirical Validity)، كما أنهم أضافوا مفهوم صدق انتقاء المجال السلوكي (Domain selection Validity)، الذي يرتبط الذي يرتبط بمفاهيم نظرية التعميم (Generalizability) (عماد عابنة، ٢٠٠٩، ص ١٥٢).

وتم حساب صدق مفردات الاختبار التحصيلي نموذج (أ) عن طريق حساب معاملات الارتباط بين درجات كل مفردة ودرجة البعد الذي تنتمي إليه وذلك بعد حذف المفردة من الدرجة

الكلية لهذا البعد باعتبار أن بقية المفردات محكًا لهذه المفردة ويسمى هذا بالصدق الداخلي (فؤاد البهى السيد، ١٩٧٩، ص ٤٥٧). ويتضح من الملاحق أرقام (١، ٢، ٣) أن جميع معاملات الارتباط دالة إحصائيًا مما يحقق الصدق الداخلي لجزئي الاختبار التحصيلي (الاختبار من متعدد - الصواب والخطأ). ومن ثم فإن مقياس الاختبار التحصيلي ككل يتميز بالصدق الداخلي.

٢/٥ ثبات وصدق الاختبار التحصيلي في النموذج (ب) المرتب عشوائياً

أ) طريقة معامل ألفا كرونباخ، تم حساب ثبات كل جزء من جزئي الاختبار التحصيلي في النموذج (ب) بالإضافة إلى ثبات الدرجة الكلية للاختبار ككل، حيث إن قيمة معامل الثبات ألفا للاختبار الاختيار من متعدد في نموذج (ب) عند حذف المفردة تتراوح من (٠,٧٧٣) إلى (٠,٧٧٠) وهى قيم مرتفعة ولم يتم حذف أية مفردات ومعامل ثبات اختبار الاختيار من متعدد ككل قيمته (٠,٧٧٠) كما في ملحق رقم (٤)، بالإضافة إلى حساب قيمة معامل الثبات ألفا للاختبار الصواب والخطأ نموذج (ب) عند حذف المفردة تتراوح من (٠,٧٧١) إلى (٠,٧٧٨) وهى قيم مرتفعة ولم يتم حذف أية مفردات ومعامل ثبات اختبار الصواب والخطأ في نموذج (ب) قيمته (٠,٧٧٨) كما في ملحق رقم (٥)، كما أن قيمة معامل الثبات ألفا للاختبار التحصيلي ككل في نموذج (ب) عند حذف المفردة تتراوح من (٠,٥٩٨) إلى (٠,٦٠١) وهى قيم مرتفعة ولم يتم حذف أية مفردات ومعامل ثبات الاختبار التحصيلي ككل نموذج (ب) قيمته (٠,٦٠١) كما في ملحق رقم (٦).

ب) طريقة التجزئة النصفية: تم حساب ثبات الأبعاد الفرعية والثبات الكلي للاختبار التحصيلي فى مقرر الفروق الفردية والقياس النفسى بطريقة التجزئة النصفية لـ " سبيرمان / براون" فكانت معامل ثبات اختبار الاختيار من متعدد في النموذج (ب) (٠,٧٧٦)، ومعامل ثبات اختبار الصواب والخطأ في النموذج (ب) (٠,٧٧٩) ويتضح من السابق ثبات جزئي الاختبار التحصيلي وثبات المقياس ككل. وعموماً فإن المقياس ككل على درجة عالية من الثبات.

ج) الصدق الداخلى (صدق المفردات): تم حساب صدق مفردات الاختبار التحصيلي نموذج (ب) عن طريق حساب معاملات الارتباط بين درجات كل مفردة ودرجة البعد الذى تنتمى إليه وذلك بعد حذف المفردة من الدرجة الكلية لهذا البعد باعتبار أن بقية المفردات محكًا

لهذه المفردة ويسمى هذا بالصدق الداخلي (فؤاد البهي السيد، ١٩٧٩، ٤٥٧). ويتضح من الملاحق أرقام (٤، ٥، ٦) أن جميع معاملات الارتباط دالة إحصائياً مما يحقق الصدق الداخلي لجزئي الاختبار التحصيلي (الاختبار من متعدد- الصواب والخطأ). ومن ثم فإن مقياس الاختبار التحصيلي ككل يتميز بالصدق الداخلي.

يتضح مما سبق أن معاملات ثبات وصدق الاختبار التحصيلي في النموذج (أ) المرتب بتسلسل معرفياً أعلى من معاملات ثبات وصدق الاختبار التحصيلي في النموذج (ب) المرتب عشوائياً. ولهذا يوصى الباحثان بترتيب المفردات طبقاً للمحتوى لأنه يرفع من ثباتها وصدقها.

نتائج البحث وتفسيرها ومناقشتها:

نتائج الفرضين الأول والثاني وتفسيرها ومناقشتها:

ينص الفرض الأول على أنه " توجد فروق دالة إحصائية بين ترتيب المفردات طبقاً لتسلسل المحتوى معرفياً أو عدم تسلسله (عشوائية الترتيب) في معاملات الصعوبة للاختبار ومفرداته طبقاً للنموذج ثنائي البارامتر (المُعَلَّم) لطلاب الدبلومة العام بكلية التربية.

وينص الفرض الثاني على أنه " توجد فروق دالة إحصائية بين ترتيب المفردات طبقاً لتسلسل المحتوى معرفياً أو عدم تسلسله (عشوائية الترتيب) في معاملات التمييز للاختبار ومفرداته طبقاً للنموذج ثنائي البارامتر (المُعَلَّم) لطلاب الدبلومة العام بكلية التربية.

ولاختبار صحة الفرضين الأول والثاني تم استخدام برنامج الحاسب الآلي (Bilog-Mg v3.0) لحساب تقديرات قيم الصعوبة للمفردات، وكذلك قدرات الأفراد والأخطاء المعيارية لكل منهما، وقيم معاملات التمييز وأخطائها المعيارية، وقيم مربع كاي باستخدام نماذج الاستجابة للمفردة (النموذج ثنائي المُعَلَّم)، ويتم عرض نتائج الفرضين على النحو التالي:

تم حساب قيم معاملات الصعوبة وأخطائها المعيارية، وقيم معاملات التمييز، والأخطاء المعيارية لها وقيم مربع كاي ودرجات الحرية لكل مفردة، وتم الحصول على قيم هذه البارامترات بعد عدد (١٢) دورة من المحاولات/الاشتقاقات (SUBTEST TEST0001; ITEM PARAMETERS AFTER CYCLE 12)، وتم تقدير القدرات والأخطاء المعيارية لها بطريقة (بييز المشروطة المسبقة) ((EAP) Bayes (Expected A posteriori)) والجدول التالي يوضح هذه القيم.

جدول (٧)

معاملات الصعوبة وأخطاؤها المعيارية، ومعاملات التمييز وأخطاؤها المعيارية باستخدام النموذج ثنائي البارامتر (المعلم) للنموذجين (أ) المرتب معرفياً، (ب) المرتب عشوائياً

درجات الحرية	مربع كاي وتعتمد الدلالة على درجات الحرية لكل نموذج		الخطأ المعياري لمعاملات التمييز		معاملات التمييز SLOPE		الخطأ المعياري لمعاملات الصعوبة E.S		معاملات الصعوبة THRESHOID		الخاصية	
	نموذج (ب)	نموذج (أ)	نموذج (ب)	نموذج (أ)	نموذج (ب)	نموذج (أ)	نموذج (ب)	نموذج (أ)	العشوائي نموذج	المتسلسل نموذج (أ)		
٧	٩	١,١	٧,٠	٠,٠٦٦	٠,٠٥٠	٠,٢٦٢	٠,٥٣١	٠,٧١٥	٠,٣٧٠	-٢,٨٧٣	١,٢٩٣	١
٧	٧	١,٩	٠,٤٨٠	٠,٠٧٦	٠,٠٨٥	٠,٣١٢	٠,٤٤٩	٠,٦٥٢	٠,٢٧٧	-٢,٧٥٥	-١,٢٤٨	٢
٥	٨	١,٣	٨,١	٠,١٠٧	٠,٠٧٥	٠,٤٢٣	٠,٤٢١	٠,٦٣٩	٠,١٦٨	-٢,٨٣٩	-٠,٦٦٧	٣
٦	٨	١٠,٦	١,٠	٠,٠٧٩	٠,٠٧١	٠,٣٩٧	٠,٣٢٦	٠,٢٠٦	٠,٤٤٥	-٠,٨٣٦	-٢,٠٨٣	٤
٧	٧	٨,٧	١١,٤	٠,٠٧١	٠,٠٧٣	٠,٣٤٥	٠,٣٤٤	٠,٥٣٩	٠,٤٦٥	-٢,٦٤٣	-٢,٢٤٩	٥
٦	٨	٢,٦	٦,٥	٠,٠٨٣	٠,٠٥٩	٠,٣٢٣	٠,٢٩١	١,١٣٩	٠,٣٩٤	-٤,٥٩٠	-١,٧٨٧	٦
٧	٧	٢,٧	١٠,٧	٠,٠٧١	٠,٠٨٩	٠,٢٧٢	٠,٤٧٢	٠,٨٧٥	٠,٢٨٤	-٣,٤١٥	-١,٦٦٧	٧
٧	٧	٩,٥	٧,٥	٠,٠٧٢	٠,٠٨٢	٠,٣٨٢	٠,٣٤٥	٠,٢٣٩	٠,٧٣٥	٠,٩١٠	-٣,٢٤٩	٨
٧	٧	٢,٨	٦,٣	٠,٠٧٦	٠,٠٦٩	٠,٣٣٣	٠,٢٥١	٠,٥٥٨	١,١٧٤	-٢,٥٣٠	-٤,٢٩٧	٩
٧	٧	٢,٢	٧,٠	٠,٠٧٨	٠,٠٨٦	٠,٣٥٦	٠,٣٢٨	٠,٣٨٤	٠,٩٧٩	-١,٧٩٣	-٣,٩٠٥	١٠
٧	٧	٦,٥	٨,٧	٠,٠٨٤	٠,٠٨٤	٠,٤٤٤	٠,٤٩٤	٠,٢٥٩	٠,٢٢٩	-١,٣٧٤	-١,٤٤٤	١١
٧	٨	٦,٩	٧,٢	٠,٠٧٨	٠,٠٦٤	٠,٤٦٥	٠,٢٩٢	٠,١٥٧	٠,٤١٧	-٠,٦١١	-١,٨٢٦	١٢
٦	٨	٤,٠١	١٢,٨	٠,٠٧٧	٠,٠٦٩	٠,٢٩٢	٠,٣٢٧	٠,٨٦٢	٠,٣٧٥	-٣,٤٠٨	-١,٧٧٧	١٣
٧	٨	*١٤,١	٧,٦	٠,٠٦٦	٠,٠٦٥	٠,٢٤٦	٠,٣٧٤	٠,٨٣٦	٠,١٥٧	-٣,١٣٦	-٠,١٢٦	١٤
٥	٨	٢,١٣٦	٩,٥	٠,١٤٠	٠,٠٧٥	٠,٥٠٨	٠,٤١١	٠,٦٩٨	٠,٢٢٣	-٣,٠٤٠	-١,١٣٠	١٥
٥	٨	٦,٢	١٣,٠	٠,١٢٦	٠,٠٧٦	٠,٤٥٠	٠,٤٩٩	٠,٧٨٠	٠,١٣٢	-٣,١٩٩	-٠,٤٥٨	١٦
٦	٦	٣,٥	٥,٨	٠,٠٩٢	٠,٠٧٦	٠,٣٦٢	٠,٣١٧	٠,٧٧٢	٠,٦٣١	-٣,٢٨٣	-٢,٧٤٨	١٧
٨	٨	١٠,٥	١٥,٥	٠,٠٦٦	٠,٠٧٠	٠,٣٣٢	٠,٣٣٩	٠,٠٦٥	٠,٣٥٦	٠,١٨٥	-١,٧٠٢	١٨
٦	٦	٧,٢	٢,٧	٠,٠٩٨	٠,١٠٠	٠,٣٧١	٠,٣٥٣	٠,٧٣٤	١,٠٠٤	-٣,٠٤٤	-٣,٨١٦	١٩
٨	٩	١١,٠٤٤	٦,٤	٠,٠٦٣	٠,٠٥٧	٠,٢٩٦	٠,٢٧٣	٠,٢٥٢	٠,٩٢٣	-٠,٨٠٣	-٢,٧١١	٢٠
٧	٦	١٢,١	١,٣	٠,٠٧٧	٠,٠٨٩	٠,٣٢٣	٠,٣١٩	٠,٤٦١	١,٠٠٢	-١,٩٩٦	-٣,٨٠٤	٢١
٦	٦	٨,٣	٢,٥	٠,٠٧٩	٠,٠٨٢	٠,٣٢٩	٠,٢٧٨	٠,٤٩٨	١,٣٨٥	-٢,١٨٥	-٤,٧٩٦	٢٢
٨	٩	٢,٧	٥,٢	٠,٠٦٩	٠,٠٥٩	٠,٣٣٤	٠,٣٢٠	٠,٢٦١	٠,٢٠٤	-١,٠٢٠	٠,٤٩٨	٢٣
٨	٨	٥,٠٢٠	٩,٠٢٠	٠,٠٥٨	٠,٠٦٦	٠,٢٧٧	٠,٣٤٩	٠,٢٢٣	٠,٢٠٧	٠,٢٣٨	-٠,٧٦٩	٢٤
٦	٨	٩,٦	١٣,٠١	٠,٠٨٣	٠,٠٧١	٠,٣٧٦	٠,٤٣٤	٠,٤٠٨	٠,١٣٨	-١,٩٤٨	-٠,٢١٥	٢٥

درجات الحرية		مربع كاي وتعتمد الدلالة على درجات الحرية لكل نموذج		الخطأ المعياري لمعاملات التمييز		معاملات التمييز SLOPE		الخطأ المعياري لمعاملات الصعوبة E.S		معاملات الصعوبة THRESHOID		الخاصية
												رقم المفردة
٧	٨	٣,١	٣,٤	٠,٠٧٥	٠,٠٦٢	٠,٢٩١	٠,٢٦٤	٠,٧٦٠	٠,٤٥٢	-٣,٠٣٧	-١,٨٢٢	٢٦
٧	٧	٢,٣	٣,٨	٠,٠٨٣	٠,٠٧٢	٠,٢٨٩	٠,٣١٤	٠,٤١٠	٠,٧٠٨	-٢,٠٥٢	-٣,١٦٨	٢٧
٧	٧	٤,٤٧	٤,٦	٠,٠٧٦	٠,٠٧٨	٠,٣٩١	٠,٣٠١	٠,١٥٩	٠,٩٦٣	-٠,١٤٧	-٣,٨٥٩	٢٨
٨	٨	٨,١١	٣,١٤٧	٠,٠٧١	٠,٠٥٩	٠,٣٦٠	٠,٢٨٥	٠,١٧٦	٠,٢٩٢	-٠,٣٢٩	-١,١١٥	٢٩
٧	٨	٨,٦	٥,٥	٠,٠٦٧	٠,٠٧١	٠,٢٨٧	٠,٣٦٨	٠,٤٥٦	٠,٢٩٣	-١,٨٧٨	-١,٤٧٩	٣٠
٨	٧	٤	٦,٢	٠,٠٥٧	٠,٠٧٣	٠,٢٥٥	٠,٣٣٣	٠,٢٣٦	٠,٥٤٩	٠,١٤١	-٢,٦٠٦	٣١
٧	٨	٤,٤	٥,٦	٠,٠٧١	٠,٠٥٨	٠,٣٠٤	٠,٢٦٠	٠,٤٢٦	٠,٤٠٨	-١,٨٠٨	-١,٦٤٢	٣٢
٧	٨	٥,٩	٢,٨٠٨	٠,٠٧٦	٠,٠٥٩	٠,٣٥٥	٠,٢٥٨	٠,٢٨٨	٠,٥١٤	-٦,٢٥٨	-٢,١٤١	٣٣
٧	٧	١,٦	٢,٩	٠,٠٧٤	٠,٠٧٩	٠,٣٠٣	٠,٣٨٤	٠,٨١٣	٠,٣٦٨	-٣,٤٤٢	-١,٨٨٧	٣٤
٨	٨	٧,٧	٥,٣	٠,٠٧١	٠,٠٦٤	٠,٣٥١	٠,٣٣٣	٠,٢٦٣	٠,١٨٧	-١,١١٣	-٠,٤٧٢	٣٥
٦	٨	*١٣,٣	٤,٥	٠,٠٩٩	٠,٠٦٢	٠,٤٦٠	٠,٣٥١	٠,٤٦٤	٠,١٦٧	-٢,٥٧٠	-٠,٣١٠	٣٦
٤	٨	٤,٧	١١,٩	٠,١٠٠	٠,٠٦٤	٠,٣١١	٠,٣٣٢	١,٣٩٦	٠,١٩٤	-٤,٥٨٠	-٠,٥٤٢	٣٧
٧	٦	٤,٤	٠,٩٨٠	٠,٠٦٦	٠,٠٩٩	٠,٣٢٧	٠,٣٨٥	٠,٢٧٢	٠,٨١٩	-١,٠٩٨	-٣,٤٦٧	٣٨
٧	٧	٦,٣	٣,٧	٠,٠٧١	٠,٠٦٢	٠,٣٠٠	٠,٣٣٥	٠,٦٤٣	٠,١٦٦	-٢,٧٧٤	-٠,١٧٣	٣٩
٨	٧	١٣,٣	١١,١	٠,٠٦٧	٠,٠٧٠	٠,٣٢٩	٠,٢٧٣	٠,٣٠٣	٠,٨٤٨	-١,٢٢٩	-٣,٣٦٨	٤٠
٧	٨	٢,٩	٦,٤	٠,٠٦٩	٠,٠٦٥	٠,٢٨٤	٠,٣٣٨	٠,٤٢٣	٠,١٩٢	-١,٦٧٠	-٠,٥٦٨	٤١
٤	٧	٠,٥	٢,٩	٠,٠٨٥	٠,٠٨١	٠,٣٥٢	٠,٤١٩	٠,٦٥٧	٠,٣٢٦	-٢,٨٨٩	-١,٧٩٥	٤٢
٦	٦	٠,٩٠٨	٤,١	٠,١٦٨	٠,٠٨٣	٠,٥٥٠	٠,٣٨٩	٠,٧١٩	٠,٤٦٦	-٢,٩٦٣	-٢,٣٦٧	٤٣
٨	٨	*١٤,٩٣	٤,٩٦٣	٠,٠٦٢	٠,٠٥٧	٠,٢٩٦	٠,٢٧٨	٠,٠١٤	٠,٤٥٠	٠,٢٠٤	-٢,٠٤٦	٤٤
٧	٨	٤,٣	٣,٣	٠,٠٧٤	٠,٠٦٩	٠,٣٥٨	٠,٣٣٥	٠,٣٢٥	٠,٣١٩	-١,٥٣٠	-١,٤٨٦	٤٥
٦	٩	٢,٣	٦,٦	٠,٠٨٢	٠,٠٥٠	٠,٣٥٨	٠,٢٣٩	٠,٥٩٣	٠,٢٤٣	-٢,٧٥٧	٠,٢٩٣	٤٦
٨	٣	٨,٦	٢,٥	٠,٠٥٥	٠,٢١٣	٠,٢٦٧	٠,٥٥٨	٠,٢٨٣	١,٢٠٨	-٠,٨٥٤	-٣,٨٠٥	٤٧
٦	٧	٦,٣	**١٨	٠,٠٧٤	٠,٠٨٥	٠,٣٧٥	٠,٣٣٧	٠,٢٣٣	٠,٦٦٦	-٠,٩٨٣	-٢,٨٤٠	٤٨
٩	٨	١٢,١	٧,٧	٠,٠٦٣	٠,٠٧٠	٠,٣٠٧	٠,٣٨٤	٠,٣٣٦	٠,٠٧٠	١,٤١٨	٠,١٥٥	٤٩
٨	٧	١٢,٤	٨,٤	٠,٠٥٩	٠,٠٧٤	٠,٢٨٤	٠,٣٣٣	٠,٢١٦	٠,٤٩٦	٠,٢٣٢	-٢,٢٩٠	٥٠
٦	٦	٠,٤١٨	٨,٨	٠,٠٩٥	٠,١١١	٠,٣٢٣	٠,٤٣٥	١,٢٣٦	٠,٧١٧	-٤,٤٢٦	-٣,١٥٦	٥١
٧	٨	١,٨	١٠,٨	٠,٠٦٧	٠,٠٦١	٠,٢٩٤	٠,٣٣٥	٠,٥٠٢	٠,٢٥٠	-٢,٦٧	-١,٠٥١	٥٢
٦	٨	٩,٤٢٦	١٠,٤٢٦	٠,٠٧٠	٠,٠٧٠	٠,٣٣٢	٠,٣٢٩	٠,٦٨٥	٠,٣٨٥	-٣,٣١٤	-١,٨٠٧	٥٣
٦	٨	٢,٩	٢,٦٧	٠,١٠٣	٠,٠٥٩	٠,٤٠٦	٠,٣١١	٠,٧٠٣	٠,١٧٣	-٣,٠٧٢	-٠,١٨٦	٥٤
٨	٨	٣,٨	١٠,٦	٠,٠٦٥	٠,٠٥٨	٠,٣١٨	٠,٣٢١	٠,١٩٦	٠,٢١٠	-٠,٢٨٢	-٠,٦٤٤	٥٥
٦	٧	٤,٦	٢,٢	٠,٠٧٦	٠,٠٨٦	٠,٣٠٣	٠,٣٦٧	٠,٧٢٦	٠,٦١٠	-٢,٩٩٤	-٢,٧٩٥	٥٦
٨	٨	٣,٤	١٢,٠١	٠,٠٦٢	٠,٠٥٥	٠,٣٠٣	٠,٣٠٢	٠,٢٥٤	٠,٢٥١	-٠,٨٥١	-٠,٩٦٢	٥٧

درجات الحرية	مربع كاي وتعتمد الدلالة على درجات الحرية لكل نموذج	الخطأ المعياري لعمليات التمييز		معاملات التمييز SLOPE		الخطأ المعياري لعمليات الصعوبة E.S		معاملات الصعوبة THRESHOID		الخاصية رقم المادة		
٨	٧	٩,٩٩٤	٦,٣	٠,٠٦٥	٠,٠٧٠	٠,٢٩٦	٠,٣٣٦	٠,٣٩٦	٠,٤٥٠	-١,٦٥٥	-٢,١٨٢	٥٨
٦	٧	٤,٠١	٩,٨٥١	٠,٠٩٥	٠,٠٩٧	٠,٤٠٦	٠,٥٢٩	٠,٥٤٥	٠,٢٤١	-٢,٥٨٧	-١,٥٠٠	٥٩
٨	٥	٧,٦٥٥	٢,٢	٠,٠٦٦	٠,١٢٣	٠,٣٧٩	٠,٤١٦	٠,١٩٧	١,٠٥٣	٠,٦٨٥	-٣,٩٢٦	٦٠
٩	٩	١٣,٧	١٢,٩	٠,٠٣١	٠,٠٣٨	٠,١١٤	٠,١٥٠	٠,٨٢٠	٠,٤٩١	٢,٣٠٣	١,٢٥٧	٦١
٩	٦	١٣,٥١	٩,٢	٠,٠٣٨	٠,٠٥٢	٠,١٤٥	٠,١٧٧	٠,٦٣١	٢,٠٤١	١,٨٢٦	-٦,٦٩٠	٦٢
٨	٨	**٢١,٤	٨,٩	٠,٠٣٩	٠,٠٥٨	٠,١٤٢	٠,٢٢٠	٠,٧١٢	٠,٨٨٠	-٢,١٣٢	-٣,٣٢٨	٦٣
٨	٩	١٢,٣	*١٦,٤	٠,٠٤١	٠,٠٤٠	٠,١٣٨	٠,١٤٤	٢,١٢٨	١,٤٥٢	-٦,٩١٣	-٤,٩٥١	٦٤
٧	٨	١٢,٧	٦,٣	٠,٠٣٧	٠,٠٥٧	٠,١٢٩	٠,٢٢٤	١,٩٧٣	٠,٦٢٥	-٦,٥٠٩	-٢,٣٧٧	٦٥
٨	٩	١٢,١	٨,٣	٠,٠٤٨	٠,٠٤٢	٠,١٦٨	٠,١٦٩	١,١٧٦	٠,٠٦٩	-٣,٩٧٤	٠,٣٢٩	٦٦
٧	٦	٧,٥١	١,٥١	٠,٠٤٧	٠,٠٧٩	٠,١٦٢	٠,٢٦٣	١,٤١٢	١,٤٩٥	-٤,٦٩٨	-٥,٠٥٥	٦٧
٨	٩	١٤,٣١	٩,٩١	٠,٠٣٧	٠,٠٣٩	٠,١٣٣	٠,١٤٣	١,٠٩٩	١,٣٤٤	-٣,٥٤٣	-٤,٥٩١	٦٨
٨	٩	**٢٠,٢	١٤,٦	٠,٠٣٤	٠,٠٤٨	٠,١١٧	٠,١٩٣	١,٨١٣	٠,٤٠٨	-٥,٨٨٢	-١,٢٢٥	٦٩
٨	٧	**٢٨,٢	٨,٧	٠,٠٣٤	٠,٠٥٥	٠,١٢٠	٠,١٨٦	١,٦٨٩	١,٧٤٩	-٥,٥١٠	-٥,٨٢٢	٧٠
٧	٩	٩,٥	**٢٢,٧	٠,٠٥٠	٠,٠٣٥	٠,١٦٦	٠,٥١٧	١,٧٧٩	٠,٨٧١	-٥,٨٥٣	-٢,٧٥٥	٧١
٩	٨	*١٧,٤	٨,٦٧	٠,٠٤٠	٠,٠٦٢	٠,١٤٩	٠,٢٤٥	٠,٥٧٦	٠,٦٦٦	-١,٥٨٧	-٢,٦٢٦	٧٢
٩	٧	١٤,٦	١٠,٦	٠,٠٣٤	٠,٠٤٣	٠,١٢٤	٠,١٥١	٠,٦٧٣	١,٩٥٧	-١,٧٤٧	-٦,٦٢٥	٧٣
٥	٩	*١١,٦	١٢,٢	٠,٠٦٠	٠,٠٣٨	٠,١٨٨	٠,١٤٣	٢,٤٦٠	٠,٧٤٣	-٧,٤٤٩	-٢,٣٨٧	٧٤
٨	٩	*١٨,٤	**٢٤,٨	٠,٠٣٦	٠,٠٣٨	٠,١٣٠	٠,١٤٣	٠,٨٧٣	٠,٨٤٩	-٢,٦٧٦	-٢,٨٢٤	٧٥
٨	٩	**٢٩,٨	**١٨,٨	٠,٠٣٦	٠,٠٣١	٠,١٢٤	٠,١١٧	١,٧١٦	٠,٧٦٧	-٥,٥٧١	-٢,٢٢٢	٧٦
٨	٨	٦,٧٢	٦,٦١	٠,٠٤٢	٠,٠٥٦	٠,١٥٧	٠,٢١٨	٠,٦٠٩	٠,٨١٥	-١,٨٢٤	-٣,١١٩	٧٧
٩	٨	١٤,٢	**١٨,١	٠,٠٣٣	٠,٠٤٩	٠,١٢٠	٠,١٨٠	٠,٥٠٦	٠,٧٥١	-٠,٥٦٧	-٢,٦١٢	٧٨
٨	٨	**١٦,٨	٦,١	٠,٠٤٤	٠,٠٥٨	٠,١٧٣	٠,٢٣٠	٠,٦٦٧	٠,٦٩٨	٢,٢٠٥	-٢,٧٠٠	٧٩
٨	٩	*١٤,٨	**٢٦,٨	٠,٠٣٣	٠,٠٣٥	٠,١١٧	٠,١٣٠	٢,١٥٢	٠,٥٥٢	-٧,٠٩٢	-١,٣٣٩	٨٠
٨	٩	٩,٨	*١٦	٠,٠٤٢	٠,٠٣٦	٠,١٥٦	٠,١٣٠	٠,٥٢٣	١,١١٤	-١,٤٠٠	-٣,٧١٢	٨١
٨	٨	٦,٧١	١٢,٦١	٠,٠٤٣	٠,٠٥٠	٠,١٦٥	٠,١٨٤	٠,٠٩٨	٠,٩٩٧	-٠,٣٥٤	-٣,٥٥٦	٨٢
٩	٨	**٢٨,٤	٩,٧	٠,٠٣٦	٠,٠٦٠	٠,٠٩٤	٠,٢٢١	٠,٩٠٠	١,٠٦٢	-٢,٢٩٣	-٣,٩٣٠	٨٣
٧	٩	٧,٦	١٠,٣	٠,٠٤١	٠,٠٣٨	٠,١٤٤	٠,١٥١	١,٤٠٧	٠,٤٩٦	-٤,٦٥٣	-١,٣٠٤	٨٤
٦	٩	٤,٦	١١,٣	٠,٠٤٨	٠,٠٣٦	٠,١٦٣	٠,١٣٣	١,٧٩٥	١,٠٨٠	-٥,٩١٠	-٣,٦٤٠	٨٥
٥	٧	**١٢,٥	٦,٥١	٠,٠٤٨	٠,٠٦٥	٠,١٥٨	٠,٢١٨	٢,٢٨٣	١,٥٥٥	-٧,١٩٨	-٥,٢٥٥	٨٦
٨	٨	٧,٩١	**٢٣,٦	٠,٠٤١	٠,٠٤٠	٠,١٥٥	٠,١٣٩	٠,٤٠٠	١,٦٢٠	٠,٥٠٥	-٥,٣٩٠	٨٧
٨	٩	*١٤,٦	١٣,٦	٠,٠٣٨	٠,٠٥٤	٠,١٣٤	٠,٢٥٩	١,٢٩٠	٠,٤٤١	-٤,١٩٧	١,٨٣٥	٨٨
٩	٩	**٢٠,١	١٥,٣	٠,٠٣٩	٠,٠٣٩	٠,١٤٣	٠,١٤٥	٠,٦٦٨٨	١,٠٨٩	-١,٩٥٦	-٣,٧٣٦	٨٩

درجات الحرية	مربع كاي وتعتمد الدلالة على درجات الحرية لكل نموذج	الخطأ المعياري لمعاملات التمييز		معاملات التمييز SLOPE		الخطأ المعياري لمعاملات الصعوبة E.S		معاملات الصعوبة THRESHOID		الخاصية	رقم المفردة	
		٠,٠٣٥	٠,٠٤٧	٠,١٢٧	٠,١٧٥	١,١١٦	٠,٧٤٢	-٣,٦٠٧	-٢,٥٧٠			
٨	٨	١٢,٥	**٢١,٧	٠,٠٣٥	٠,٠٤٧	٠,١٢٧	٠,١٧٥	١,١١٦	٠,٧٤٢	-٣,٦٠٧	-٢,٥٧٠	٩٠
٨	٩	*١٥,٨	١٤,١	٠,٠٣٨	٠,٠٤٥	٠,١٤٢	٠,١١٨	٠,٤٥٦	٠,٧٣٨	-٠,٧٧٥	-٢,٧٠٠	٩١
٦	٩	**١٦,٦	٧,٦	٠,٠٥٦	٠,٠٤٤	٠,١٩٦	٠,١٧٦	١,٣٧٤	٠,٥٠٥	-٤,٧٢٥	-١,٦٠١	٩٢
٦	٨	١٢,٢	١١,٢	٠,٠٤٦	٠,٠٥٣	٠,١٥٥	٠,١٩٢	٢,٢٩٤	١,٠٩٣	-٧,٤٨٠	-٣,٨٦٨	٩٣
٧	٩	٤,٠١	٨,٠١	٠,٠٦٣	٠,٠٤٨	٠,٢٣١	٠,١٨٠	١,٠٩٩	٠,٩٤٢	-٣,٩٧٣	-٣,٣٥٩	٩٤
٨	٦	١١,٣	٤,٣	٠,٠٤٣	٠,٠٩٤	٠,١٦١	٠,٣١١	٠,٤٥٣	١,٢٢٨	-١,٠٥٩	-٤,٢٧٣	٩٥
٨	٦	١٣,٨٥١	٣,٨٥١	٠,٠٤٥	٠,٠٧٣	٠,١٦٨	٠,٢٣٥	٠,٧٣٠	١,٥٩٥	-٢,٤٢٨	-٥,٢٦٤	٩٦
٨	٦	٧,٤	٧,٢	٠,٠٣٦	٠,٠٥٥	٠,١٣١	٠,١٨٩	٠,٩٨٠	١,٨٤٣	-٣,١٤٥	-٦,١٥٤	٩٧
٩	٩	**٣٣,٩	٩,٣	٠,٠٣٠	٠,٠٤١	٠,١٠٧	٠,١٦٤	٠,٨٧١	٠,٣٥٩	-٢,٤٠٦	٠,٤٦١	٩٨
٩	٦	٩,٧	**١٤,٢	٠,٠٣٨	٠,٠٤٩	٠,١٤٢	٠,١٦٩	٠,٥٥٦	١,٦٠٥	-١,٤٠٣	-٥,٣٤١	٩٩
٩	٩	*١٦,٧	*١٦,٥	٠,٠٣٣	٠,٠٣٥	٠,١١٨	٠,١٣٥	٠,٥٩٥	٠,٤٢٤	-١,٢٢٢	-٠,٤٥٨	١٠٠

** مستوى الدلالة عند (٠,٠١)

* مستوى الدلالة عند (٠,٠٥)

يُلاحظ من الجدول رقم (٧) أن:

- معاملات الصعوبة انحصرت في الفترة [- ٦,٦٩ ، ١,٨٣٥] لوجيت، وأقل معامل صعوبة كان للمفردة رقم (٦٢)، وأعلى معامل صعوبة كان للمفردة رقم (٨٨) وذلك في نموذج (أ) المرتب بتسلسل طبقاً للمحتوى الذي تم تدريسه. بينما معاملات الصعوبة انحصرت في الفترة [- ٧,٤٨ ، ٢,٣٠] لوجيت، وأقل معامل صعوبة كان للمفردة رقم (٩٣)، وأعلى معامل صعوبة كان للمفردة رقم (٦١) وذلك في نموذج (ب) العشوائي.
- الأخطاء المعيارية لمعاملات الصعوبة انحصرت في الفترة بين [٠,٠٧ ، ٢,٠٤]، وأقل خطأ معياري لمعامل الصعوبة كان للمفردة رقم (٦٦)، وأعلى خطأ معياري كان للمفردة رقم (٦٢) وذلك في نموذج (أ) المرتب بتسلسل طبقاً للمحتوى الذي تم تدريسه. بينما في نموذج (ب) المرتب عشوائياً فإن الأخطاء المعيارية لمعاملات الصعوبة انحصرت في الفترة [٠,٠١ ، ٢,٤٦]، وأقل خطأ معياري لمعامل الصعوبة كان للمفردة رقم (٤٤)، وأعلى خطأ معياري لمعامل الصعوبة كان للمفردة رقم (٧٤).
- معاملات التمييز انحصرت في الفترة [٠,١٢ ، ٠,٥٦]، وأقل معامل تمييز كان للمفردة رقم (٧٦)، وأعلى معامل تمييز كان للمفردة رقم (٤٧) وذلك في نموذج (أ) المرتب بتسلسل

طبقا للمحتوى الذى تم تدريسه. بينما في نموذج (ب) المرتب عشوائياً فإن معاملات التمييز انحصرت في الفترة [٠,٠٩ ، ٠,٥٥]، وأقل معامل تمييز كان للمفردة رقم (٤٣)، وأعلى معامل تمييز كان للمفردة رقم (٨٣).

■ الأخطاء المعيارية لمعاملات التمييز انحصرت في الفترة بين [٠,٠٣ ، ٠,٢١]، وأقل خطأ معياري لمعامل التمييز كان للمفردة رقم (٧٦)، وأعلى خطأ معياري كان للمفردة رقم (٤٧). وذلك في نموذج (أ) المرتب بتسلسل طبقا للمحتوى الذى تم تدريسه. بينما في نموذج (ب) المرتب عشوائياً فإن الأخطاء المعيارية لمعاملات التمييز انحصرت في الفترة [٠,٠٣ ، ٠,١٧]، وأقل خطأ معياري لمعامل التمييز كان للمفردة رقم (٨٣)، وأعلى خطأ معياري لمعامل التمييز كان للمفردة رقم (٤٣).

■ عدد المفردات التي لها دالة مميزة (٨٨) مفردة تزيد قيم χ^2 المحسوبة عن قيمة χ^2 الجدولية عند مستوى دلالة (٠,٠١) أى غير دالة إحصائية وتختلف الدلالة حسب درجات الحرية لكل مفردة على حده باستخدام النموذج ثنائي المعلم وذلك في نموذج (أ) المرتب بتسلسل طبقا للمحتوى الذى تم تدريسه. بينما في نموذج (ب) المرتب عشوائياً فإن عدد المفردات التي لها دالة مميزة (٨٠) مفردة تزيد قيم χ^2 المحسوبة عن قيمة χ^2 الجدولية عند مستوى دلالة (٠,٠١) أى غير دالة إحصائية.

■ عدد المفردات التي ليس لها دالة مميزة (غير مطابقة) (١٢) مفردة تقل قيم χ^2 المحسوبة عن قيمة χ^2 الجدولية عند مستوى دلالة (٠,٠١) أى دالة إحصائية وذلك في نموذج (أ) المرتب بتسلسل طبقا للمحتوى الذى تم تدريسه. بينما في نموذج (ب) المرتب عشوائياً فإن عدد المفردات التي ليس لها دالة مميزة (غير مطابقة) (٢٠) مفردة تقل قيم χ^2 المحسوبة عن قيمة χ^2 الجدولية عند مستوى دلالة (٠,٠١) أى دالة إحصائية وتختلف الدلالة حسب درجات الحرية لكل مفردة على حده باستخدام النموذج ثنائي المعلم.

وأشارت النتائج أيضاً إلى أن متوسط صعوبة الاختبار ككل نموذج (أ) = (١,٦٠٣)، والخطأ المعياري لمعامل صعوبة الاختبار ككل نموذج (أ) = (١,٠٣٣)، ومعامل تمييز الاختبار ككل نموذج (أ) باللوجيت = (٠,٥٣٣) والخطأ المعياري لمعامل تمييز الاختبار ككل نموذج (أ) باللوجيت = (٠,٤٧٤). بينما متوسط صعوبة الاختبار ككل نموذج (ب) = (١,٩٤٠)، والخطأ المعياري لمعامل صعوبة الاختبار ككل نموذج (ب) = (١,٦٠١)، ومعامل

تميز الاختبار ككل نموذج (ب) باللوغيت = (0,512) والخطأ المعياري لمعامل تمييز الاختبار ككل نموذج (ب) باللوغيت = (0,344).

ويتضح من جدول رقم (٧) السابق أنه يوجد اختلاف بين نموذجي الاختبار في صعوبة وتمييز المفردات باستخدام النموذج ثنائي البارامتر. ولكي يتم دراسة الفروق بين نموذجي الاختبار في معالم المفردات بشكل أدق بعد حذف المفردات غير الملائمة للنموذج ثنائي البارامتر تم استخدام اختبار (ت) للعينات المستقلة لدراسة تلك الفروق من خلال برنامج (SPSS) وكانت النتائج كالتالي:

جدول (٨)

نتائج اختبار (ت) لدلالة الفرق في قيم صعوبة وتمييز المفردات والخطأ المعياري لتقدير صعوبة وتمييز كل منهما بين نموذجي الاختبار

المتغيرات	النموذج العشوائي (B)			النموذج المرتب معرفياً (A)		
	الانحراف المعياري	المتوسط الحسابي	عدد المفردات	الانحراف المعياري	المتوسط الحسابي	عدد المفردات
معامل صعوبة المفردات	1,827	2,650	80	1,584	2,396	88
تقدير الخطأ المعياري لصعوبة المفردة	0,509	0,737	80	0,421	0,628	88
معامل تمييز المفردات	0,107	0,273	80	0,109	0,296	88
تقدير الخطأ المعياري لتمييز المفردة	0,0251	0,0657	80	0,0244	0,0659	88

يتضح من الجدول رقم (٨) أنه لا توجد فروق دالة إحصائية عند مستوى (0,01) أو (0,05) بين نموذجي الاختبار (A, B) في تقديرات كل من صعوبة المفردة وتمييز المفردة والخطأ المعياري لصعوبة المفردة والخطأ المعياري لتمييز المفردة.

وتختلف هذه النتيجة مع نتائج بحث حسان العمري (2011)، وبحث عدنان اللحياني (2010) التي أشارت إلى أن معاملات الصدق والثبات والصعوبة والتمييز للاختبار ومفرداته تختلف باختلاف ترتيب مفردات الاختبار، وتتفق نتيجة البحث الحالي مع نتائج بحث خالد سليمان (2015) وسناء عباينة (2017) التي أظهرت أنه لا يوجد اختلاف في معاملات الصعوبة أو التمييز ترجع إلى ترتيب مفردات الاختبار.

ربما يعود ذلك إلى أن الفقرات في نموذجي الاختبار تحمل نفس المحتوى فلا يتأثر معامل الصعوبة لمفردات الاختبار بأنماط ترتيب الفقرات، حيث إن المستويات العقلية التي تقيسها الفقرات هي واحدة أيضا بغض النظر عن الترتيب، وكل فقرة تحتاج إلى نفس القدرات العقلية للإجابة عنها بغض النظر عن موقعها في الاختبار. كما يمكن تفسير أنه لا توجد فروق دالة إحصائية بين نموذجي الاختبار في معامل تمييز مفردات الاختبار إلى أن معلّمة تمييز المفردة تقسم الطلاب إلى متمكنين وغير متمكنين وبالتالي تعتمد إجابة الطالب في مادة الفروق الفردية والقياس النفسي على قدرته العقلية، فالطالب المتقن للمادة يكون متمكنا من المحتوى ويستطيع الإجابة على أية مفردة بغض النظر عن موقعها بالاختبار، وعلى النقيض من ذلك الطالب غير المتمكن. وبالتالي موقع أو ترتيب المفردات ليس له تأثير واضح على معلّمة تمييز مفردات اختبار مقرر الفروق الفردية والقياس النفسي.

نتائج الفرض الثالث وتفسيرها ومناقشتها:

ينص الفرض الثالث على أنه "لا يوجد تأثير دال إحصائياً لترتيب مفردات الاختبار طبقاً للمحتوى المعرفي (متسلسل أو عشوائي) في الأداء الأكاديمي لطلاب الدبلومة العام في التربية.

ولاختبار هذا الفرض يتم استخدام اختبار (ت) للعينات المستقلة من خلال برنامج

Spss25 ويتضح ذلك من الجدول التالي:

جدول (٩)

نتائج اختبار (ت) لدلالة الفروق بين متوسطي درجات الطلاب في نموذجي الاختبار (أ، ب) في الأداء الأكاديمي

درجات الاختبار التحصيلي	ترتيب المفردات	العدد	المتوسط	الانحراف المعياري	قيمة (ت) ودلالاتها
الدرجة الكلية	نموذج أ	٤٢٥	٧٠,٤٦	٧,٦٠	*٢,١٨٤
للاختبار التحصيلي	نموذج ب	٤٠٧	٦٩,٢٨	٦,٨٥	دالة إحصائياً عند مستوى (٠,٠٥)

يتضح من الجدول رقم (٩) السابق أنه توجد فروق دالة إحصائياً بين متوسطي درجات

طلاب النموذج (أ) المرتب معرفياً، وطلاب النموذج (ب) المرتب عشوائياً في الأداء الأكاديمي

لصالح طلاب النموذج (أ) المرتب معرفياً. وتتفق هذه النتيجة مع نتائج ما توصل إليه كل من سناء عبابنة (٢٠١٧)، ورمضان رمضان (٢٠١٠)، وDoerner & Calhoun (2009) من وجود تأثير دال إحصائياً لنماذج ترتيب فقرات الاختبار على الأداء الأكاديمي والخصائص السيكومترية للاختبار، وتختلف مع ما توصل إليه بحث كل من Kennedy & Butter (2013)، وخالد سليمان (٢٠١٥) من أنه لا يوجد تأثير دال إحصائياً لنماذج ترتيب فقرات الاختبار على الأداء الأكاديمي والخصائص السيكومترية للاختبار.

ويمكن تفسير هذه النتيجة من خلال أن الطلاب عندما استقبلوا أسئلة الاختبار التحصيلي المرتب بتسلسل طبقاً للمحتوى الذي درسوه شعروا بالارتياح والاطمئنان وأن المعلومات يتم استرجاعها بنفس التنظيم والترتيب الزمني الذي تم تخزينها فيه وأصبح من السهل عليهم الحصول على درجات مرتفعة وهذا يتماشى مع قوانين تنظيم المجال الإدراكي عند النظرية المجالية. وعلى النقيض من ذلك أسئلة الاختبار التحصيلي المرتب عشوائياً يجد الطلاب صعوبة في استرجاع معلوماته.

وبالتالي يعتبر ترتيب مفردات الاختبار في النموذج (أ) المرتب بتسلسل معرفي يحقق أداء تحصيلي أعلى من مفردات النموذج (ب) المرتب عشوائياً، وهذا يتفق مع طبيعة الطلاب بناء على المعيشة معهم أثناء تدريس المحتوى حيث إنهم يفضلون أن تكون الأسئلة مرتبة طبقاً لترتيب المقرر الذي درسوه وهذا النوع من الاختبار المرتب بتسلسل معرفي أشبع ميولهم ولهذا زادت درجة الطلاب على الاختبار، وعلى النقيض من ذلك مفردات الاختبار المرتبة بعشوائية يجد الطلاب فيها صعوبة في اشباع رغباتهم ويشعرون بالتشتت واستغراق الوقت في استرجاع المعلومة المخزنة ولذا كانت درجاتهم التحصيلية منخفضة.

ربما قد يعود ذلك إلى أن الترتيب حسب المحتوى المعرفي يكون محسوباً بطريقة دقيقة بما يحقق التدرج في الأسئلة بشكل منطقي ويعمل على زيادة دافعية الطالب لإعطاء أفضل إجابة لديهم، مما يؤدي إلى زيادة قدرته على مواصلة الإجابة حتى نهاية فترة الاختبار، وتساوده على زيادة قدرته على الضبط النفسي عند مواجهة أي مشكلة في الاختبار، كما يساعد ترتيب الفقرات على التهيؤ العقلي للطالب مما يجعل الاختبار يقيس قدرة الطالب على التحصيل بدرجة مقبولة من الدقة أو بأقل خطأ، بمعنى أن يكون أكثر ثباتاً. وهذا يوفر للطالب الدافعية للاستمرار في محاولة للإجابة عندما يشعر بالتعزيز الفوري بسبب قدرته على الإجابة عن الأسئلة الأولى التمهيدية من الاختبار والتي تسمى بالأسئلة التشجيعية أو الأسئلة الماصة

للصدمة. على العكس من النموذج العشوائي التي يجد الطالب صعوبة في استرجاع معلوماته نظرًا لاختلاف طريقة التخزين عن طريق الاسترجاع.

التوصيات:

- يمكن الوصول من هذا البحث إلى مجموعة من التوصيات منها:
- ١- تبصرة العاملين في مراكز التقويم وإعداد الامتحانات بنتائج البحث من أجل إعداد بنوك أسئلة في المقررات الدراسية وصياغة صور متكافئة للاختبار عن طريق اختلاف النماذج.
 - ٢- عمل دورات تدريبية على برامج تحليل مفردات الاختبار مثل برنامج Bilog-mg3
 - ٣- عقد ورش عمل لمعدي الامتحانات عن كيفية إعداد اختبار تحصيلي مرجعي المحك من أجل الوصول إلى مدى تحقق الأهداف السلوكية للمقرر الدراسي.
 - ٤- ضرورة اهتمام القائمين على إعداد الاختبارات بمستوى تحقيق الأهداف عند صياغتهم مفردات الاختبار، ومراعاة المفردات مستقلة عن بعضها، بمعنى أكثر تحديداً، أن تكون الإجابة عن فقرة ما غير مرتبطة بالإجابة عن فقرة أخرى.

البحوث المقترحة:

- يمكن الوصول من هذا البحث إلى مجموعة من البحوث المقترحة منها:
- ١- دراسة عن تأثير ترتيب مفردات الاختبار على الزمن المستغرق في الإجابة على الاختبار
 - ٢- أثر نوع المفردة على الخصائص السيكمترية للاختبار.
 - ٣- دور ترتيب مفردات الاختبار في التقليل من التخمين.
 - ٤- دراسة عن تأثير ترتيب مفردات الاختبار على الخصائص السيكمترية للاختبار ومفرداته في ضوء نظريتي القياس التقليدية والحديثة.
 - ٥- أثر نوع المفردة على دقة نتائج الاختبارات مرجعية المحك.

المراجع

- أبو المجد إبراهيم الشوربجي، عزت عبدالحميد محمد (٢٠٠٥). أسلوب مقترح لاتخاذ قرارات الأتقان في ضوء مفهوم درجة القطع وثبات الاختبارات التحصيلية مرجعية المحك. مجلة كلية التربية، جامعة طنطا، (٣٤)، ص ص ٥٥٦ - ٥٩٧.
- أبو المجد إبراهيم الشوربجي، وعزت عبد الحميد محمد حسن (٢٠١٢). القياس والإحصاء التربوي والنفسي، الرياض: مكتبة الرشد.
- أحمد الرفاعي غنيم (١٩٨٥). تطبيقات على ثبات الاختبار، القاهرة، مكتبة نهضة الشرق.
- أحمد الرفاعي غنيم، ونصر محمود صبرى (١٩٩٩). التحليل الإحصائي للبيانات النفسية والتربوية باستخدام SPSS، القاهرة، دار قباء للطباعة والنشر.
- أحمد عودة (٢٠٠٥). القياس والتقويم في العملية التدريسية، عمان الأردن: دار الأمل.
- إسماعيل حسن الوليلي (٢٠٠٢). دراسة سيكومترية مقارنة لبعض نماذج الاستجابة للمفردة في انتقاء مفردات الاختبارات مرجعية المحك. رسالة دكتوراه غير منشورة، كلية التربية، جامعة الأزهر.
- أحمد بوزيان تيغزة (٢٠١٢). التحليل العاملي الاستكشافي والتوكيدي، مفاهيمها ومنهجيتها بتوظيف حزمة Spss وليزرل LISREL. ط١، دار المسيرة للنشر والتوزيع، عمان: الأردن.
- أمين علي محمد سليمان، رجاء محمود أبو علام (٢٠١١). القياس والتقويم في العلوم الإنسانية: أسسه وأدواته وتطبيقاته، القاهرة: دار الكتاب الحديث.
- أمينة محمد كاظم (١٩٨٨). استخدام نموذج راش في بناء اختبار تحصيلي في مادة علم النفس لطلاب كلية الآداب بجامعة الكويت وتحقيق التفسير الموضوعي للسلوك، في: أنور الشرقاوي، سليمان الخضري الشيخ، أمينة محمد كاظم، نادية محمد عبدالسلام (١٩٩٦) (محررون): اتجاهات معاصرة في القياس والتقويم النفسي والتربوي. القاهرة: مكتبة الأنجلو المصرية.
- أيمن عودة الحجاج (٢٠١٣). أثر ترتيب الفقرات حسب خصائص السيكمترية على أداء الأفراد والخصائص السيكمترية للاختبار، رسالة ماجستير، جامعة مؤتة، عمادة الدراسات العليا، الأردن.

إيهاب محمد نجيب السيد (٢٠١٢). معادلة الاختبارات وعلاقتها ببعض المتغيرات السيكومترية- دراسة محاكاة وتطبيقية. رسالة دكتوراه غير منشورة، كلية التربية، جامعة عين شمس.

باسل خميس أبو فودة، نجاتي أحمد بني يونس (٢٠١٢). الاختبارات التحصيلية (مفهومها وكيفية إعدادها وأسس بنائها وتكوينها وتطبيقات ميدانية). الطبعة الأولى، الأردن: دار المسيرة للنشر والتوزيع.

حسان العمري (٢٠١١). أثر ترتيب الفقرات في المقاييس النفسية على التجانس الداخلي للمقياس وخصائص فقراته. مجلة جامعة دمشق، ٢٧ (٣)، ص ١١١-١٢٩.

حمدي يونس أبو جراد (٢٠١٢). الخصائص السيكومترية لاختبارات الاختيار من متعدد والتكميل: دراسة مقارنة باستخدام نظرية الاستجابة للمفردة، مجلة العلوم التربوية والنفسية بالبحرين، ٣١ (٣)، ص ٣٧٥-٤٠٤.

خالد شاهر سليمان (٢٠١٥). أثر ثلاثة نماذج لترتيب فقرات اختبار تحصيلي في التقييم التربوي على الخصائص السيكومترية للاختبار وفقراته. مجلة كلية التربية والآداب - جامعة تبوك، السعودية.

رمضان صالح رمضان (٢٠١٠). أثر ترتيب فقرات الاختبار حسب درجة صعوبتها على تحصيل طلاب المرحلة الثانوية لمادة الرياضيات. مجلة التربية للبحوث التربوية والنفسية والاجتماعية، كلية التربية، جامعة الأزهر، (٨)، ص ١٥٧-١٧٦.

سامية سمير شحاتة (٢٠١٢). دروس في القياس النفسي والتربوي، القاهرة: دار إيتراك للنشر والتوزيع. سناء حسين عبابنة (٢٠١٧). أثر ترتيب فقرات اختبار اختيار من متعدد تبعا لعدد خطوات حلها على الخصائص السيكومترية للاختبار وفقراته وأداء الطلبة عليه، رسالة ماجستير، كلية التربية، جامعة اليرموك، الأردن.

"صلاح الدين" محمود علام (٢٠٠١). اختبار تشخيصي مرجعي المحك للمهارات الأساسية في بناء الاختبارات التحصيلية (الصيغة أ)، (الصيغة ب)، ودليل الاختبار التشخيصي مرجعي المحك للمهارات الأساسية في بناء الاختبارات التحصيلية. القاهرة، مكتبة الأنجلو المصرية.

"صلاح الدين" محمود علام (٢٠٠٥). نماذج الاستجابة للمفردة الاختبارية: أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي. ط ١، القاهرة: دار الفكر العربي.

"صلاح الدين" محمود علام (٢٠٠٩). الاختبارات التشخيصية مرجعية المحك في المجالات: التربوية والنفسية والتدريبية. القاهرة: دار الفكر العربي.

صلاح أحمد مراد، وأمين على سليمان (٢٠٠٢). الاختبارات والمقاييس في العلوم النفسية والتربوية: خطوات إعدادها وخصائصها، القاهرة، دار الكتاب الحديث.

صلاح شريف عبد الوهاب (٢٠٠١). أثر بعض الطرق الوزنية لتقدير الدرجات على صدق الاختبارات مرجعية المحك ذات الاختيار من متعدد، أكتوبر، مجلة كلية التربية، بنها: ص ص ٢٠٢ - ٢٥٥.

عادل سعد خضر (٢٠٠٧). بنوك الأسئلة بين النظرية والتطبيق، القاهرة، دار السحاب للنشر والتوزيع

عدنان اللحياني (٢٠١٠). أثر ترتيب فقرات الاختبار على صدق وثبات اختبار اختيار من متعدد البدائل في مادة الرياضيات لطلاب المرحلة الثانوية بمكة المكرمة، رسالة ماجستير غير منشورة، جامعة أم القرى بمكة المكرمة.

عزت عبد الحميد حسن (٢٠١٦). الإحصاء النفسي والتربوي: تطبيقات باستخدام برنامج Spss 18. القاهرة: دار الفكر العربي.

عماد غصاب عبابنة (٢٠٠٩). الاختبارات محكية المرجع (فلسفتها وأسس تطويرها)، ط ١، الأردن: دار المسيرة للنشر والتوزيع والطباعة.

على ماهر خطاب (٢٠٠١). القياس والتقويم في العلوم النفسية والتربوية والاجتماعية، ط ٢، القاهرة: مكتبة الانجلو المصرية.

فاروق عبدالفتاح على موسى (١٩٩٠). القياس النفسي والتربوي للأسوياء والمعوقين. ط ١، القاهرة: مكتبة النهضة المصرية.

فريال محمود (٢٠١٥). أثر ترتيب فقرات الاختبار ومستوى القلق على أداء الطلاب في اختبار من نوع الاختيار من متعدد، مجلة العلوم التربوية والنفسية، ١٦ (٢)، ص ص ٥٨٩-٦٠٣.

فؤاد البهى السيد (١٩٧٩). علم النفس الإحصائى وقياس العقل البشرى، ط٣، القاهرة، دار الفكر العربى.

محمد محمد "فتح الله" سيد احمد (٢٠٠٥). مدى فاعلية نموذج تعليمي باستخدام الاختبارات مرجعية المحك على بعض المهارات الابتكارية لدى طلاب المرحلة الثانوية. رسالة دكتوراه غير منشورة، كلية التربية، جامعة الأزهر.

محمد مصطفى عليوة (٢٠١٣). طرق تحليل البنود في ضوء النظرية الكلاسيكية ونماذج الاستجابة للمفردة لاختبار مرجعي المحك في مادة العلوم لتلاميذ الصف السادس الابتدائي. رسالة ماجستير غير منشورة، كلية التربية، جامعة الزقازيق.

محمد منصور محمد الشافعي (٢٠١١). تأثير نماذج التحليل اللوغاريتمية (الأحادية والثلاثية والثلاثية) البارامتر على تقدير دوال معلومات المفردات الاختبارية، وعلاقتها بدقة تقديرات القدرة (دراسة مقارنة). المجلة المصرية للدراسات النفسية، ٢١ (٧٣)، ص ٥٤٩ - ٦١٥.

هناء عوض الخرشة (٢٠١٥). أثر ترتيب فقرات الاختبار على دقة معالم الفقرات وقدرات الأفراد والخصائص السيكمترية للاختبار، رسالة ماجستير (غير منشورة)، عمادة الدراسات العليا، الأردن.

هند عبد المجيد الحموري، أحمد إسماعيل الكحلوت (٢٠٠٩). استراتيجية مقترحة لبناء اختبارات تشخيصية محكية المرجع. مجلة العلوم التربوية والنفسية، ١٠ (٢)، ص ١٨٥ - ٢١٤.

Baker, F. (2001). *The Basics of Item Response Theory*. ERIC clearing house on Assessment and Evaluation, University of Masy Land, College Park. M. D.

Barnett, J., Nigg, C., Bourdeaudhuij, B., Maglione, C., Maddock, J., & Jay, M. (2007). The Effect of Item Order on Physical Activity Estimates Using the IPAQ, *Californian Journal of Health Promotion*, 5, 1, Pp. 23-29 .

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

- Chen ,H. (2012). The Moderating Effects of Item Order Arranged by Difficulty on the Relationship between Test Anxiety and Test Performance, *Creative Education*, 3(3),Pp. 328-333.
- Cunningham,G.K.(1986). *Educational and Psychological Measurement*. MacMillan Publishers Company, New York.
- Doerner, W., & Calhoun P. (2009). The impact of the order of test questions in introductory economics., Retrieved from, SSRN :<http://ssrn.com/abstract>.
- Downing , S. M. (2003) . Item Response theory applications of modern test theory in medical education, *The Matric of Medical Education*,37(1), pp.739-745.**
- Duncan,J. M., and W. Kerry Mummery, W.K.(2013). The Effect of Changes to Question Order on the Prevalence of ‘Sufficient’ Physical Activity in an Australian Population Survey Christine Hanley, *Journal of Physical Activity and Health*, 10, Pp.390-396**
- Gronlund. & Linn,R. (1990). *Measurement and Evaluation in Teaching*, New York-Macmillan Publishing, U.S.A.
- Hambleton & Swaminathan, H. (1989). *Item Response Theory :Principles and Applications*. the Netherlands: Kluwer Nijhoft publishing.**
- Henson, R. K. (1999) . Understanding the one – parameter rasch model of item response theory. *Paper Presented At the Annual Meeting of The South West Educational Research Association, San Antonio:pp.1-51.***
- Kennedy. &Butter, A.(2013).Changing the order of mathematics test item: Helping or hindering student performance? *Journal of Humanistic Mathematics*,3(1), pp.21-32.
- Lee, S.H.(2007). Multidimensional Item Response Theory : A sas Mdir empirical study of Piat Math test, Ph.D., University of Macro and Aklahoma, Norman, Aklahoma.
- Lord, F. M.; Stocking, M. L. (1990) . Item Response Theory, Walberg H. J. & Haertel, G. D. (EDs). *The International Encyclopedia of Educational Evaluation*,PP. 281 – 284.**
- McCamy, R.(2002): A primer on the one-parameter Rasch model, *paper presented at the annual meeting of the South West Educational Research Association, Austin,Pp. 1 – 18,ED.462 -440.***

- Neely, D., Springston, R., & McCann, S. (1994). Does item order affect performance on multiple-choice exams? *Teaching of Psychology*, 21(1), PP. 44- 45.
- Owidha, A. A. (2007). **Acomparison of the Rasch model and The Three Parameter Logistic Model Applied to Quantityre Subtest of the General Aptitude Test. Sand Larabla. Ph. D. , University of Denver.**
- Pei-Huachen, B.S. (2005). **IRT- based Automated Test Assembly : A sampling and stratification perspective , Ph.D., University of Texas, Austin, August.**
- Pettijohn, T. & Sacco, M. (2007). Multiple-choice exam question order influences on student performance, Completion time, and Perceptions, *Journal of Instructional Psychology*, 34(3), pp. 142-149.
- Reeve ,B. (2002): **An Introduction to Modern Measurement Theory, Applied Reseach Program ,Division of Cancer Control and Population Sciences . Pp.1-67 .**
- Shawer, F.S. (2011). **Standardized Assessment and Test Construction without Anguish: The complete step by step Guide to test Design Administration, Scoring, Analysis, and Interpretation. Dar El Fikr El Arabi, Cairo, Egypt.**
- Sue, D. (2009). The effect of scrambling test questions on student performance in a small class setting. *Journal for Economic Educators*, 9(1), pp.32-41.
- Yen, M. & Edwardson, S. (1999). **Item Response Theory Approach in Scale Development, Officale. Journal of The Eastern Nursing Research Society and The Western Institute of Nursing**, 48, pp.234 – 238.

Abstract in English

The aim of recent study is to figure out the statistical impact of different distribution in the sequence order of items of Criterion-Referenced Test on Difficulty and Discernment factors of items and test, in terms of Second parameter Model (IRT). Sequence order of items was according to the cognitive content (ordered/serial, ordered randomly).. To achieve the objective of the research, Criterion-Referenced Test (CRT) was prepared in "Individual differences and psychological measurement" course for the university students from the general diploma in education in the faculty of Education- Zagazig University for the academic year 2017/2018. The impact of the difference distribution of the item was revealed in two types of items: MCQ, and True/False (T/F). In its initial form, the test consists of (60) items of 4-choices MCQ type and (40) items of T/F type. Two models of the test (A,B). Items of A model were distributed in order sequence according to arrangement cognitive content, Items of model B were distributed randomly scrambling. Pilot sample was collected to calculate reliability and validity of the test. The total final sample was randomly selected which consists of (858) students; students' ability estimates, their standard error estimates, and psychometric properties of items calculated according to, two-parameter model, where the test prepared in "individual differences and psychological measurement" course and corrected automatically through the scanner system, and then used the appropriate statistical treatments to ensure that the data conformed to the assumptions of the IRT (unidimensionality and local independence,). Most results showed that:1) There is no statistical significance at ($\alpha = .05$) in the estimates of item's parameters, their standard errors, due to the sequence order of items distribution according to content arrangement or random scramble type through two parameters model.(2) there differences in the empirical reliability coefficient (3) There is statistical significance at ($\alpha = .05$) in the academic achievement due to the sequence order of items distribution according to content arrangement or random scramble type in favor of content arrangement type,

Key words: Sequence order of items- Criterion-Referenced Test -Item Response Theory.