أثر التفاعل بين نمط الاختبار وجنس المستجيب على ثبات الاختبار وبعض الخصائص السيكومترية للفقرات

إعداد

أ/ صالح موسى محمد الشمراني

طالب دكتوراة تخصص قياس وتقويم

قسم علم النفس - كلية التربية جامعة ام القرى

أثر التفاعل بين نمط الاختبار وجنس المستجيب على ثبات الاختبار وبعض الخصائص السيكومترية للفقرات

صالح موسى محمد الشمراني
قسم علم النفس، كلية التربية، جامعة أم القرى، مكة المكرمة، المملكة العربية السعودية.
البريد الإلكتروني: zadmath1@gmail.com

ملخص الدراسة:

هدفت الدراسة الحالية إلى التعرّف على نمط الاختبار الأكثر مناسبةً لجنس المستجيب للوصول لاختبار ذو خصائص سيكومترية جيدة ودراسة أثر التفاعل بين نمط الاختبار وجنس المستجيب على الخصائص السيكومترية للاختبار. وقد استخدم الباحث المنهج الوصفي في الإجابة على أسئلة الدراسة. قام الباحث بصياغة اختبار تحصيلي مكوّن من ثلاثة أنماط هي (الاختبار من متعدد، الصواب والخطأ و التكميل) حيث تضمن خمسون فقرة لكل نمط من أنماط الاختبار الثلاثة وعليه تمت الدراسة الاستطلاعية على طلاب الصف ألأول ثانوي في ثانوية المدائن التابعة لتعليم مكة المكرمة وعددهم (81) طالب. وللتحقق من صدق وثبات وجودة مفردات الاختبارات التحصيلية قام الباحث بتطبيق الاختبارات على عينة من الطلاب بواقع 26 طالب لاختبار الاختيار من متعدد و28 طالب لاختبار الصواب والخطأ و27 طالب لاختبار الإكمال وفي ضوء نتائج التحليل الإحصائية أصبح العدد النهائي لمفردات الاختبارات الثلاث (الاختيار من متعدد، والصواب والخطأ، والإكمال) هو 33 مفردة. تكوّنت عينة الدراسة من (446) طالب وطالبة من الصف الأول الثانوي بمدينة مكة المكرمة بواقع (226) طالب و(220) طالبة من 3 مدارس ثانوية حكومية طلاب(ذكور) و3 مدارس ثانوية حكومية طالبات(إناث) . واستخدم الباحث الإحصاءات والأساليب الإحصائية لاختبار صحة الفروض وهي حساب معامل الصعوبة التقليدي لكل مفردة ومعامل التمييز باستخدام معامل الارتباط التسلسلي الأصيل وحساب معاملات ثبات الاتساق الداخلي لكل فقرة كمؤشر للصدق وثبات الاختبار ذلك من خلال طريقة ألفا –كرونباخ والتجزئة النصفية المصححة بمعادلة (سبيرمان- براون- جتمان) واستخدام تحليل التباين الثنائي Two-way ANOVA لدراسة التفاعل و استخدام الإحصائي (M) لحساب الفروق في معاملات الثبات. وقد توصلت الدراسة إلى أنه لا يوجد أثر دال إحصائياً عند مستوى (0,05) للتفاعل بين نمط الاختبار (اختيار من متعدد، الصواب والخطأ، إكمال) وجنس المستجيب (ذكور، إناث) على معامل صعوبة الفقرة. كما توصلت الدراسة إلى أنه لا يوجد أثر دال إحصائيا عند مستوى (0,05) للتفاعل بين نمط الاختبار (اختيار من متعدد، صواب وخطأ، إكمال) وجنس المستجيب (ذكور، إناث) على معامل تمييز الفقرة. في حين توصلت الدراسة لوجود فروق دالة إحصائياً بين معاملات الثبات بطريقة ألفا بين نمط الاختيار من متعدد والصواب والخطأ لصالح نمط الاختيار من متعدد لدى كل من الذكور والإناث.

الكلمات المفتاحية: نمط الاختبار، جنس المستجيب، ثبات الاختبار، الخصائص السيكومترية.

Effect of interaction between item format and gender on test score reliability and the psychometric properties

Saleh Mosa Alshamrani.

Department of Psychology, College of Education, Umm Al-Qura University, Makkah Al-Mukarramah, Kingdom of Saudi Arabia.

High school Math Teacher - Ministry of Education.

E-mail: zadmath1@gmail.com

Abstract:

The aim of this study was to investigate the most suitable item format for respondents (male/female) to achieve better psychometric properties.. In addition, the study aimed at investigation of the effect of interaction between item format and gender on the psychometric properties (item difficulty and discrimination) of the test. The researcher utilized the descriptive method to answer the research questions. The researcher constructed an achievement test consisted of three item formats(multiple choice, true and false ,completion). The test initially consisted of 50 items for each format .Each test form was administered using a pilot sample of (81) students to assess test validity and reliability26 students were used for the multiple choice test 28 students for the true – false test 27 students for the completion test. Afteranalyzing the results the test final forms consisted of 33 items each. The sample size of the study was (446) male/ female secondary – stage students in Makkah:226 male, 220 female(3 schools for each gender). The researcher used the following statistical techniques to test the study's hypotheses:classical item difficulty,item discrimination using biserialcorrelation and test score reliability using α-Cronbach and split-half methods using spearman-brown and Gatman.Interaction effect tested using two-way ANOVA and the M-Statistics to test differences in reliability correlation.

Keywords: item format, respondent gender, test stability, psychometric properties.

مقدمة:

تُعد الاختبارات التحصيلية من أكثر أدوات التقويم شيوعاً في المؤسسات التعليمية التربوية في العالم حيث تُستخدم لقياس قدرات الطلبة تحصيلياً ونستدل من خلالها على قدراتهم المعرفية والمهارية، وبالاعتماد عليها نتنبأ بترتيب الطلبة ضمن الصفوف التي يلتحقون بها ومن خلالها نصف الطلبة ونحدد مستوى النجاح والرسوب ضمن الصف الواحد.

فالاختبارات ذات أهمية في عملية التقويم، وتُعد مؤشراً جيداً تعطينا فكرة عن قدرة وإمكانية الطلبة ومستوى تحصيلهم ونشاطهم، ومن خلال ذلك نتمكن من تصنيفهم ومعرفة مستواهم، حيث نستطيع أن نضع لهم الخطط والأساليب التدريسية المناسبة، ونستثير دافعيتهم للتعلّم، ونجعلهم قادرين على التفاعل في الحصة الصفية. كما يمكن القول إن الاختبارات ضرورية في تحديد إمكانية وقدرات الطلبة التحصيلية، فمن هنا نجد أن الاختبارات تعد الأساس في قياس التحصيل (عبدالهادي، 2001م).

وللاختبارات التي يعدها المعلمون بمختلف أشكالها دور أساسي في تقويم عملية التدريس حيث تستخدم نتائجها في اتخاذ القرارات التربوية، فكلما زادت أهمية القرارات الذي يتوقع اتخاذها، زادت الحاجة إلى أن تكون المعلومات التي يتم الحصول عليها من الاختبار دقيقة، وذات صلة وثيقة بالغرض الذي أُعّد له الاختبار.

والواقع أن الاختبارات تُعطي عملية التعليم مجموعة فوائد، منها تحسين اتخاذ القرارات الخاصة بعملية التدريس، وتحسين دافعية التلاميذ، وتزيد من مستوى الاحتفاظ بالمعلومات، وانتقال أثر التعلّم، أو تزيد من معرفة الطلبة بمستواهم، كما أنها توفر التغذية الراجعة في عملية التدريس (Eidger,2000). فالاختبار الجيد يجب أن يتصف بالدقة، حتى يتسنى لنا تطبيقه كما يجب أن يتسم بالموضوعية وسهوله التصحيح والتفسير والصدق والثبات .

إن شكل وأنماط فقرات الاختبار التي يستخدمها المعلم في قياس مستوى التحصيل لدى طلبته متنوعة ومتعددة . و أهمها الاختبارات الموضوعية حيث تشير الدراسات في مجال تصميم الاختبارات، إلى أن هذا النوع من الاختبارات إذا أُعِد إعداداً شاملاً ودقيقاً فإن نجاح الطالب أو المفحوص لا يخضع للصدفة أو الخطأ، ومن ضمنها الاختبارات المكتوبة التي تصنّف إلى فئتين أساسيتين الفئة الأولى الفقرات ذات الإجابة المصاغة(supply-type): وتشمل (فقرات التكميل – الإجابة القصيرة – المقالية) وفيها الطالب يُعطي الإجابة بلغته الخاصة. والفئة الثانية ذات الإجابة المنتقاة (selected - type) وتشمل فقرات الاختيار من بديلين الصواب والخطأ(True-False Item) و الاختيار من متعدد (Multiple-Choice Item) (الصمادي وَ الزاملي، 2004م).

ورغم أهمية نمط الاختبار وتنوع العوامل التي تؤثر به، إلا أن معظم المعلمين والمعلمات عند اختيارهم نمط الاختبار ونوع الفقرة يستخدمون نمط دون الآخر مع

عدم الاهتمام بتلك العوامل ومدى تفاعل أداء كلاً من الطلاب والطالبات على نمط الاختبار المستخدم أو مناسبته لهم.

وذكر النبهان (2004م) أنه لا يوجد نمط من الفقرات يفضل على النمط الأخر بشكل مطلق، بل أن الاختيار يتم في ضوء استعراض عوامل قوة أو ضعف كل نمط منها. وعليه يجب أن تهتم الدراسات بتحديد تلك العوامل ودراستها بشكل متعمق للتعرف عليها وتحديد أثرها على خصائص الاختبارات، وما الدراسة الحالية إلا محاولة لذلك.

مشكلة الدراسة وتساؤلاتها:

تناولت الكثير من الدراسات أثر أنماط الاختبار في الخصائص السيكومترية للاختبار و فقراته مثل دراسة (الدبك ،1998م؛ العلي،1989م ؛ الكحلوت،2004م ؛ سليمان ،2001م ؛ Anderson &Hyers, 1991؛ Frisbi and Sweeney ,1982؛ Frisbie&Ebel, 1972؛ Ebel,1971؛ Erickson,1988). وجميع تلك الدراسات اهتمت بجانب واحد وهو جانب أنماط الاختبارات بأنواعها كمتغير مستقل مع إهمال أو تجاهل دور جنس المستجيب كمتغير آخر مؤثر على الخصائص السيكومترية للاختبار وفقراته على الرغم من أن دراسات كلٌ من(Andreson,1998؛DeMars,2000؛ Kyriakides& Antoniou ,2009) أكدت أن هناك تفاعل بين نمط الاختبار وجنس المستجيب (ذكور ،إناث) وأن الخصائص السيكومترية للاختبار وفقراته تتغير بتغير النمط جنس المستجيب معاً، في حين أن دراسات أخرى مثل دراسة (Tasdemir, 2010 ؛ Simens&Globni, 2001) توصلوا إلى أنه لا يوجد فروق بين الجنسين (ذكور ،إناث) على الخصائص السيكومترية لفقرات الاختبار (معامل الصعوبة والتمييز).

ومما سبق نجد أن هناك تضارباً واضحاً في الدراسات في أن الخصائص السيكومترية للاختبار وفقراته تتغير بتغير نمط الاختبار وجنس المستجيب بالرغم من ذلك فقد وجد الباحث عدداً قليلاً من الدراسات الأجنبية اهتمت بدراسة أثر التفاعل بين نمط الاختبار وجنس المستجيب على الخصائص السيكومترية للاختبار وفقراته.

ومن هذا المنطلق وبناءً على الحيثيات السابقة فإن الدراسة الحالية تحاول الإجابة على التساؤل الرئيس التالي:

ما أثر التفاعل بين نمط الاختبارات (الاختبار من متعدد، والصواب والخطأ، والإكمال) وجنس المستجيب (ذكور، إناث) على بعض الخصائص السيكومترية للفقرات وثبات الاختبار؟

ويتفرع من هذا التساؤل التساؤلات الفرعية التالية:

1) هل هناك أثر للتفاعل بين نمط الاختبار (الاختبار من متعدد ذو الأربع بدائل والصواب والخطأ والإكمال) وجنس المستجيب (ذكور، إناث) على معامل الصعوبة للفقرة؟

2) هل هناك أثر للتفاعل بين نمط الاختبار (الاختبار من متعدد ذو الأربع بدائل والصواب والخطأ والإكمال) وجنس المستجيب (ذكور وإناث) على معامل التمييز للفقرة؟

3) هل هناك أثر للتفاعل بين نمط الاختبار (الاختبار من متعدد ذو الأربع بدائل والصواب والخطأ والإكمال) وجنس المستجيب (ذكور وإناث) على ثبات الاختبار؟

أهداف الدراسة:

تهدف الدراسة الحالية إلى فحص أثر التفاعل بين نمط الاختبار وجنس المستجيب على الخصائص السيكومترية للاختبار والتحقق من أن نمط الاختبار قد يكون غير كافٍ في تفسير اختلاف الخصائص السيكومترية للاختبارات وأن جنس المستجيب قد يكون عاملاً آخر مؤثراً في ثبات الاختبار والخصائص السيكومترية للفقرة . ويمكن تحديد أهداف الدراسة فيما يلي:

1) التعرّف على نمط الاختبار الأكثر مناسبةً لجنس المستجيب للوصول لمستويات مرتفعة من الثبات.

2) التعرّف على نمط الاختبار الأكثر مناسبةً لجنس المستجيب للوصول لاختبار فقراته ذات خصائص سيكومترية جيدة.

أهمية الدراسة :

تستمد الدراسة أهميتها من خلال ما يلي :

3) المقارنة بين ثلاثة أنواع من الاختبارات (الاختيار من متعدد ذو الاربعة بدائل الصواب والخطأ والإكمال) في المرحلة الثانوية تعتبر من أكثر الاختبارات انتشاراً في مادة الرياضيات.

4) تُساعد نتائج هذه الدراسة في الوصول لنمط أو شكل الاختبار الأكثر مناسبةً بين طلاب وطالبات المرحلة الثانوية.

5) تُساعد في ترشيح إحدى النمطين أو الشكلين (الصواب والخطأ والإكمال) لاستخدامه في الاختبارات التحصيلية لمادة الرياضيات.

6) كما أن الدراسة تُقدم أداة قياس جديدة لمعلمي ومعلمات المدارس الثانوية في المناهج الجديدة المطورة صالحة للاستخدام مستقبلاً لغرض تقويم تحصيل طلاب وطالبات الصف الأول ثانوي لاسيما أنها تتمتع بدرجات عالية من الثبات والفقرات ذات الخصائص السيكومترية الجيدة.

مصطلحات الدراسة:

التفاعل (Interaction):

هو التأثير المشترك للمتغير المستقل على المتغير التابع. (أبوعلام ،2000م).

إجرائياً: في الدراسة الحالية يتم قياس أثر التفاعل بين المتغيرات المستقلة والمتمثلة في جنس المستجيب (ذكور/إناث) ونمط فقرات الاختبار (فقرات الاختيار من متعدد – الصواب والخطأ - الإكمال) على المتغيرات التابعة والمتمثلة في ثبات الاختبار – الخصائص السيكومترية للفقرات (معامل الصعوبة –التمييز) حيث تم استخدام الأسلوب الإحصائي تحليل التباين الثنائي نموذج التأثيرات الثابتة (Two-Way ANOVA,Fixed Effects Model) ذو التصميم العاملي (2×3) .

جنس المستجيب (Gender):

تُشير في الدراسة إلى (ذكور/إناث) أو (طلاب/ طالبات)

نمط الاختبار(Type Test):

أ) فقرات التكميل أو الإكمال (Completion items).

"وهي من اختبارات الاستدعاء أو التذكر (Recall). وتكون عادة عل شكل جملة ناقصة يكمل الجواب "(أبو لبدة،2008م)

إجرائياً: هو اختبار فقراته تتطلب من الطالب وضع الكلمة المناسبة في الفراغ المناسب.

ب) فقرات الصواب والخطأ (True-false items).

وهي "جملة كاملة الترتيب إما يكون معناها صحيحاً أو خطأ وهي عبارات والمطلوب تعيين فيما إذا كانت العبارة صادقة أو خاطئة" "(أبو لبدة،2008م).

إجرائياً: هو اختبار فقراته ثنائية الإجابة (صواب أو خطأ).

ج) الاختيار من متعدد (Multiple- choice items).

"وهو عبارة عن جزئيين يسمى الجزء الأول المتن (Stem) أو مقدمة الفقرة، أما الجزء الثاني فهو عبارة عن بدائل (alternatives) وتتألف من الجواب الصحيح والمموهات (distracters) " (النبهان،2004م).

إجرائيا: هو اختبار فقراته مكونة من متن وأربعة بدائل وهناك بديل واحد صحيح يختار الطالب الإجابة الصحيحة من بين البدائل الأربع.

الثبات (ثبات درجات الاختبار)(Test score Reliability):

هـو" الاتـساق والانـسجام الـذي تقـيس بـه علامـات الاختبـار الـسمة أو الـشي الـذي أعد الاختبار لقياسه " (ملحم،2002).

كمـا يعـرف إجرائيـاً بأنـه ثبـات الاتـساق الـداخلي للاختبـار المحـسوب باسـتخدام معادلة ألفا كرونباخ أو التجزئة النصفية.

الخصائص السيكومترية للفقرات (Psychometric item):

ويقصد بها في هذه الدراسة معاملات الصعوبة والتمييز.

معامل صعوبة الفقرة (Item difficulty).

يقـصد بهـا "مـستوى التعقيـد الـذي يواجهـه الطالـب في الإجابـة الـصحيحة علـى الفقـرة الاختباريـة، ومـا إذا كـان ذلـك المـستوى عاليـاً أو متوسـطاً" (الزاملـي وَالـصارمي وكاظم ،2009).

إجرائيـاً: نـسبة الطلبـة الـذين أجـابوا عـن الفقـرة إجابـة صحيحة مـن بـين الطـلاب الذين حاولوا الإجابة على المفردة.

معامل تمييز الفقرات (Item Discrimination).

عـرّف اناسـتازي (الطريـري،1997م) بأنهـا الدرجـة التي تتمتـع بهـا الفقـرة مـن حيـث التمييز الدقيق والصحيح بين المفحوصين في السمة أو الخاصية التي يقيسها.

إجرائيـاً: وهـو قيمـة معامـل الارتبـاط بـين الأداء علـى الفقـرة والدرجـة الكليـة علـى الاختبـار باسـتخدام معامـل الارتبـاط التسلـسلي الأصـيل(النقطـي) Point Biserial Correlation .

حدود الدراسة:

1) الحـدود الموضـوعية: اقتـصرت هـذه الدراسـة علـى موضـوعات الفـصلين الأول والثـاني مـن مقـرر مـادة الرياضـيات للـصف الأول الثـانوي الفـصل الدراسـي الأول لعام (1432هـ/1433هـ).
2) الحـدود المكانيـة: تـم تطبيـق هـذه الدراسـة علـى عينـة مـن طـلاب وطالبـات الصف الأول الثانوي من المدارس الحكومية بتعليم مكة المكرمة.
3) الحـدود الزمانيـة: طُبقـت الدراسـة فـي الفـصل الدراسـي الأول لعـام (1432هـ/1433هـ).

الدراسات السابقة:

يشتمل هذا الفصل الدراسات التي تناولت الاستجابة على نمط فقرات الاختبارات والخصائص السيكومترية للاختبار والتي صُنّفت إلى ثلاث محاور:

المحور الأول: الفروق في استجابة المفحوصين على أنماط فقرات الاختبار وأثرها على الخصائص السيكومترية للاختبار.

في دراسة أجراها أيبل روبرت Ebel,Robert(1971م) والتي كانت بعنوان "مقارنة بين فعالية اختبارات التحصيلي لأسئلة الاختيار من المتعدد و فعالية اختبار الصواب و الخطأ" وهدفت إلى تحويل أسئلة الاختيار من متعدد إلى أسئلة الصواب و الخطأ بدون تغيير ما تقيسه الأسئلة بالإضافة إلى تحسين الكفاءة، وعليه تم إعادة صياغة 90 سؤال من نوع الاختيار من متعدد لقسم العلوم الطبيعية إلى أسئلة الصواب و الخطأ، و نتج عن ذلك180 سؤال تم توزيعه على نموذجين، وكل نموذج متضمنا 90 سؤال طُبق على الطلبة الجامعيين، وأوضحت نتائج التحليل أن أسئلة الصواب و الخطأ اقل تمييزاً من أسئلة الاختيار من متعدد ولكنها لها نفس ثبات أسئلة الاختيار من متعدد . وهذا يؤيد الاعتقاد بأن أسئلة الصواب والخطأ لها ثبات مثل أسئلة الاختيار من المتعدد، وأنه لا يوجد اختلاف بينهما في القياس.

وفي دراسةأخرى قام بها أوسترهوف وجلاسناب Oosterhof&Glasnapp (1972م) والتي كان عنوانها هو" مقارنة بين ثبات نماذج اختبار الاختيار من المتعدد و نماذج الصواب والخطأ " واشتملت عينة الدراسة على 101 من الطلاب الجامعيين المسجلين في السنة التحضيرية للمنهج الدراسي القياس . نتائج ذلك البحث توضح أن أسئلة الصواب و الخطأ (خاصة الجمل الصحيحة) أقل ثباتاً من أسئلة الاختيار من المتعدد. ومن الواضح أن التصحيح طبقاً لمعادلة التخمين تم تطبيقه لتساوى الدرجات المتعلقة بالأسئلة الصحيحة، وأوضحت الدراسة أن أسئلة الاختيار من المتعدد أسهل من أسئلة الصواب والخطأ.

كما أجرى فريسبي وايبل،Frisbi&Ebel(1972م) دراسة هدفت للمقارنة بين اختبار الاختيار من متعدد و اختبار الصح والخطأ من حيث الثبات والصدق التلازمي والوقت اللازم على الفقرة وكان حجم العينة (1018) طالباً من ستة مدارس حكومية في ولاية ميتشغان . وأشارت النتائج أن هناك فروق ذات دلالة إحصائية في معامل الثبات(KR20) باستخدام اختبار "ت"، للفروق بين معامل ثبات فقرات اختبار الصح والخطأ وفقرات اختبار الاختيار من متعدد عند مستوى(0,01) وعدم وجود فروق ذات دلالة إحصائية في معامل الثبات(KR20) باستخدام اختبار "ت"، للفروق بين معامل ثبات فقرات اختبار الصح والخطأ المحول بطريقة التحكيم ومعامل ثبات فقرات الصح والخطأ المحول بطريقة التمييز عند مستوى(0,05). أما فيما يتعلق بالصدق التلازمي فقد أشارت النتائج أن معاملات ارتباط بيرسون بين الدرجة على اختبار الاختيار من متعدد والدرجة على اختبار الصح خطأ لنفس الموضوع كانت واحد صحيح في ستة من النماذج الثمانية، بعد تصحيح معامل الارتباط . ومن حيث الوقت

وجد أن الطالب الذي يجيب على ثلاث فقرات صح وخطأ، فإنه يجيب في نفس المدة الزمنية على فقرتين من نوع الاختيار من متعدد.

كذلك وقد أجرى أوسترهوف و كوتس Oosterhof& Coats (1981م) دراسة هدفت إلى مقارنة أسئلة الاختيار من متعدد والتكميل في مستويات الصعوبة والثبات وذلك في قياس المهارات الكمية. كما اهتمت الدراسة بمقارنة نوعين من أشكال أسئلة الاختيار من متعدد وهي تلك التي تحتوي على "لا شيء مما سبق كبديل" وتلك التي تضع بديل مستقل لمتن سؤال الاختيار من متعدد. وقد تم إعداد اختبارات من 40 مفردة تم تطبيقها على 250 طالب وطالبة. وقد توصلت الدراسة إلى أن استخدام بديل "لا شيء مما سبق" كان له تأثير طفيف جداً على صعوبة وثبات أسئلة الاختيار من متعدد. كما توصلت الدراسة إلى أن ثبات أسئلة التكميل كان أعلى من ثبات أسئلة الاختيار من متعدد وأن عدد أقل من أسئلة التكميل يكفي للوصول لنفس الثبات الذي يحققه عدد أكبر من أسئلة الاختيار من متعدد.

وفي دراسة فريسبي وسويني Sweeney &Frisbi(1982م) المشار إليها في دراسة (العلي ، 1998م) بعنوان المميزات النسبية للاختبارات التحصيلية، وكان هدفها مقارنة تقديرات كل من وقت وثبات وصدق وصعوبة اختبار الاختيار من نوع اختيار من متعدد بنوع اختبار الصح والخطأ. وقد أجريت الدراسة على عينة مكونة من (574) طالباً. وكانت نتائج الدراسة من حيث الوقت الذي استغرقه المفحوص في إجابة فقرة واحدة من نوع الاختبار الاختيار من متعدد تعادل الإجابة على (3,44) فقرة من نوع الاختبار صح-الخطأ. ومن حيث الثبات كان معامل ثبات فقرات الاختبار من نوع الصح والخطأ أعلى منه في الاختبار من نوع الاختيار من متعدد، ويعزو الباحثان ذلك إلى الفرق في طول الاختبار والوقت الذي استغرقه المفحوص في الإجابة. ومن حيث الصدق وجد أن معامل الصدق عالياً في كلا النوعين من الاختبار، ولا توجد فروق ذات دلالة إحصائية بين معامل الصدق في الاختبار من نوع الاختيار من متعدد ومعامل الصدق في الاختبار من نوع صح وخطأ. ومن حيث الصعوبة عدم توفر دلالة حاسمة أثناء صعوبة أي من نوعي الاختبار على الأخر، وعند سؤال الطلبة عن درجة صعوبة الاختبارين أجاب (47%) منهم بأن الاختبار من نوع الصح والخطأ أصعب و(22%) أجاب أن الاختبار من نوع الاختيار من متعدد هو الأصعب وأن (31%) أجاب بأن للاختبارين نفس الدرجة من الصعوبة.

كما قام أريكسون Erickson(1988م) بدراسة هدفت إلى دراسة أثر شكل الاختبار ونوع الفقرة على مستوى الصعوبة، وعلى مستوى معاملات التمييز للفقرات في اختبارات التحصيل للتمريض. ولتحقيق هذا الهدف تم تطبيق ثلاث اختبارات (الصواب والخطأ المتعدد والإجابة القصيرة والاختيار من متعدد) على عينة حجمها 89 طالبة تمريض، وأشارت النتائج إلى أن معامل ثبات ألفا كان متقارباً في جميع الاختبارات. أما من حيث الصعوبة والتمييز فكان اختبار الصواب والخطأ المتعدد له أعلى متوسط صعوبة، وأعلى متوسط تميز ثم اختبار الاختيار من متعدد.

وأجرى العلي (1989م) دراسة بعنوان المقارنة بين أثر اختبارات الاختيار من متعدد و اختبارات الإجابة القصيرة واختبارات المزيج معاً على بعض الخصائص السيكومترية للاختبار، حيث تكونت عينة الدراسة من 197 طالباً من طلاب الصف الأول الثانوي. وقد أظهرت نتائج الدراسة، وجود فروق بين معاملات ثبات الاختبار ذات دلالة إحصائية في مستوى الأداء الكلي ، و مستوى التذكر و كانت هذه الفروق لصالح الإجابة القصيرة في مستوى الأداء الكلي، و لصالح الاختبار من نوع المزيج بين النوعين معاً، في مستوى التذكر، كما أظهرت النتائج وجود فروق ذات دلالة إحصائية بين مستوى الأداء الكلي، و مستوى الاستيعاب لصالح الإجابة القصيرة في مستوى التذكر، في حين لم تكن هنالك فروق ذات دلالة إحصائية لباقي المعاملات و أظهرت النتائج أن استخدام الاختبار من نوع الإجابة القصيرة أو المزيج من النوعين معاً، أفضل من استخدام الاختبار من نوع الاختيار من متعدد، و لجميع المستويات من الأهداف.

وفي دراسة أندرسون و هايرز Anderson &Hyers (1991م) التي هدفت إلى مقارنة اختبار الاختيار من متعدد و اختبار التكميل من حيث الصعوبة والتمييز و الثبات. ولتحقيق هذا الهدف تم تطبيق نوعي الاختبار على عينة حجمها 500 طالباً، وأشارت النتائج إلى أن معامل ثبات اختبار التكميل أعلى من معامل ثبات اختبار الاختيار من متعدد. ومن حيث الصعوبة توصلت الدراسة إلى أن اختبار التكميل أصعب من اختبار الاختيار من متعدد. ومن حيث معامل التمييز كان اختبار التكميل أعلى من متوسط معامل التمييز لاختبار الاختيار من متعدد.

كما أجرى الدبك (1998م) دراسة بعنوان أثر شكل الفقرات (تكميل، اختبار من متعدد) وعدد البدائل (ثلاثة، وأربعة بدائل) في الخصائص السيكومترية للفقرة (الصعوبة، والتمييز). ولتحقيق الهدف من الدراسة قام الباحث بإعداد اختبار تحصيلي في مادة الفيزياء لطلبة الصف التاسع وكانت العينة مكونة من (300) طالب. وقد كانت نتائج الدراسة من حيث معامل الصعوبة أنه توجد فروق ذات دلالة إحصائية عند مستوى $(\alpha = 0.05)$ في معاملات الصعوبة تعزى لشكل الفقرات بين النموذجين ذي الفقرات من نوع(نمط) التكميل وذي الأربعة بدائل لصالح نموذج التكميل، وبين النموذجين ذي الفقرات من نوع(نمط) التكميل وذي الثلاث بدائل لصالح نموج التكميل. بالنسبة لمعامل التمييز وجود فروق ذات دلالة إحصائية عند مستوى $(\alpha = 0.05)$ في معاملات التمييز تعزى لشكل الفقرات بين النموذجين ذي الفقرات من نوع(نمط) التكميل وذي الأربعة بدائل لصالح نموذج التكميل، ولم تظهر النتائج فروق بين النموذجين ذي الفقرات من نوع(نمط) التكميل وذي الثلاث بدائل. بالنسبة لمعامل ثبات الاتساق الداخلي لم يظهر هناك فرق ذات دلالة إحصائية بين معاملات ثبات الاتساق الداخلي للنماذج الثلاثة.

وفي دراسة تهاني جادالله Gadalla, (1999م) والتي هدفت إلى دراسة التكافؤ بين فقرات الاختيار من متعدد وفقرات الاستجابات المنشأة عندما طُبقت في مجال

الرياضيات للطلاب والطالبات من الصف الثاني إلى السادس. وقد قامت الباحثة بمقارنة الدرجة الكلية على كل نوع من نوعي الأسئلة وكذلك البناء العاملي للمفردات في كل نوع من نوعي الأسئلة. وقد تكونت عينة الدراسة من 1028 طالباً وطالبةً من الصف الثاني إلى السادس. وقد تم تطبيق اختباران على كل طالب بينهما فترة زمنية مناسبة. وقد توصلت الدراسة إلى أنه في الصفوف الثاني والثالث كان متوسط الطلبة أعلى في أسئلة الاختيار من متعدد مقارنة بالأسئلة المنشأة في حين لم تتوصل الدراسة لفروق دالة بين درجات الطلاب في نوعي الاختبار في الصفوف من الرابع إلي السادس.

كما أجرى سليمان (2001م) دراسة هدفت إلى استقصاء بنية فقرات الصح والخطأ والتكميل على الخصائص السيكومترية للفقرة، ممثلة بالصعوبة والتمييز وعلى الخصائص السيكومترية للاختبار، ممثلة بالصدق والثبات. حيث طبقت الاختبارات التحصيلية على عينة من 1440 طالباً وطالبةً في محافظة إربد، وتوصلت الدراسة إلى وجود فروق دالة إحصائياً بين متوسطات معاملات الصعوبة لصالح الفقرات سليمة البنية. بينما توصلت الدراسة لعدم وجود فروق دالة إحصائياً بين متوسطات معاملات التمييز للفقرات سليمة ومخالفة البنية. وتوصلت الدراسة إلى وجود فروق دالة إحصائياً بين متوسطات معاملات الثبات والصدق المرتبط بمحك لصالح الفقرات سليمة البنية.

وفي دراسة الكحلوت (2002م) التي هدفت إلى مقارنة بين الخصائص السيكومترية لكل من اختبارات الاختيار من متعدد واختبارات التكميل حيث تم تطبيق الاختبارين على عينة مكونة من 451طالبا وطالبة من الصف السادس الأساسي من مجتمع الدراسة الذي يضم مدارس وكالة الغوث الدولية في الأردن. وأظهرت نتائج الدراسة أن معاملي ثبات اختبار التكميل المحسوبين بطريقة الفا كرونباخ وطريقة إعادة الاختبار أكبر من معاملي ثبات اختبار الاختيار من متعدد كما بينت النتائج أن متوسط معاملات الصعوبة لاختبار الاختيار من متعدد أكبر من متوسط معاملات الصعوبة لاختبار التكميل، وعلى مستوى معاملات التمييز ظهرت الدراسة أن معاملات التمييز لاختبار التكميل أكبر من متوسط معاملات الصعوبة لاختيار من متعدد.

كما أجرى تاسديمير Tasdemir(2010م) دراسة بعنوان مقارنة مستويات الصعوبة وقوى التمييز وقوى انجاز الاختبار في كل من اختبارات الاختيار من متعدد واختبارات الصواب والخطأ. تم إجراء البحث على 252 طالباً من طلاب مرحلة البكالوريوس. وقد أظهرت التحليلات التالية لتطبيق وتحليل اختبارات الاختيار من متعدد والصواب والخطأ التي تم تحضيرها بشكل متساوي أن ثباتها كان عالياً. كما لم تظهر النتائج وجود فروق ذات دلالة إحصائية بين متوسطات اختبار الاختيار من متعدد والصواب والخطأ على أساس الجنس. وكان الفرق بين متوسطات الإجابات على أسئلة الاختبار الـ28 ذات دلالة إحصائية. في حين كان الفرق في متوسط 13 سؤال لصالح أسئلة الاختيار من متعدد، وكان الفرق في متوسط 15 سؤال من صالح أسئلة الصواب والخطأ. ومع ذلك، فإن الفرق بين متوسطات 22 سؤال لم يكن فيها دلالة إحصائية. وكانت النتائج أن اختبارات الاختيار من متعدد ليست متفوقة على

اختبارات الصواب والخطأ إذا تم المقارنة بينهما في الكشف عن نتائج تحصيل الطلبة ونفس الشيء بالعكس. أما بالنسبة لأسئلة التراكيب الـ24 لاختبارات الاختيار من متعدد والصواب والخطأ فقد تم تطبيقها في العام الدراسي على أن لها نفس المستوى من حيث مؤشرات التميز ومستويات الصعوبة. وأظهرت القيم المحسوبة لكل زوج من الأسئلة دلالة إحصائية.

المحور الثاني: الفروق بين الجنسين في الاستجابة على نمط فقرات الاختبار وأثرها على الخصائص السيكومترية للاختبار.

أجرى ويلسون وَ زانج Wilson & Zhang (1998م) دراسة هدفت الى الفروق بين الجنسين في الاستجابة على المفردات الاختبارية من نوع الاختيار من متعدد ومن نوع الاستجابات المنشأة. وكان عدد العينة 29809 من الصفوف الثالث والخامس والثامن والعاشر. وقد توصلت الدراسة إلى أن متوسط درجات الطلاب كانت أعلى من درجات الطالبات في كلا الاختباران. كما توصلت الدراسة إلى أنه بالرغم من أن متوسط درجات الذكور كانت أعلى من متوسط درجات الإناث إلا أن الفرق بين المتوسطين في اختبار الصواب والخطأ كان أقل بكثير من الفرق بين المتوسطين في الاستجابات المنشأة. كما كان الفرق بين متوسطي الذكور والإناث في المفردات ذات المتطلبات المعرفية الأعقد (الأصعب) أكبر بكثير من الفرق بين المتوسطين في المفردات ذات المتطلبات المعرفية البسيطة (الأسهل نسبياً). كما توصلت الدراسة أيضا إلى أن الفرق يزيد مع زيادة الفرقة الدراسية من الثالث إلى العاشر.

وفي دراسة أخرى لبيلر وجافاني Beller&Gafni(2000م) والتي هدفت لفحص الأداء الفارق للذكور والإناث في أسئلة الاختيار من متعدد والأسئلة المفتوحة في أحد الاختبارات الدولية في الرياضيات وهو الاختبار الدولي للتقدم التعليمي في الرياضيات International Assessment of Educational Progress (IAEP) mathematics test وذلك في الأعوام 1988 م و 1991م. في عام 1988م تكونت عينة الدراسة من 1000 طالب وطالبة من كل دولة من الدول الست المشاركة كان متوسط أعمارهم 13 عاما. أما في عام 1991م فقد تكونت عينة الدراسة من 1650 طالب وطالبة من كل فئة عمرية (9 أعوام و13 عام) من 20 من الدول المشاركة في التقييم عام 1991. وقد توصلت الدراسة إلى أن أداء الذكور كان أفضل من أداء الإناث في الرياضيات بصفة عامة. في عام 1988م كان أداء الذكور أفضل من أداء الإناث في أسئلة الاختيار من متعدد مقارنة بالأسئلة المفتوحة. في حين كانت نتائج عام 1991م معاكسة حيث كان أداء الذكور أفضل من أداء الإناث في الأسئلة المفتوحة مقارنة بأسئلة الاختيار من متعدد مما يدل على أن الفروق بين الجنسين لا تكفي لتفسير أداء الطلاب في مادة الرياضيات. وقد توصلت الدراسة لمعاملات ارتباط منخفضة بين صعوبة المفردات وجنس المستجيب حيث تراوحت بين 0.26 إلى 0.53 في الأعمار المختلفة. كما توصلت الدراسة إلى أنه كلما زادت صعوبة المفردات كلما كان أداء الطلاب الذكور أفضل من الإناث.

المحور الثالث: التفاعل بين نمط الجنس ونمط فقرات الاختبار وأثرها على الخصائص السيكومترية للاختبار

في دراسة أجراها أندرسون,Andreson (1998م) وكان عنوانها دراسة التفاعل بين جنس المستجيب ونمط السؤال على صدق وثبات الاختبار. وتكونت العينة من 345 طالب وطالبة في مادة الرياضيات في الصف الثاني عشر. وكان من بين نتائج الدراسة وجود تفاعل بين الجنس ونمط السؤال، حيث كانت درجات الطالبات أعلى من درجات الطلاب في أسئلة الصواب والخطأ في حين كانت درجات الطلاب أعلى من الطالبات في أسئلة الاختيار من متعدد. كما توصلت الدراسة إلى أن الأسئلة الصعبة تبدو أصعب لدى الطالبات وكذلك الأسئلة السهلة تبدو أسهل لدى الطالبات. كما توصلت الدراسة لعدم وجود فروق دالة إحصائيا في ثبات وصدق الاختبار تعزى للجنس.

كما أجرى ديمارسDeMars (2000م) دراسة هدفت إلى فحص أثر ترتيب الاختبار ونمط الاختبار (أسئلة الاختيار من متعدد، وأسئلة الإجابات المنشأة) وجنس المستجيب وذلك في اختبارات الرياضيات والعلوم لأحد دبلومات الدراسات العليا. وقد تكونت عينة الدراسة من 422 طالب وطالبة من طلاب الدراسات العليا. وقد توصلت الدراسة من بين نتائجها إلى وجود تفاعل بين نمط الاختبار وجنس المستجيب فقد كانت درجات الطلاب أعلى من درجات الطالبات على اسئلة الاختيار من متعدد، في حين كانت درجات الطالبات أعلى من الطلاب على أسئلة الاجابات المنشأة.

وفي دراسة بيلنسكي وَ دافيسون Bielinski& Davison (2001م) والتي هدفت إلى دراسة التفاعل بين صعوبة المفردات والجنس لدى الطلبة في مادة الرياضيات. وقد تكونت عينة الدراسة من طلاب من صفوف الرابع والثامن والعاشر والثاني عشر في عام 1992م لاختبارات دولية وهي NEAP, TIMSS and NELS. وقد تم فحص التفاعل بين صعوبة المفردات والجنس من خلال معامل الارتباط بين الفرق بين معامل الصعوبة لدى الذكور والإناث ومعامل الصعوبة للعينة الكلية. وتوصلت الدراسة إلى أن المفردات السهلة كانت أسهل لدى الإناث منها لدى الذكور في حين أن المفردات الصعبة كانت أصعب لدى الإناث منها لدى الذكور. كما وصلت الدراسة لمعاملات ارتباط سالبة ودالة بين الفرق في معامل الصعوبة لدى الذكور والإناث من ناحية والإناث ومعامل الصعوبة للعينة الكلية من ناحية أخرى.

كما أجرى سيمنز وَ جلوبني Simens&Globni (2001م)دراسة كان من أهدافها استقصاء الفروق بين الجنسين في ثلاثة أشكال من أسئلة الاختيار من متعدد وأثر ذلك على صعوبة وتمييز فقرات الاختبار. وتكونت عينة الدراسة من 1028 طالب وطالبة في المرحلة المتوسط في اختبارات الرياضيات والعلوم. وكان من بين نتائج الدراسة عدم وجود فروق بين الجنسين في الصعوبة والتمييز في أي من أنماط أسئلة الاختيار من متعدد الثلاثة.

وفي دراسة كيركيدس وَ انتونيو Kyriakides& Antoniou (2009م) والتي هدفت لدراسة التفاعل بين الجنس وصعوبة المفردات لاختبارات الرياضيات في قبرص.

وتكونت عينة الدراسة من 3242 طالب وطالبة. وقد توصلت الدراسة إلى أن أداء الطلاب كان أفضل في المفردات الصعبة في حين كان أداء الطالبات أفضل في المفردات السهلة. وقد توصلت الدراسة لمعاملات ارتباط دالية إحصائياً بين جنس المستجيب والصعوبة وتراوحت بين 0.53 إلى 0.72 وفقا للفئات العمرية المختلفة.

التعليق على الدراسات السابقة :

من خلال استعراض الباحث الدراسات السابقة يمكن ملاحظة ما يلي:

1) بعض الدراسات السابقة اعتمدت على المقارنة بين أنماط الاختبارات الموضوعية مع تجاهل جنس المستجيب (ذكور إناث) على الخصائص السيكومترية للفقرات و ثبات الاختبار مثل الدراسات (العلي 1989؛ دبك، 1998م؛ Ebel,1971؛Frisbie,&Ebel,1972؛ Erickson 1988؛ Anderson & Hyers 1991 , Frisbi& Sweeney 1982 ؛ امريزق ، 1992م؛ سليمان، 2001م) والتي بحثت في ضوء متغير مستقل واحد هو أنماط الاختبارات.

2) هناك دراسات اهتمت بالفروق بين الجنسين على الخصائص السيكومترية للاختبار وفقراته ولم تدرس التفاعل بينهما مثل دراسات (Gadalla, 1999 ؛ Wilson & Zhang, 1998؛ Beller&Gafni, 2000)

3) يوجد على حد علم الباحث دراسات أجنبية قليلة جداً اهتمت بالتفاعل بين وجنس المستجيب وأنماط الاختبار ومنها دراسة (Andreson,1998؛ DeMars,2000؛ Kyriakides& Antoniou ,2009) . كما أن معظم هذه الدراسات قامت بأخذ نمطين من أنماط الاختبارات وهي إما (الاختيار من متعدد والصح والخطأ) أو (الاختيار من متعدد والإكمال) حيث تختلف الدراسة الحالية عن الدراسات السابقة بدراسة التفاعل بين أنماط الاختبارات الأكثر استخدام وهي الاختبار (فقرات الاختيار من متعدد – الصواب والخطأ - التكميل) كمتغير مستقل أول وجنس المستجيب(ذكور/إناث) كمتغير مستقل آخر والمتغيرات التابعة والمتمثلة في ثبات الاختبار والخصائص السيكومترية للفقرات (معامل الصعوبة –التمييز).

4) تتفق الدراسة الحالية مع الدراسات السابقة في معظم الجوانب مثل المنهج المستخدم وبعض الأساليب الإحصائية وتختلف عنها في طبيعة المادة ودراسة التفاعل بين متغيرين هما أنماط فقرات الاختبار والجنس (ذكور / إناث) وأثرها على المتغيرات التابعة والمتمثلة في ثبات الاختبار والخصائص السيكومترية للفقرات (معامل الصعوبة –التمييز) واستخدام الأسلوب الإحصائي الأمثل (تحليل التباين الثنائي Two-way ANOVA) ذو التصميم العاملي (2×3).

5) أن الدراسة الحالية تتناول جانب تكميلي لما سبقها من دراسات وذلك من خلال التركيز على أثر التفاعل وأنواعه المختلفة بين المتغيرات المستقلة والمتمثلة في

جنس المستجيب (ذكور-إناث) ونمط فقرات الاختبار (فقرات الاختيار من متعدد – الصواب والخطأ - الإكمال) وأثرها على المتغيرات التابعة والمتمثلة في ثبات الاختبار والخصائص السيكومترية للفقرات (معامل الصعوبة –التمييز).

فروض الدراسة:

1) لا يوجد أثر دال إحصائياً للتفاعل عند مستوي (0.05) بين نمط الاختبار وجنس المستجيب (ذكور،إناث) على معامل الصعوبة للفقرة .
2) لا يوجد أثر دال إحصائياً للتفاعل عند مستوي (0.05) بين نمط الاختبار وجنس المستجيب (ذكور،إناث) على معامل التمييز للفقرة .
3) لا يوجد أثر دال إحصائياً للتفاعل عند مستوي (0.05) بين نمط الاختبار وجنس المستجيب (ذكور،إناث) على ثبات الاختبار.

إجراءات الدراسة:

أولاً: منهج الدراسة:

استخدمت الدراسة الحالية المنهج الوصفي، حيث يذكر عطوي (2007م) بأنه "منهج يعتمد على دراسة الظاهرة كما توجد في الواقع ، كما يهتم بوصفها وصفاً دقيقاً ويعبر عنها تعبير كيفياً أو تعبيراً كمياً ، بهدف الوصول إلى استنتاجات تساهم في فهم الواقع وتطوره ".

ثانياً: مجتمع الدراسة:

تتمثل في جميع طلاب وطالبات مدارس التعليم العام للمرحلة الثانوية بمكة المكرمة بمدينة مكة المكرمة للعام 1433/1432هـ .

ثالثاً: عينة الدراسة:

تم اختيار عينة عشوائية عنقودية من مجتمع الدراسة حيث بلغت (446) طالب وطالبة بواقع (226)طالب و(220) طالبة من 3 مدارس ثانوية حكومية طلاب وهي (ثانوية الملك عبدالعزيز وثانوية الملك فهد وثانوية قباء) و3 مدارس ثانوية حكومية طالبات وهي (ثانوية الرابعة والعشرون وثانوية الثلاثون وثانوية الثامنة والعشرون) حيث تم اختيار ثلاثة فصول دراسية داخل المدرسة عشوائياً و تطبيق الأنماط الثلاثة للاختبار (الاختيار من متعدد والصواب والخطأ والإكمال) في الفصل الواحد بحيث تم تقسيم عدد الطلاب على الأنماط الثلاثة بمتوسط 10 طلاب لكل نمط داخل كل فصل.

جدول رقم (1): عدد أفراد العينة من بنين وبنات في أنماط الاختبارات الثلاثة

الجنس	نمط الاختبار	نمط الاختبار: الاختيار من متعدد	الصواب والخطأ	الإكمال	المجموع
جنس المستجيب	ذكور	78	75	73	226
	إناث	77	75	68	220
المجموع		155	150	141	446

رابعاً: أدوات الدراسة:

أولاً: تحديد الموضوعات المراد قياسها:

اختار الباحث الفصلين الأول والثاني لمقرر الرياضيات في المرحلة الثانوية للصف الأول الثانوي لعام 1432هـ/1433هـ ، وهي: الفصل الأول: التبرير والبرهان و الفصل الثاني: التوازي والتعامد.

ثانياً: الأهداف السلوكية والفقرات الاختبارية:

قام الباحث بصياغة الأهداف السلوكية وكتابتها على المستويات المعرفية لتصنيف بلوم بصياغة الفقرات الاختبارية بالثلاثة الأنماط (الاختبار من متعدد، الصواب والخطأ والإكمال) لضمان التكافؤ في الفقرات الاختبارية.

جدول رقم (2):

عدد الدروس والأهداف السلوكية

الفصول (الأبواب)	عدد الدروس	الأهداف السلوكية
الباب الأول: التبرير والبرهان	8 دروس	43 هدف سلوكي
الباب الثاني: التوازي والتعامد	6 دروس	29 هدف سلوكي
المجموع	14 درس	72 هدف

ثالثاً: تحكيم الأهداف السلوكية والفقرات الاختبارية (الصدق الظاهري):

تم عرض تحليل المحتوى والفقرات الاختبارية للأنماط الثلاثة هي :(الاختبار من متعدد ، الصواب والخطأ و التكميل) لضمان التكافؤ في بناء الفقرات الاختبارية على نخبة من أساتذة الجامعات في القياس والتقويم من جامعة أم القرى وجامعة الطائف وجامعة الباحة ونخبة من المشرفين التربويين متخصصين في مجال الرياضيات منهم رئيس قسم الرياضيات بتعليم مكة المكرمة ومشرفين تربويين ونخبة من المعلمين المتميزين وبلغ عددهم (15 محكم) وكانت نتائج اراء المحكمين تراوحت

النسبة المئوية للاتفاق بين المحكمين على ملائمة كل هدف سلوكي لمستواه المعرفي بين 93% إلى 100%. وتم تعديل بعض الفقرات بناء على أراء المحكمين.

رابعا: تكوين جدول المواصفات (على صدق المحتوى):

تم إعداد جدول المواصفات في ضوء الوزن النسبي لكل موضوع والوزن النسبي لمستويات الأهداف.

جدول رقم (3):

جدول المواصفات

الباب	موضوع المقرر الدراسي	عدد الحصص	مستوى الأهداف المعرفية السلوكية			عدد الفقرات لكل موضوع	الأوزان النسبية للموضوعات
			التذكر	الفهم	التطبيق		
الباب الأول : التبرير والبرهان	التبرير الإستقرائي والتخمين الرياضي	2	1	1	1	3	%6
	المنطق الرياضي	2	1	2	1	4	%6
	العبارت الشرطية	2	1	2	1	4	%7
	التبرير الإستنتاجي	3	2	2	1	5	%9
	المسلمات والبراهين الحرة	3	2	2	1	5	%9
	البرهان الجبري	2	1	-	1	2	%6
	إثبات علاقات بين القطع المستقيمة	2	1	-	1	2	%6
الباب الثاني : التوازي والتعامد	إثبات علاقات الزوايا	3	1	2	1	4	%9
	المستقيمان المتوازيان والمستقيمات المستعرضة	2	1	3	-	4	%6
	الزوايا والمستقيمات المتوازية	2	1	2	-	3	%6
	ميل المستقيم	3	1	2	1	4	%9
	معادلة المستقيم	3	2	2	1	5	%9
	إثبات توازي المستقيمات	2	1	1	1	3	%6
	الأعمدة والمسافة	2	1	-	1	2	%6
	المجموع	33	17	21	12	50	-
	مجموع الأوزان النسبية	-	(%42)	(%40)	(%18)	-	%100

خامساً: كتابة وصياغة الأسئلة:

قام الباحث بكتابة الفقرات الاختبارية بحيث تضمن خمسون فقرة لكل نمط من أنماط الاختبار الثلاثة (الاختبار من متعدد ذو الأربع بدائل، الصواب والخطأ والإكمال).

سادساً: توفير أدوات الدراسة:

قام الباحث بتوفير نسخ كافية من نماذج الاختبار لثلاثة الأنماط وأوراق التصحيح اليدوي لنمط الصواب والخطأ ونمط الاختيار من متعدد.

سابعاً: الدراسة الاستطلاعية:

قام الباحث بتطبيق الاختبار على عينة من الطلاب هم طلاب الصف الأول ثانوي في ثانوية المدائن التابعة لتعليم مكة المكرمة وعددهم (81) طالباً حيث كان الهدف من ذلك هو تحليل فقرات الاختبار وضمان تكافؤ الاختبارات بأنماطها الثلاثة كما هو موضح في الجدول التالي:

جدول رقم (4):

عدد طلاب العينة الاستطلاعية

نمط الاختبار	عدد الطلاب
1) التكميل	28 طالب
2) الاختيار من متعدد	26 طالب
3) الصواب والخطأ	27 طالب
المجموع	81 طالب

- موعد الاختبار بتاريخ 1432/12/19هـ

زمن الاختبار = (زمن انتهاء أول طالب + زمن انتهاء آخر طالب) / 2

- وكان الوقت المخصص للاختبار هو: (ساعة ونصف) (90 دقيقة) وتم حساب الوقت الفعلي للإجابة كما هو موضح في الجدول التالي:

جدول رقم (5): زمن الاختبار للأنماط الثلاثة في الدراسة الاستطلاعية

أنماط الاختبارات الثلاثة	زمن انتهاء أول طالب	زمن انتهاء آخر طالب	زمن الاختبار
1)الصواب والخطأ	18 د	70 د	44 د
2)الاختيار من متعدد	23د	84 د	54 د
3)الإكمال	17 د	90د	53,5 د

ثامناً: قام الباحث تصحيح الاختبار للعينة الاستطلاعية للثلاثة الأنماط يدوياً

تاسعاً: التحليل الإحصائي للعينة الاستطلاعية:

تم استخدام الأساليب الإحصائية الآتية للتحقق من الصدق والثبات والخصائص السيكومترية للأنماط الثلاثة للاختبار: معامل ألفا والتجزئة النصفية (سبيرمان، براون- وجتمان) - معاملي الصعوبة والتمييز لكل فقرة - قوة تشتيت البدائل لفقرات الاختيار من متعدد- معاملات الاتساق الداخلي لكل فقرة كمؤشر للصدق وذلك في سبيل الوصول على أنماط متكافئة للاختبار في صورته النهائية.

الخصائص السيكومترية ومعاملات الصعوبة والتمييز وإحصاءات مفردات الاختبارات التحصيلية:

أولاً: معامل الصعوبة والتمييز لأنماط الاختبارات الثلاثة في الدراسة الاستطلاعية المكونة من 50 فقرة:

1) معاملات الصعوبة والتمييز للاختبار الاختيار من متعدد:

تباينت معاملات صعوبة مفردات الاختيار من متعدد، حيث تراوحت معاملات الصعوبة بين (0,25) للمفردة (3) وللمفردة (11) التي تعد أصعب مفردات اختبار الاختيار من متعدد حيث أجاب عليها (6) أفراد فقط إجابة صحيحة ووصلت معاملات الصعوبة إلى (0.85) للمفردة (7) والتي تعد أسهل مفردات اختبار الاختيار من متعدد حيث أجاب عليها (22) فرد إجابة صحيحة.

تباينت معاملات التمييز كما هو في ملحق رقم (1-1) وإن كانت معظمها مرتفع إذا ما استثنينا المفردات (2) والتي لها معامل تمييز (0.11)، والمفردة (7) والتي لها معامل تمييز (0.09)، والمفردة (20) والتي لها معامل تمييز (0.12)، حيث تراوحت معاملات التمييز بين (0.31) للمفردة (4) والمفردة (40) إلى (0.79) للمفردة (13) وهي معاملات تمييز مقبولة إلى مرتفعة بعد حذف المفردات الثلاث المنخفضة التمييز السابقة الذكر.

ومن ملحق رقم (1-1) نجد أن متوسط الدرجة الكلية للطلاب الذين أجابوا على المفردة بشكل صحيح كان أكبر من متوسط الدرجة الكلية للطلاب الذين أجابوا

بـشكل خطـأ علـى المفـردة وذلـك صـحيح لجميـع مفـردات الاختبـار ممـا يـشير لأن جميـع المفردات استطاعت التمييز بين الطلاب، ولكن بدرجات متفاوتة.

ملحق رقم (1-1):

جـدول معـاملات الـصعوبة والتمييز وإحـصاءات مفـردات اختبـار الاختيـار مـن متعـدد للعينـة الاستطلاعية (50 مفردة)

رقم السؤال	معامل الصعوبة	معامل التمييز	عدد الطلاب المجيبين علي المفردة بشكل صحيح	عدد الطلاب المجيبين علي المفردة بشكل غير صحيح	متوسط الدرجة الكلية للطلاب المجيبين علي المفردة بشكل صحيح	متوسط الدرجة الكلية للطلاب المجيبين علي المفردة بشكل غير صحيح
1	0.73	0.35	19	7	29.05	19.43
2	0.73	0.11	19	7	27.32	24.14
3	0.25	0.32	6	18	31.83	22.94
4	0.72	0.31	18	7	28.78	20.14
5	0.40	0.69	10	15	36.10	18.93
6	0.73	0.44	19	7	29.79	17.43
7	0.85	0.09	22	4	26.91	24.00
8	0.73	0.50	19	7	30.21	16.29
9	0.54	0.55	14	12	32.79	19.08
10	0.50	0.69	13	13	35.00	17.92
11	0.25	0.37	6	18	32.67	22.56
12	0.35	0.71	9	17	38.56	20.06
13	0.36	0.79	9	16	39.89	19.81
14	0.69	0.48	18	8	30.39	17.63
15	0.58	0.61	15	11	32.93	17.64
16	0.54	0.60	14	12	33.29	18.50
17	0.65	0.52	17	9	31.12	17.67
18	0.58	0.64	15	11	33.20	17.27

رقم السؤال	معامل الصعوبة	معامل التمييز	عدد الطلاب المجيبين علي المفردة بشكل صحيح	عدد الطلاب المجيبين علي المفردة بشكل غير صحيح	متوسط الدرجة الكلية للطلاب المجيبين علي المفردة بشكل صحيح	متوسط الدرجة الكلية للطلاب المجيبين علي المفردة بشكل غير صحيح
19	0.50	0.53	13	13	33.00	19.92
20	0.35	0.12	9	17	28.56	25.35
21	0.44	0.46	11	14	32.27	20.86
22	0.60	0.47	15	10	30.47	18.80
23	0.69	0.61	18	8	31.50	15.13
24	0.73	0.50	19	7	30.21	16.29
25	0.54	0.52	14	12	32.36	19.58
26	0.69	0.43	18	8	30.00	18.50
27	0.36	0.49	9	16	34.56	21.75
28	0.38	0.46	10	16	33.60	22.00
29	0.54	0.67	14	12	34.14	17.50
30	0.69	0.56	18	8	31.06	16.13
31	0.62	0.57	16	10	32.06	17.50
32	0.58	0.63	15	11	33.13	17.36
33	0.54	0.77	14	12	35.29	16.17
34	0.50	0.52	13	13	32.92	20.00
35	0.62	0.54	16	10	31.75	18.00
36	0.62	0.57	16	10	32.06	17.50
37	0.56	0.55	14	11	31.86	18.27
38	0.38	0.67	10	16	36.90	19.94
39	0.38	0.60	10	16	35.90	20.56
40	0.32	0.31	8	17	31.38	23.29
41	0.42	0.79	11	15	37.82	18.13

رقم السؤال	معامل الصعوبة	معامل التمييز	عدد الطلاب المجيبين علي المفردة بشكل صحيح	عدد الطلاب المجيبين علي المفردة بشكل غير صحيح	متوسط الدرجة الكلية للطلاب المجيبين علي المفردة بشكل صحيح	متوسط الدرجة الكلية للطلاب المجيبين علي المفردة بشكل غير صحيح
42	0.65	0.45	17	9	30.53	18.78
43	0.35	0.39	9	17	33.11	22.94
44	0.42	0.53	11	15	34.09	20.87
45	0.69	0.59	18	8	31.33	15.50
46	0.46	0.64	12	14	35.00	19.14
47	0.46	0.53	12	14	33.58	20.36
48	0.42	0.54	11	15	34.18	20.80
49	0.50	0.71	13	13	35.23	17.69
50	0.46	0.63	12	14	34.83	19.29

فاعلية المشتتات لاختبار الاختيار من متعدد:

يتضح من ملحق رقم (1-2) أن متوسط الدرجة الكلية للطلاب الذين اختاروا البديل الصحيح في جميع مفردات الاختيار من متعدد كان أكبر من متوسط الدرجة الكلية للطلاب الذين اختاروا البدائل الخاطئة. وفي حين كان الفرق كبير بين متوسط الدرجة الكلية للطلاب الذين اختاروا البديل الصحيح ومتوسط الدرجة الكلية للطلاب الذين اختاروا البدائل الخاطئة لمعظم مفردات الاختيار من متعدد، كان الفرق متواضع في ثلاث مفردات أرقام 2، 7، 25 حيث كان الفرق متواضع بين متوسط الدرجة الكلية للطلاب الذين اختاروا البديل الصحيح ومتوسط الدرجة الكلية للطلاب الذين اختاروا البدائل الخاطئة. على سبيل المثال المفردة الثانية كان متوسط الدرجة الكلية للطلاب الذين اختاروا البديل الصحيح (ب) هو 27,32 في حين كان متوسط الدرجة الكلية للطلاب الذين اختاروا البديل غير الصحيح (أ) متقارب جدا وهو 26. وفي حالة المفردة 34 كان متوسط الدرجة الكلية للطلاب الذين اختاروا البديل الصحيح (ج) هو 32,36 وهو أقل من متوسط الدرجة الكلية للطلاب الذين اختاروا البديل غير الصحيح (أ) وهو 34 ولذلك فإن حذف تلك المفردات يؤثر ايجابياً على الدرجة الكلية لمفردات الاختيار من متعدد.

ملحق رقم (1-2):

جدول قوة تشتيت البدائل لمفردات اختبار الاختيار من متعدد للعينة الاستطلاعية(50مفردة)

رقم السؤال	الإجابة الصحيحة	عدد الذين اختاروا البديل (أ)	عدد الذين اختاروا البديل (ب)	عدد الذين اختاروا البديل (ج)	عدد الذين اختاروا البديل (د)	متوسط الدرجة الكلية للطلاب الذين اختاروا البديل (أ)	متوسط الدرجة الكلية للطلاب الذين اختاروا البديل (ب)	متوسط الدرجة الكلية للطلاب الذين اختاروا البديل (ج)	متوسط الدرجة الكلية للطلاب الذين اختاروا البديل (د)
1	أ	19	5	0	2	29.05	21.40		14.50
2	ب	3	19	4	0	26.00	27.32	22.75	
3	د	8	6	4	6	29.00	20.00	15.25	31.83
4	ب	4	18	1	2	20.00	28.78	24.00	18.50
5	ج	8	7	10	0	18.25	19.71	36.10	
6	د	1	2	4	19	14.00	19.00	17.50	29.79
7	أ	22	4	0	0	26.91	24.00		
8	ج	2	3	19	2	18.50	15.67	30.21	15.00
9	ج	9	2	14	1	17.33	30.00	32.79	13.00
10	ب	6	13	5	2	17.83	35.00	17.60	19.00
11	أ	6	10	7	1	32.67	18.70	27.29	28.00
12	أ	9	3	12	2	38.56	21.67	20.25	16.50
13	ج	8	6	9	2	17.38	24.17	39.89	16.50
14	د	4	4	0	18	18.25	17.00		30.39
15	ج	5	3	15	3	16.00	21.00	32.93	17.00
16	ج	6	6	14	0	14.83	22.17	33.29	
17	ج	2	7	17	0	15.50	18.29	31.12	
18	ب	9	15	2	0	17.44	33.20	16.50	
19	ب	7	13	2	4	25.43	33.00	14.00	13.25
20	ب	9	9	6	2	23.89	28.56	27.67	25.00
21	ب	6	11	6	2	19.33	32.27	24.33	15.00
22	ج	7	3	15	0	19.43	17.33	30.47	
23	ب	3	18	3	2	17.33	31.50	13.67	14.00
24	ب	1	19	4	2	15.00	30.21	17.00	15.50
25	ج	1	5	14	6	34.00	23.00	32.36	14.33
26	ج	3	3	18	2	16.67	17.00	30.00	23.50
27	أ	9	8	6	2	34.56	20.00	22.00	28.00
28	ب	3	10	8	5	17.67	33.60	26.38	17.60
29	ج	6	3	14	3	17.33	13.33	34.14	22.00
30	ب	4	18	1	3	16.75	31.06	14.00	16.00

ملحق رقم (1-2):

جدول قوة تشتيت البدائل لمفردات اختبار الاختيار من متعدد للعينة الاستطلاعية(50مفردة)

رقم السؤال	الإجابة الصحيحة	عدد الذين اختاروا البديل (أ)	عدد الذين اختاروا البديل (ب)	عدد الذين اختاروا البديل (ج)	عدد الذين اختاروا البديل (د)	متوسط الدرجة الكلية للطلاب الذين اختاروا البديل (أ)	متوسط الدرجة الكلية للطلاب الذين اختاروا البديل (ب)	متوسط الدرجة الكلية للطلاب الذين اختاروا البديل (ج)	متوسط الدرجة الكلية للطلاب الذين اختاروا البديل (د)
31	أ	16	4	4	2	32.06	16.75	18.25	17.50
32	ج	4	5	15	2	16.00	18.80	33.13	16.50
33	د	4	4	4	14	17.00	15.00	16.50	35.29
34	ج	0	11	13	2		21.64	32.92	11.00
35	ج	2	5	16	3	20.50	18.00	31.75	16.33
36	ب	3	16	3	4	18.33	32.06	15.00	18.75
37	د	2	6	3	14	20.00	19.67	14.33	31.86
38	ب	7	10	5	4	18.57	36.90	22.00	19.75
39	أ	10	7	4	5	35.90	18.43	24.25	20.60
40	ب	5	8	7	5	22.00	31.38	22.57	25.60
41	د	4	6	5	11	17.50	21.67	14.40	37.82
42	ب	3	17	2	4	20.67	30.53	16.50	18.50
43	ب	6	9	7	4	27.33	33.11	19.14	23.00
44	ج	8	3	11	4	21.38	23.67	34.09	17.75
45	ج	2	3	18	3	21.00	13.33	31.33	14.00
46	أ	12	4	4	6	35.00	18.50	16.50	21.33
47	ب	6	12	6	2	16.17	33.58	26.00	16.00
48	أ	11	7	2	6	34.18	22.43	26.00	17.17
49	ب	6	13	5	2	15.33	35.23	19.60	20.00
50	أ	12	6	4	4	34.83	22.00	13.00	21.50

2) معاملات الصعوبة والتمييز للاختبار الصواب والخطأ :

بالرجوع الى ملحق رقم (1-3) تباين معاملات الصعوبة حيث تراوحت معاملات الصعوبة بين (0,36) للمفردة (11) التي تعد أصعب مفردات اختبار الصواب والخطأ حيث أجاب عليها (10) أفراد فقط إجابة صحيحة ووصلت معاملات الصعوبة إلى (0.93) للمفردة (12) والتي تعد أسهل مفردات اختبار الصواب والخطأ حيث أجاب عليها (25) فرد إجابة صحيحة.

وبالرجوع ايضاً الى ملحق رقم (1-3) نجد تباين معاملات التمييز وإن كانت معظمها مرتفع إذا ما استثنينا المفردات (3) والتي لها معامل تمييز (0.11) ، والمفردة

(4) والتي لها معامل تمييز (0.18)، والمفردة (11) والتي لها معامل تمييز (0.15)، والمفردة (12) والتي لها معامل تمييز (0.16)، والمفردة (16) والتي لها معامل تمييز (صفر)، والمفردة (21) والتي لها معامل تمييز (0.10)، والمفردة (23) والتي لها معامل تمييز (0.19)، والمفردة (30) والتي لها معامل تمييز (0.03)، والمفردة (31) والتي لها معامل تمييز (0.12)، والمفردة (36) والتي لها معامل تمييز (0.06)، والمفردة (40) والتي لها معامل تمييز (-0.05)، والمفردة (45) والتي لها معامل تمييز (0.18)، والمفردة (48) والتي لها معامل تمييز (-0.10)، حيث تراوحت معاملات التمييز بين (0.20) للمفردة (43) إلي (0.66) للمفردة (22) وهي معاملات تمييز مقبولة إلى مرتفعة بعد حذف المفردات السابقة المنخفضة التمييز.

كما يتضح من ملحق رقم (1-3) أن متوسط الدرجة الكلية للطلاب الذين أجابوا على المفردة بشكل صحيح كان أكبر من متوسط الدرجة الكلية للطلاب الذين أجابوا بشكل خطأ على المفردة وذلك صحيح لجميع مفردات الاختبار عدا المفردتين 45، 48 مما يشير لأن جميع المفردات استطاعت التمييز بين الطلاب ولكن بدرجات متفاوتة.

ملحق رقم (1-3):

جدول معاملات الصعوبة والتمييز وإحصاءات مفردات اختبار الصواب والخطأ للعينة الاستطلاعية (50 مفردة)

رقم السؤال	معامل الصعوبة	معامل التمييز	عدد الطلاب المجيبين علي المفردة بشكل صحيح	عدد الطلاب المجيبين علي المفردة بشكل غير صحيح	متوسط الدرجة الكلية للطلاب المجيبين علي المفردة بشكل صحيح	متوسط الدرجة الكلية للطلاب المجيبين علي المفردة بشكل غير صحيح
1	0.79	0.48	22	6	34.00	25.83
2	0.57	0.61	16	12	35.94	27.33
3	0.57	0.11	16	12	32.94	31.33
4	0.74	0.18	20	7	32.50	29.71
5	0.57	0.53	16	12	35.44	28.00
6	0.50	0.45	14	14	35.43	29.07
7	0.79	0.22	22	6	33.05	29.33
8	0.89	0.29	25	3	32.96	26.33
9	0.75	0.29	21	7	33.43	28.71
10	0.86	0.25	24	4	32.96	28.00
11	0.36	0.15	10	18	33.70	31.44
12	0.93	0.16	25	2	32.76	28.50
13	0.54	0.34	15	13	34.47	29.69
14	0.57	0.47	16	12	35.13	28.42
15	0.61	0.27	17	11	33.76	29.91
16	0.82	0.00	23	5	32.26	32.20

رقم السؤال	معامل الصعوبة	معامل التمييز	عدد الطلاب المجيبين علي المفردة بشكل صحيح	عدد الطلاب المجيبين علي المفردة بشكل غير صحيح	متوسط الدرجة الكلية للطلاب المجيبين علي المفردة بشكل صحيح	متوسط الدرجة الكلية للطلاب المجيبين علي المفردة بشكل غير صحيح
17	0.54	0.38	15	13	34.73	29.38
18	0.86	0.26	24	4	33.00	27.75
19	0.39	0.43	11	17	36.00	29.82
20	0.50	0.47	14	14	35.57	28.93
21	0.86	0.10	24	4	32.54	30.50
22	0.36	0.66	10	18	38.40	28.83
23	0.68	0.19	19	9	33.16	30.33
24	0.82	0.35	23	5	33.39	27.00
25	0.50	0.53	14	14	35.93	28.57
26	0.50	0.36	14	14	34.79	29.71
27	0.79	0.24	22	6	33.14	29.00
28	0.54	0.42	15	13	35.00	29.08
29	0.61	0.61	17	11	35.71	26.91
30	0.82	0.03	23	5	32.35	31.80
31	0.68	0.12	19	9	32.84	31.00
32	0.71	0.54	20	8	34.65	26.25
33	0.68	0.35	19	9	33.95	28.67
34	0.71	0.27	20	8	33.45	29.25
35	0.89	0.36	25	3	33.12	25.00
36	0.86	0.06	24	4	32.42	31.25
37	0.44	0.60	12	15	36.67	28.20
38	0.64	0.31	18	10	33.89	29.30
39	0.57	0.31	16	12	34.13	29.75
40	0.67	-0.05	18	9	31.72	32.44
41	0.61	0.29	17	11	33.88	29.73
42	0.71	0.37	20	8	33.90	28.13
43	0.46	0.20	13	15	33.77	30.93
44	0.71	0.29	20	8	33.55	29.00
45	0.71	0.18	20	8	33.05	30.25
46	0.61	0.25	17	11	33.65	30.09
47	0.64	0.43	18	10	34.50	28.20
48	0.36	-0.10	10	18	31.30	32.78
49	0.63	0.28	17	10	33.24	29.30
50	0.46	0.52	13	15	36.15	28.87

3) معاملات الصعوبة والتمييز للاختبار الإكمال:

تباينت معاملات الصعوبة كماهو موضّح في ملحق رقم(1-4) حيث تراوحت معاملات الصعوبة بين (0,15) للمفردة (27) التي تعد أصعب مفردات اختبار الإكمال حيث أجاب عليها (4) أفراد فقط إجابة صحيحة ووصلت معاملات الصعوبة إلى (0.93) للمفردة (31) والتي تعد أسهل مفردات اختبار الإكمال حيث أجاب عليها (25) فرد إجابة صحيحة .

كما تباينت معاملات التمييز كماهو في ملحق رقم (1-4) وإن كانت معظمها مرتفع إذا ما استثنينا المفردات (7) والتي لها معامل تمييز (0.09)، والمفردة (31) والتي لها معامل تمييز (0.19)، حيث تراوحت معاملات التمييز بين (0.14) للمفردة (22) إلي (0.80) للمفردة (44) وهي معاملات تمييز مقبولة إلى مرتفعة بعد حذف المفردات المنخفضة التمييز السابقة الذكر.

كما نجد من ملحق رقم (1-4) أن متوسط الدرجة الكلية للطلاب الذين أجابوا على المفردة بشكل صحيح كان أكبر من متوسط الدرجة الكلية للطلاب الذين أجابوا بشكل خطأ على المفردة وذلك صحيح لجميع مفردات الاختبار مما يشير لأن جميع المفردات استطاعت التمييز بين الطلاب، ولكن بدرجات متفاوتة. وبذلك يتضح ضرورة حذف المفردات الـ (17) الآتية:

2، 3، 4، 7، 11، 12، 16، 20، 21، 23، 25، 30، 31، 36، 40، 45، 48 وبذلك يصبح العدد النهائي لمفردات الاختبارات الثلاث (الاختيار من متعدد، والصواب والخطأ، والإكمال) هو 33 مفردة.

ملحق (1-4) :

جدول معاملات الصعوبة والتمييز وإحصاءات مفردات اختبار الإكمال للعينة الاستطلاعية (50مفردة)

رقم السؤال	معامل الصعوبة	معامل التمييز	عدد الطلاب المجيبين علي المفردة بشكل صحيح	عدد الطلاب المجيبين علي المفردة بشكل غير صحيح	متوسط الدرجة الكلية للطلاب المجيبين علي المفردة بشكل صحيح	متوسط الدرجة الكلية للطلاب المجيبين علي المفردة بشكل غير صحيح
1	0.74	0.58	20	7	33.60	17.14
2	0.78	0.56	21	6	33.10	16.17
3	0.30	0.69	8	19	42.75	23.68
4	0.48	0.61	13	14	37.31	21.93
5	0.59	0.68	16	11	36.38	19.09
6	0.78	0.56	21	6	33.10	16.17
7	0.78	0.09	21	6	29.95	27.17
8	0.74	0.58	20	7	33.65	17.00
9	0.52	0.72	14	13	38.07	19.92

رقم السؤال	معامل الصعوبة	معامل التمييز	عدد الطلاب المجيبين علي المفردة بشكل صحيح	عدد الطلاب المجيبين علي المفردة بشكل غير صحيح	متوسط الدرجة الكلية للطلاب المجيبين علي المفردة بشكل صحيح	متوسط الدرجة الكلية للطلاب المجيبين علي المفردة بشكل غير صحيح
10	0.67	0.43	18	9	33.17	21.67
11	0.37	0.56	10	17	38.50	23.94
12	0.59	0.47	16	11	34.25	22.18
13	0.56	0.62	15	12	36.33	20.58
14	0.78	0.22	21	6	30.81	24.17
15	0.67	0.56	18	9	34.33	19.33
16	0.33	0.48	9	18	37.89	25.06
17	0.56	0.55	15	12	35.53	21.58
18	0.81	0.69	22	5	33.45	11.20
19	0.37	0.63	10	17	39.60	23.29
20	0.19	0.50	5	22	42.40	26.36
21	0.56	0.54	15	12	35.40	21.75
22	0.59	0.53	16	11	34.88	21.27
23	0.70	0.39	19	8	32.53	21.75
24	0.78	0.44	21	6	32.29	19.00
25	0.89	0.42	24	3	31.21	14.33
26	0.67	0.24	18	9	31.50	25.00
27	0.15	0.28	4	22	37.75	27.82
28	0.37	0.65	10	17	40.00	23.06
29	0.63	0.69	17	10	36.00	18.00
30	0.52	0.60	14	13	36.57	21.54
31	0.93	0.19	25	2	30.00	21.00
32	0.52	0.52	14	13	35.57	22.62
33	0.81	0.47	22	5	32.14	17.00
34	0.48	0.60	13	14	37.08	22.14
35	0.67	0.46	18	9	33.44	21.11
36	0.56	0.37	15	12	33.53	24.08
37	0.52	0.56	14	13	36.14	22.00
38	0.63	0.73	17	10	36.35	17.40
39	0.63	0.58	17	10	34.94	19.80
40	0.44	0.58	12	15	37.50	22.80
41	0.67	0.57	18	9	34.39	19.22
42	0.74	0.56	20	7	33.50	17.43

رقم السؤال	معامل الصعوبة	معامل التمييز	عدد الطلاب المجيبين علي المفردة بشكل صحيح	عدد الطلاب المجيبين علي المفردة بشكل غير صحيح	متوسط الدرجة الكلية للطلاب المجيبين علي المفردة بشكل صحيح	متوسط الدرجة الكلية للطلاب المجيبين علي المفردة بشكل غير صحيح
43	0.30	0.71	8	19	43.13	23.53
44	0.44	0.80	12	15	40.50	20.40
45	0.63	0.45	17	10	33.65	22.00
46	0.56	0.70	15	12	37.20	19.50
47	0.59	0.69	16	11	36.50	18.91
48	0.41	0.59	11	16	38.18	23.25
49	0.70	0.57	19	8	33.95	18.38
50	0.67	0.62	18	9	34.83	18.33

ثانياً: معامل الصعوبة والتمييز لأنماط الاختبارات في شكلها النهائي (33فقرة):

4) معاملات الصعوبة والتمييز للاختبار الاختيار من متعدد :

تباينت معاملات صعوبة مفردات الاختيار من متعدد كما هو في ملحق رقم (1-5) حيث تراوحت معاملات الصعوبة بين (0,35) للمفردة (43) التي تعد أصعب مفردات اختبار الاختيار من متعدد حيث أجاب عليها (9) أفراد فقط إجابة صحيحة ووصلت معاملات الصعوبة إلى (0.73) للمفردات (1)، (6)، (8)، (24) والتي تعد أسهل مفردات اختبار الاختيار من متعدد حيث أجاب عليها (19) فرد إجابة صحيحة.

كما تباينت معاملات التمييز في شكلها النهائي(33مفردة) ويظهر ذلك في ملحق رقم (1-5) وإن كانت معظمها مرتفع، حيث تراوحت معاملات التمييز بين (0.28) للمفردة (1) إلى (0.80) للمفردة (41) وهي معاملات تمييز مقبولة إلى مرتفعة.

ويتضح من ملحق رقم (1-5) أيضاً أن متوسط الدرجة الكلية للطلاب الذين أجابوا على المفردة بشكل صحيح كان أكبر من متوسط الدرجة الكلية للطلاب الذين أجابوا بشكل خطأ على المفردة وذلك صحيح لجميع مفردات الاختبار مما يشير لأن جميع المفردات استطاعت التمييز بين الطلاب، ولكن بدرجات متفاوتة. كما كانت قيمة معامل الثبات لكيودر -ريتشاردسون مرتفعة جدا (0.94) وانخفض الخطأ المعياري للقياس لمفردات اختبار الاختيار من متعدد وكان قيمته (2.73).

ملحق رقم (1-5):

جدول معاملات الصعوبة والتمييز وإحصاءات مفردات الاختيار من متعدد في شكله النهائي (33 مفردة)

رقم السؤال	معامل الصعوبة	معامل التمييز	عدد الطلاب المجيبين علي المفردة بشكل صحيح	عدد الطلاب المجيبين علي المفردة بشكل غير صحيح	متوسط الدرجة الكلية للطلاب المجيبين علي المفردة بشكل صحيح	متوسط الدرجة الكلية للطلاب المجيبين علي المفردة بشكل غير صحيح
1	0.73	0.28	19	7	19.05	13.29
5	0.40	0.70	10	15	24.70	11.87
6	0.73	0.48	19	7	20.16	10.29
8	0.73	0.57	19	7	20.63	9.00
9	0.54	0.59	14	12	22.50	11.67
10	0.50	0.70	13	13	23.85	11.15
13	0.36	0.81	9	16	27.67	12.56
14	0.69	0.48	18	8	20.39	11.00
15	0.58	0.64	15	11	22.47	10.73
17	0.65	0.55	17	9	21.12	10.67
18	0.58	0.66	15	11	22.67	10.45
19	0.50	0.54	13	13	22.38	12.62
22	0.60	0.47	15	10	20.33	11.90
24	0.73	0.49	19	7	20.21	10.14
26	0.69	0.39	18	8	19.89	12.13
27	0.36	0.46	9	16	23.00	14.19
28	0.38	0.47	10	16	22.90	14.13
29	0.54	0.65	14	12	23.00	11.08
32	0.58	0.59	15	11	22.07	11.27
33	0.54	0.78	14	12	24.07	9.83
34	0.50	0.54	13	13	22.38	12.62
35	0.62	0.59	16	10	21.75	10.70

رقم السؤال	معامل الصعوبة	معامل التمييز	عدد الطلاب المجيبين علي المفردة بشكل صحيح	عدد الطلاب المجيبين علي المفردة بشكل غير صحيح	متوسط الدرجة الكلية للطلاب المجيبين علي المفردة بشكل صحيح	متوسط الدرجة الكلية للطلاب المجيبين علي المفردة بشكل غير صحيح
37	0.56	0.59	14	11	21.71	11.00
38	0.38	0.70	10	16	25.60	12.44
39	0.38	0.61	10	16	24.50	13.13
41	0.42	0.80	11	15	26.00	11.27
42	0.65	0.43	17	9	20.35	12.11
43	0.35	0.42	9	17	22.78	14.71
44	0.42	0.53	11	15	23.09	13.40
46	0.46	0.68	12	14	24.17	11.79
47	0.46	0.56	12	14	23.00	12.79
49	0.50	0.70	13	13	23.92	11.08
50	0.46	0.64	12	14	23.83	12.07
KR-20 reliability = 0.94 , standard error of measurement (SEM) = 2.73						

فاعلية المشتتات لاختبار الاختيار من متعدد :

نجد من ملحق رقم (1-6) أن فاعلية المشتتات لاختبار الاختيار من متعدد في شكلها النهائي (33 مفردة)، أن متوسط الدرجة الكلية للطلاب الذين اختاروا البديل الصحيح في جميع مفردات الاختيار من متعدد كان أكبر من متوسط الدرجة الكلية للطلاب الذين اختاروا البدائل الخاطئة. وقد كان الفرق كبير بين متوسط الدرجة الكلية للطلاب الذين اختاروا البديل الصحيح ومتوسط الدرجة الكلية للطلاب الذين اختاروا البدائل الخاطئة لكل مفردات الاختيار من متعدد، وبالرغم من ذلك فقد كانت قوة تشتيت بعض البدائل لبعض المفردات ضعيفة مثل المفردات 1، 5، 14، 17، 18، 22، 34 ولكن كانت قوة تشتيت باقي بدائل تلك المفردات جيدة ولذلك كان يجب الاحتفاظ بها خاصة وأنها ذات قوة تمييز مرتفعة. أما باقي مفردات الاختبار فقد كانت قوة تشتيت جميع بدائلها مرتفعة بالإضافة لخصائصها السيكومترية الجيدة التي تم عرضها فيما سبق.

ملحق رقم (1-6) :

جدول قوة تشتيت البدائل لمفردات اختبار الاختيار من متعدد في شكله النهائي (33 مفردة)

رقم السؤال	الإجابة الصحيحة	عدد الذين اختاروا البديل (أ)	عدد الذين اختاروا البديل (ب)	عدد الذين اختاروا البديل (ج)	عدد الذين اختاروا البديل (د)	متوسط الدرجة الكلية للطلاب الذين اختاروا البديل (أ)	متوسط الدرجة الكلية للطلاب الذين اختاروا البديل (ب)	متوسط الدرجة الكلية للطلاب الذين اختاروا البديل (ج)	متوسط الدرجة الكلية للطلاب الذين اختاروا البديل (د)
1	أ	19	5	0	2	19.05	14.20		11.00
5	جـ	8	7	10	0	11.00	12.86	24.70	
6	د	1	2	4	19	8.00	11.00	10.50	20.16
8	جـ	2	3	19	2	10.00	9.00	20.63	8.00
9	جـ	9	2	14	1	10.56	19.00	22.50	7.00
10	ب	6	13	5	2	11.33	23.85	11.00	11.00
13	جـ	8	6	9	2	10.63	15.83	27.67	10.50
14	د	4	4	0	18	11.75	10.25		20.39
15	جـ	5	3	15	3	10.40	13.00	22.47	9.00
17	جـ	2	7	17	0	8.50	11.29	21.12	
18	ب	9	15	2	0	10.44	22.67	10.50	
19	ب	7	13	2	4	16.43	22.38	8.00	8.25
22	جـ	7	3	15	0	12.29	11.00	20.33	
24	ب	1	19	4	2	10.00	20.21	9.75	11.00
26	جـ	3	3	18	2	11.33	10.67	19.89	15.50
27	أ	9	8	6	2	23.00	12.75	14.17	20.00
28	ب	3	10	8	5	11.00	22.90	17.38	10.80
29	جـ	6	3	14	3	11.17	8.33	23.00	13.67
32	جـ	4	5	15	2	10.75	12.20	22.07	10.00
33	د	4	4	4	14	10.25	8.75	10.50	24.07
34	جـ	0	11	13	2		13.55	22.38	7.50
35	جـ	2	5	16	3	13.00	11.20	21.75	8.33
37	د	2	6	3	14	12.50	11.67	8.67	21.71

رقم السؤال	الإجابة الصحيحة	عدد الذين اختاروا البديل (أ)	عدد الذين اختاروا البديل (ب)	عدد الذين اختاروا البديل (ج)	عدد الذين اختاروا البديل (د)	متوسط الدرجة الكلية للطلاب الذين اختاروا البديل (أ)	متوسط الدرجة الكلية للطلاب الذين اختاروا البديل (ب)	متوسط الدرجة الكلية للطلاب الذين اختاروا البديل (ج)	متوسط الدرجة الكلية للطلاب الذين اختاروا البديل (د)
38	ب	7	10	5	4	11.29	25.60	13.80	12.75
39	أ	10	7	4	5	24.50	11.43	15.25	13.80
41	د	4	6	5	11	11.00	13.67	8.60	26.00
42	ب	3	17	2	4	13.67	20.35	10.50	11.75
43	ب	6	9	7	4	17.33	22.78	12.14	15.25
44	جـ	8	3	11	4	14.25	14.67	23.09	10.75
46	أ	12	4	4	6	24.17	12.00	9.75	13.00
47	ب	6	12	6	2	9.00	23.00	17.67	9.50
49	ب	6	13	5	2	10.00	23.92	11.80	12.50
50	أ	12	6	4	4	23.83	14.17	7.25	13.75

1) معاملات الصعوبة والتمييز للاختبار الصواب والخطأ:

يبين ملحق رقم (1-7) تباين معاملات صعوبة مفردات الصواب والخطأ في شكلها النهائي (33 مفردة) حيث تراوحت معاملات الصعوبة بين (0,36) للمفردة (22) التي تعد أصعب مفردات اختبار الصواب والخطأ حيث أجاب عليها (10) أفراد فقط إجابة صحيحة ووصلت معاملات الصعوبة إلى (0.89) للمفردة (35) والتي تعد أسهل مفردات اختبار الصواب والخطأ حيث أجاب عليها (25) فرد إجابة صحيحة.

كما تباينت معاملات التمييز في شكلها النهائي(33مفردة) كما في ملحق رقم (1-7) وإن كانت معظمها مرتفع، حيث تراوحت معاملات التمييز بين (0.20) للمفردة (46) إلى (0.64) للمفردة (29) وهي معاملات تمييز مقبولة إلى مرتفعة.

ويتضح من ملحق رقم (1-7) أن متوسط الدرجة الكلية للطلاب الذين أجابوا على المفردة بشكل صحيح كان أكبر من متوسط الدرجة الكلية للطلاب الذين أجابوا بشكل خطأ على المفردة وذلك صحيح لجميع مفردات الاختبار مما يشير لأن جميع المفردات استطاعت التمييز بين الطلاب، ولكن بدرجات متفاوتة. كما كانت قيمة معامل الثبات لكيودر ريتشاردسون مرتفعة جدا (0.83) وانخفض الخطأ المعياري للقياس لمفردات اختبار الصواب والخطأ وكان قيمته (2.51).

ملحق رقم (1-7) :

جدول معاملات الصعوبة والتمييز وإحصاءات مفردات الصواب والخطأ في شكله النهائي (33 مفردة)

رقم السؤال	معامل الصعوبة	معامل التمييز	عدد الطلاب المجيبين علي المفردة بشكل صحيح	عدد الطلاب المجيبين علي المفردة بشكل غير صحيح	متوسط الدرجة الكلية للطلاب المجيبين علي المفردة بشكل صحيح	متوسط الدرجة الكلية للطلاب المجيبين علي المفردة بشكل غير صحيح
1	0.79	0.52	22	6	22.55	15.00
5	0.57	0.47	16	12	23.38	17.67
6	0.50	0.49	14	14	23.86	18.00
8	0.89	0.29	25	3	21.52	16.00
9	0.75	0.26	21	7	21.81	18.29
10	0.86	0.34	24	4	21.75	16.00
13	0.54	0.42	15	13	23.27	18.23
14	0.57	0.56	16	12	23.81	17.08
15	0.61	0.33	17	11	22.53	18.45
17	0.54	0.36	15	13	22.93	18.62
18	0.86	0.23	24	4	21.50	17.50
19	0.39	0.37	11	17	23.64	19.18
22	0.36	0.62	10	18	25.90	18.17
24	0.82	0.39	23	5	22.00	16.00
26	0.50	0.32	14	14	22.86	19.00
27	0.79	0.32	22	6	21.91	17.33
28	0.54	0.43	15	13	23.33	18.15
29	0.61	0.64	17	11	24.00	16.18
32	0.71	0.59	20	8	23.15	15.38
33	0.68	0.30	19	9	22.16	18.33
34	0.71	0.26	20	8	21.90	18.50
35	0.89	0.27	25	3	21.48	16.33
37	0.44	0.60	12	15	24.67	17.53

رقم السؤال	معامل الصعوبة	معامل التمييز	عدد الطلاب المجيبين علي المفردة بشكل صحيح	عدد الطلاب المجيبين علي المفردة بشكل غير صحيح	متوسط الدرجة الكلية للطلاب المجيبين علي المفردة بشكل صحيح	متوسط الدرجة الكلية للطلاب المجيبين علي المفردة بشكل غير صحيح
38	0.64	0.42	18	10	22.78	17.60
39	0.57	0.37	16	12	22.81	18.42
41	0.61	0.32	17	11	22.47	18.55
42	0.71	0.39	20	8	22.40	17.25
43	0.46	0.29	13	15	22.77	19.33
44	0.71	0.36	20	8	22.30	17.50
46	0.61	0.20	17	11	21.88	19.45
47	0.64	0.42	18	10	22.78	17.60
49	0.63	0.38	17	10	22.24	17.70
50	0.46	0.53	13	15	24.31	18.00

KR-20 reliability = 0.83 , standard error of measurement (SEM) = 2.51

2) معاملات الصعوبة والتمييز للاختبار الإكمال:

تباينت معاملات صعوبة مفردات الإكمال في شكلها النهائي(33مفردة) كما في ملحق رقم (1-8) حيث تراوحت معاملات الصعوبة بين (0,15) للمفردة (27) التي تعد أصعب مفردات اختبار الإكمال حيث أجاب عليها (4) أفراد فقط إجابة صحيحة ووصلت معاملات الصعوبة إلى (0.81) للمفردتين (18)، (33) والتي تعد أسهل مفردات اختبار الإكمال حيث أجاب عليها (22) فرد إجابة صحيحة.

يوضّح ملحق رقم (1-8) أن معاملات التمييز في شكلها النهائي(33مفردة) تباينت وإن كانت معظمها مرتفع، حيث تراوحت معاملات التمييز بين (0.22) للمفردة (26) إلى (0.80) للمفردة (44) وهي معاملات تمييز مقبولة إلى مرتفعة.

و يتضح من ملحق رقم (1-8) أن متوسط الدرجة الكلية للطلاب الذين أجابوا على المفردة بشكل صحيح كان أكبر من متوسط الدرجة الكلية للطلاب الذين أجابوا بشكل خطأ على المفردة وذلك صحيح لجميع مفردات الاختبار مما يشير لأن جميع المفردات استطاعت التمييز بين الطلاب ولكن بدرجات متفاوتة. كما كانت قيمة معامل الثبات لكيودر -ريتشاردسون مرتفعة جدا (0.94) وانخفض الخطأ المعياري للقياس لمفردات اختبار الإكمال وكان قيمته (2.25).

ملحق رقم (1-8):

معاملات الصعوبة والتمييز وإحصاءات مفردات الإكمال في شكله النهائي (33 مفردة)

رقم السؤال	معامل الصعوبة	معامل التمييز	عدد الطلاب المجيبين علي المفردة بشكل صحيح	عدد الطلاب المجيبين علي المفردة بشكل غير صحيح	متوسط الدرجة الكلية للطلاب المجيبين علي المفردة بشكل صحيح	متوسط الدرجة الكلية للطلاب المجيبين علي المفردة بشكل غير صحيح
1	0.74	0.60	20	7	23.05	10.86
5	0.59	0.71	16	11	25.13	12.27
6	0.78	0.54	21	6	22.48	10.83
8	0.74	0.62	20	7	23.15	10.57
9	0.52	0.73	14	13	26.14	13.15
10	0.67	0.49	18	9	23.00	13.67
13	0.56	0.64	15	12	25.00	13.50
14	0.78	0.24	21	6	21.05	15.83
15	0.67	0.58	18	9	23.56	12.56
17	0.56	0.52	15	12	24.07	14.67
18	0.81	0.69	22	5	22.82	7.00
19	0.37	0.61	10	17	27.00	15.71
22	0.59	0.55	16	11	23.94	14.00
24	0.78	0.46	21	6	22.10	12.17
26	0.67	0.22	18	9	21.28	17.11
27	0.15	0.30	4	22	26.25	18.68
28	0.37	0.64	10	17	27.30	15.53
29	0.63	0.66	17	10	24.41	12.20
32	0.52	0.50	14	13	24.21	15.23
33	0.81	0.47	22	5	21.91	11.00
34	0.48	0.62	13	14	25.62	14.57
35	0.67	0.41	18	9	22.50	14.67
37	0.52	0.57	14	13	24.79	14.62

رقم السؤال	معامل الصعوبة	معامل التمييز	عدد الطلاب المجيبين علي المفردة بشكل صحيح	عدد الطلاب المجيبين علي المفردة بشكل غير صحيح	متوسط الدرجة الكلية للطلاب المجيبين علي المفردة بشكل صحيح	متوسط الدرجة الكلية للطلاب المجيبين علي المفردة بشكل غير صحيح
38	0.63	0.75	17	10	25.06	11.10
39	0.63	0.63	17	10	24.24	12.50
41	0.67	0.59	18	9	23.61	12.44
42	0.74	0.59	20	7	23.00	11.00
43	0.30	0.71	8	19	29.63	15.79
44	0.44	0.80	12	15	27.92	13.47
46	0.56	0.72	15	12	25.67	12.67
47	0.59	0.67	16	11	24.81	12.73
49	0.70	0.63	19	8	23.53	11.25
50	0.67	0.62	18	9	23.78	12.11

KR-20 reliability = 0.94 , standard error of measurement (SEM) = 2.25

ثالثاً: معامل ثبات درجات الاختبار وصدق الإتساق الداخلي لأنماط الاختبارات في شكلها النهائي (33فقرة):

1) معامل ثبات نمط الاختيار من متعدد :

يبين ملحق رقم (1-9) ارتفاع معاملات الارتباط بين درجة المفردة والدرجة الكلية للاختبار بعد حذف الدرجة المفردة للاختبار والذي استخدم كمحك الصدق هنا ، حيث تراوحت قيم معاملات الاتساق الداخلي بين 0.235 (المفردة 1) و 0.792 (المفردة 13)، كما كانت قيم معاملات الثبات مرتفعة، حيث كانت قيمة الثبات بطريقة ألفا هي (0.935) والتجزئة النصفية بعد تصحيح سبيرمان براون (0.957) والتجزئة النصفية بعد تصحيح جوتمان (0.954). كما كانت قيم معاملات الثبات بطريقة ألفا بعد حذف أي مفردة متقاربة مما يدل على أن جميع المفردات تساهم بدرجة كبيرة في ثبات درجات الاختبار واتساقه الداخلي. وتشير تلك النتائج لوصول درجات الاختبار لدرجة مقبولة من الثبات والصدق ويمكن للباحث الاطمئنان لدرجاتها عند التحقق من صحة فروض الدراسة.

ملحق رقم (1-9):

جدول مؤشر صدق الاتساق الداخلي وقيمة معاملات ثبات درجات الاختبار بطرق ألفا والتجزئة النصفية (سبيرمان– براون،وجتمان) لمفردات اختبار الاختيار من متعدد في صورته النهائية (33 مفردة)

رقم السؤال	الاتساق الداخلي	قيمة ألفا في حالة حذف المفردة
1	.235	.936
5	.590	.933
6	.442	.934
8	.532	.933
9	.555	.933
10	.666	.932
13	.792	.931
14	.435	.934
15	.602	.933
17	.507	.934
18	.630	.932
19	.495	.934
22	.315	.936
24	.452	.934
26	.349	.935
27	.395	.935
28	.425	.935
29	.619	.932
32	.548	.933
33	.756	.931
34	.495	.934
35	.553	.933

رقم السؤال	الاتساق الداخلي	قيمة ألفا في حالة حذف المفردة
37	.457	.934
38	.674	.932
39	.572	.933
41	.778	.931
42	.386	.935
43	.377	.935
44	.484	.934
46	.646	.932
47	.519	.934
49	.675	.932
50	.609	.933

Alpha = 0.935, Spearman-Brown Coefficient = 0.957, Guttman Split-Half Coefficient = 0.954

2) معامل ثبات نمط الصواب والخطأ:

يبين ملحق رقم (1-10) ارتفاع معاملات الارتباط بين درجة المفردة والدرجة الكلية للاختبار بعد حذف الدرجة المفردة للاختبار والذي استخدم كمحك الصدق هنا، حيث تراوحت قيم معاملات الاتساق الداخلي بين 0.239 (المفردة 9) و0.505 (المفردة 37)، كما كانت قيم معاملات الثبات مرتفعة، حيث كانت قيمة الثبات بطريقة ألفا هي (0.807) والتجزئة النصفية بعد تصحيح سبيرمان براون (0.860) والتجزئة النصفية بعد تصحيح جوتمان (0.860). كما كانت قيم معاملات الثبات بطريقة ألفا بعد حذف أي مفردة متقاربة مما يدل على أن جميع المفردات تساهم بدرجة كبيرة في ثبات درجات الاختبار واتساقه الداخلي. وتشير تلك النتائج لوصول درجات الاختبار لدرجة مقبولة من الثبات والصدق ويمكن للباحث الاطمئنان لدرجاتها عند التحقق من صحة فروض الدراسة.

ملحق رقم (1-10) :

جدول مؤشر صدق الاتساق الداخلي وقيمة معاملات ثبات درجات الاختبار بطريقة ألفا و التجزئة النصفية (سبيرمان– براون، وجتمان) لمفردات اختبار الصواب والخطأ في صورته النهائية (33 مفردة)

رقم السؤال	الاتساق الداخلي	قيمة ألفا في حالة حذف المفردة
1	.452	.797
5	.353	.800
6	.362	.800
8	.220	.805
9	.239	.808
10	.264	.804
13	.285	.803
14	.456	.796
15	.299	.802
17	.310	.802
18	.248	.807
19	.293	.806
22	.504	.795
24	.307	.802
26	.263	.808
27	.216	.805
28	.299	.802
29	.562	.792
32	.519	.794
33	.272	.807
34	.313	.802

رقم السؤال	الاتساق الداخلي	قيمة ألفا في حالة حذف المفردة
35	.298	.805
37	.505	.794
38	.299	.802
39	.325	.801
41	.284	.803
42	.284	.803
43	.205	.806
44	.253	.804
46	.242	.812
47	.299	.802
49	.284	.803
50	.401	.798

Alpha = 0.807, Spearman-Brown Coefficient = 0.860, Guttman Split-Half Coefficient = 0.860

3) معامل ثبات نمط الإكمال:

يتضح من ملحق رقم (1-11) أن ارتفاع معاملات الارتباط بين درجة المفردة والدرجة الكلية للاختبار بعد حذف الدرجة المفردة للاختبار والذي استخدم كمحك الصدق هنا، حيث تراوحت قيم معاملات الاتساق الداخلي بين 0.263 (المفردة 27) و 0.802 (المفردة 44)، كما كانت قيم معاملات الثبات مرتفعة، حيث كانت قيمة الثبات بطريقة ألفا هي (0.942) والتجزئة النصفية بعد تصحيح سبيرمان براون (0.951) والتجزئة النصفية بعد تصحيح جوتمان (0.947). كما كانت قيم معاملات الثبات بطريقة ألفا بعد حذف أي مفردة متقاربة مما يدل على أن جميع المفردات تساهم بدرجة كبيرة في ثبات درجات الاختبار واتساقه الداخلي. وتشير تلك النتائج لوصول درجات الاختبار لدرجة مقبولة من الثبات والصدق ويمكن للباحث الاطمئنان لدرجاتها عند التحقق من صحة فروض الدراسة.

ملحق رقم (1-11) :

جدول مؤشر صدق الاتساق الداخلي وقيمة معاملات ثبات درجات الاختبار بطرق ألفا والتجزئة النصفية (سبيرمان– براون، وجتمان) لمفردات اختبار الإكمال في صورته النهائية(33 مفرد)

رقم السؤال	الاتساق الداخلي	قيمة ألفا في حالة حذف المفردة
1	.567	.940
5	.708	.939
6	.563	.940
8	.587	.940
9	.709	.939
10	.453	.941
13	.629	.940
14	.297	.943
15	.547	.941
17	.502	.941
18	.665	.940
19	.587	.940
22	.511	.941
24	.473	.941
26	.268	.944
27	.263	.943
28	.621	.940
29	.633	.940
32	.465	.941
33	.440	.941
34	.591	.940
35	.370	.942

رقم السؤال	الاتساق الداخلي	قيمة ألفا في حالة حذف المفردة
37	.536	.941
38	.766	.938
39	.605	.940
41	.556	.940
42	.557	.940
43	.691	.939
44	.802	.938
46	.704	.939
47	.661	.939
49	.638	.940
50	.585	.940

Alpha = 0.942, Spearman-Brown Coefficient = 0.951, Guttman Split-Half Coefficient = 0.947

ويمكن تلخيص معاملات الثبات للأنماط الثلاثة في الجدول رقم (6) التالي:

معاملات الثبات للأنماط الثلاثة للعينة الاستطلاعية في الشكل النهائي(33مفردة)

	معاملات الثبات		
نمط الأختبار	ألفا	سبيرمان-براون	جوتمان
1)الاختيار من متعدد	0.935	0.957	0.954
2)الصواب والخطأ	0.807	0.860	0.860
3)الإكمال	0.942	0.951	0.947

عاشراً: الصورة النهائية للأختبار :

بعد أن تم حذف المفردات الـ (17) الآتية:

2، 3، 4، 7، 11، 12، 16، 20، 21، 23، 25، 30، 31، 36، 40، 45، 48 وبـــــــــذلك يـــصبح العـــدد النهـــائي لمفـــردات الاختبـــارات الـــثلاث (الاختيـــار مـــن متعـــدد، والـــصواب والخطأ، والإكمال) هو 33 مفردة.

الحادي عشر: التطبيق النهائي:

تم مراجعة المقياس الاختبار التحصيلي بالثلاثة الأنماط المختلفة حيث تم الحصول على الموافقة لتطبيق الاختبار: حصل الباحث على خطاب من كلية التربية موجه إلى الإدارة العامة للتربية والتعليم بمنطقة مكة المكرمة وتم اختيار المدارس الموضحة وتوفير أدوات الدراسة وتم تحديد موعد تطبيق الاختبار بجدول زمني.

خامساً: الأساليب الإحصائية:

1) الأساليب الإحصائية المستخدمة في اختبار فرضيات الدراسة:

2) حساب معامل الصعوبة التقليدي لكل مفردة من مفردات الاختبارات الثلاث (اختيار من متعدد، صواب وخطأ، إكمال) من خلال النسبة بين من أجابوا على المفردة إجابة صحيحة مقسوماً على عدد الطلاب الذين حاولوا الإجابة على المفردة.

3) حساب معامل التمييز لكل مفردة من مفردات الاختبارات الثلاث (اختيار من متعدد، صواب وخطأ، إكمال) وذلك من خلال معامل الارتباط بين درجة الطلاب على كل فقرة ودرجاتهم الكلية على الاختبار باستخدام معامل الارتباط التسلسلي الأصيل(النقطي) **Point Biserial Correlation** والتي تعد مؤشرا لتمييز المفردة الاختبارية (النبهان، 2004م).

4) حساب معاملات ثبات الاتساق الداخلي لكل اختبار من الاختبارات الثلاث (اختيار من متعدد، صواب وخطأ، إكمال) من خلال طريقة ألفا –كرونباخ.

$$r_{xx} = \frac{k}{k-1}\left[1 - \frac{\sum {s_i}^2}{{s_x}^2}\right]$$

حيث:

k : عدد فقرات الاختبار.

S_i^2 : تباين الفقرة الواحدة

${s_x}^2$: تباين كل الاختبار.

4) استخدام الأسلوب الإحصائي تحليل التباين الثنائي نموذج التأثيرات الثابتة (Two-Way ANOVA,Fixed Effects Model) ذو التصميم العاملي (2×3) حيث كان المتغير المستقل الأول هو انماط الاختبارات و كان المتغير المستقل الثاني هو الجنس(ذكور / اناث) ودراسة التفاعل بينهم.

5) استخدام تحليل التباين الثنائي **Two-way ANOVA** للتحقق من الفروق بين متوسطات معاملات الصعوبة وفقا لنوع الاختبار (اختيار من متعدد، صواب

وخطأ، إكمال) وجنس المستجيب (ذكر – أنثي) والتفاعل بينهما وذلك بعد تحويل قيم معاملات الصعوبة للفقرات إلي قيما دلتا Δ المقابلة والتي تعد تحويلا خطيا بمتوسط = 13 وانحراف معياري = 4 بعد تحويل معامل الصعوبة لما يقابله من درجة معيارية وبذلك تكون معاملات الصعوبة على مقياس فئوي لسهولة تفسير النتائج ومعادلتها هي: $\Delta = 4Z + 13$ (علام،2002م).

6) استخدام تحليل التباين الثنائي **Two-way ANOVA** للتحقق من الفروق بين متوسطات معاملات التمييز وفقا لنوع الاختبار (اختيار من متعدد، صواب وخطأ، إكمال) وجنس المستجيب (ذكر – أنثي) والتفاعل بينهما وذلك بعد تحويل قيم معاملات التمييز إلى معاملات فيشر الزائية(z) **Fisher r-to-z transformation** وذلك لضمان اعتدالية شكل توزيع معاملات التمييز والتي هي في الأصل معاملات ارتباط والتي تكون ذات توزيعات ملتوية كلما زادت قيمة معاملات الارتباط.

7) استخدام الإحصائي **(M) (Hakstain& E.Whalen, 1976)** للتحقق من وجود فروق دالة بين معاملات الثبات للاتساق الداخلي بطريقة ألفا، وتتبع هذه الإحصاءه توزيع مربع كاي بدرجة حرية = عدد معاملات الثبات -1

$$M = \left[\sum_{K=1}^{K} B_K - \frac{\left[\sum_{K=1}^{K} B_K (1 - r_K)^{-1/3} \right]^2}{\sum_{K=1}^{K} (1 - r_K)^{-2/3}} \right]$$

$$B_K = \frac{(J_K - 1)}{18 J_K} \frac{(9n_K - 11)^2}{(n_K - 1)}$$

حيث n_k ، k عدد النماذج، J عدد مفردات النموذج، r معامل الثبات.

نتائج الدراسة:

يتناول الباحث عرضاً لنتائج الدراسة ومناقشتها، وفقاً لأسئلة الدراسة وتحقيقاً لأهدافها وذلك من خلال التحقق من فروضها، على النحو التالي :

عرض نتائج الدراسة ومناقشتها:

أولاً: اختبار الفرض الأول للدراسة :

ينص فرض الدراسة الأول على أنه " لا يوجد أثر دال إحصائيا للتفاعل بين نمط الاختبار (اختيار من متعدد، صواب وخطأ، إكمال) وجنس المستجيب (ذكور، إناث) على معامل صعوبة الفقرة ".

قبل استخدام تحليل التباين قام الباحث بالتحقق من افتراضات تحليل التباين: الاعتدالية وتجانس التباين. وقد قام الباحث بالتحقق من افتراض الاعتدالية

باستخدام اختبار شابيرو-ولك والذي كانت قيمته 0.051 وهي قيمة غير دالة إحصائيا عند مستوي 0.05. كما قام الباحث بالتحقق من عدم انتهاك افتراض تجانس التباين باستخدام اختبار ليفين لاختبار تساوي تباينات المجتمعات، وقد كانت قيمة اختبار ليفين غير دالة إحصائيا حيث كانت قيمة ف = 0.053 (5، 192) وهي قيمة غير دالة إحصائيا (الاحتمال = 0.998).

وللتحقق من صحة هذا الفرض قام الباحث بتحويل معاملات الصعوبة للفقرات إلى قيم دلتا Δ المناظرة والتي تعتبر درجات معيارية بمتوسط قيمته 13 وانحراف معياري قيمته 4 حيث أن قيم دلتا Δ تقع على مقياس فئوي وهو ما يتيح حساب متوسطات معاملات الصعوبة وتحليل البيانات باستخدام تحليل التباين. وقد قام الباحث بعرض معاملات صعوبة جميع فقرات الاختبارات الثلاثة (اختيار من متعدد، صواب وخطأ، إكمال) بعد تحويلها إلى قيم دلتا Δ المناظرة طبقاً لجنس المستجيب(ذكور، إناث) في معاملات الصعوبة ملحق رقم(2-1) ثم قام الباحث بحساب دلالة الفروق بين متوسطات معامل الصعوبة بعد تحويلها إلى قيم دلتا Δ المناظرة وفقا لنمط الاختبار والجنس ،والتفاعل بينهما في جدول (6).

ويتضح من معاملات الصعوبة كما في ملحق رقم (2-1) تباين معاملات الصعوبة للفقرات، حيث تراوحت معاملات الصعوبة للإناث في نمط أسئلة الاختيار من متعدد بعد تحويلها لقيم دلتا بين 4,73 للمفردة 33 إلى 20,38 للمفردة 8 بمتوسط 12,88 وانحراف معياري 4,05 في حين تراوحت معاملات الصعوبة للذكور في نمط أسئلة الاختيار من متعدد بعد تحويلها لقيم دلتا بين 5,41 للمفردة 17 إلى 20,59 للمفردة 1 بمتوسط 13,01 وانحراف معياري 4,06. كما تراوحت معاملات الصعوبة للإناث في نمط أسئلة الصواب والخطأ بعد تحويلها لقيم دلتا بين 4,97 للمفردة 26 إلى 20,68 للمفردة 30 بمتوسط 12,97 وانحراف معياري 4,09 في حين تراوحت معاملات الصعوبة للذكور في نمط أسئلة الصواب والخطأ بعد تحويلها لقيم دلتا بين 2,70 للمفردة 13 إلى 19,67 للمفردة 11 بمتوسط 13,12 وانحراف معياري 4,08. وأخيراً تراوحت معاملات الصعوبة للإناث في نمط أسئلة الإكمال بعد تحويلها لقيم دلتا بين 4,32 للمفردة 33 إلى 19,75 للمفردة 8 بمتوسط 13,00 وانحراف معياري 4,00 في حين تراوحت معاملات الصعوبة للذكور في نمط أسئلة الإكمال بعد تحويلها لقيم دلتا بين 6,03 للمفردة 29 إلى 20,42 للمفردة 1 بمتوسط 12,93 وانحراف معياري 3.91.

و من معاملات الصعوبة في ملحق رقم(2-1) يلاحظ الباحث أنه بالرغم من تقارب متوسطات معاملات الصعوبة للفقرات بعد تحويلها لقيم دلتا وقيم انحرافاتها المعيارية بين الجنسين وبين أنماط الاختبارات الثلاث إلا أن هناك تباين في معاملات صعوبة الفقرات على مستوي أنماط الاختبارات فلم تكن أسهل مفردات اختبار الاختيار من متعدد هي الأسهل بالنسبة للنمطين الآخرين للاختبار، كذلك لم تكن أصعب مفردات اختبار الاختيار من متعدد هي الأصعب بالنسبة للنمطين الآخرين

للاختبار ولذلك فإن الترتيب النسبي للمفردات من حيث معاملات الصعوبة على مستوي الأنماط الثلاث للمفردات لم يكن متساوي. وعلى نفس المنوال فقد اختلف الترتيب النسبي للمفردات وفقا للجنس فلم تكن أسهل المفردات لدى الذكور هي الأسهل لدى الإناث في كل نمط من أنماط الاختبار وكذلك لم تكن أصعب المفردات لدى الذكور هي الأصعب لدى الإناث في كل نمط من أنماط الاختبار.

ملحق رقم (2-1) :

جدول معاملات الصعوبة(Δ) لمفردات الاختبارات الثلاث (اختيار من متعدد، صواب وخطأ، إكمال) وفقا لمتغير الجنس للعينة

رقم المفردة	الاختيار من متعدد		الصواب والخطأ		الإكمال	
	ذكور	إناث	ذكور	إناث	ذكور	إناث
1	20.59	11.84	19.32	13.31	20.42	16.05
2	9.92	12.91	13.44	14.91	12.47	12.35
3	16.90	17.18	12.40	12.99	19.29	15.43
4	17.31	18.96	16.90	18.12	11.71	13.58
5	9.51	13.27	14.82	13.95	6.79	9.57
6	15.26	11.49	10.32	5.29	16.26	15.43
7	6.64	7.22	4.43	6.90	8.30	14.51
8	18.95	20.38	9.63	14.91	17.02	19.75
9	11.56	11.84	10.67	8.50	11.33	7.41
10	18.95	15.04	6.86	11.38	12.85	10.49
11	14.85	17.89	19.67	18.76	15.88	17.59
12	9.10	15.40	11.70	16.19	12.85	12.65
13	15.26	9.36	2.70	5.94	13.61	16.05
14	16.90	18.24	17.25	17.15	17.02	19.44
15	14.44	18.60	10.32	15.23	14.74	13.89
16	7.87	11.13	9.97	8.18	8.30	6.79
17	5.41	8.29	11.36	12.03	7.55	7.41
18	13.21	12.56	17.94	15.87	12.85	11.73

رقم المفردة	الاختيار من متعدد		الصواب والخطأ		الإكمال	
	ذكور	إناث	ذكور	إناث	ذكور	إناث
19	16.08	14.69	17.94	17.15	17.39	14.51
20	14.03	18.24	15.51	17.47	17.77	17.28
21	10.74	10.78	15.86	16.51	17.02	10.80
22	18.54	16.47	14.48	16.19	15.12	17.90
23	10.33	7.58	9.63	11.71	7.92	8.33
24	8.69	15.04	14.48	10.74	11.71	12.96
25	7.87	8.29	17.94	11.06	9.44	11.73
26	12.79	13.27	11.36	4.97	12.85	15.12
27	15.67	9.71	13.44	10.42	13.61	13.58
28	11.97	10.42	11.01	11.06	6.79	4.32
29	5.82	6.87	13.09	9.46	6.03	6.79
30	12.79	15.40	17.94	20.68	17.02	18.21
31	16.49	9.00	14.48	15.23	9.44	11.11
32	13.21	12.91	15.51	15.23	13.98	16.67
33	11.56	4.73	10.67	10.42	11.33	9.57
المتوسط	13.01	12.88	13.12	12.97	12.93	13.00
الانحراف المعياري	4.06	4.05	4.08	4.09	3.91	4.00
عدد الطلاب	75	75	78	77	73	68

دلالة الفروق بين المتوسطات:

للتحقق من دلالة الفروق بين المتوسطات الفروق بين متوسطات معامل الصعوبة بعد تحويلها إلى قيم دلتا Δ المناظرة وفقا لنمط الاختبار والجنس والتفاعل بينهما، قام الباحث باختبار تحليل التباين الثنائي لنمط الاختبار وجنس المستجيب والتفاعل بينهما وتم تلخيص نتائج تحليل التباين الثنائي في جدول (7).

جدول (7) :

دلالة الفروق بين المتوسطات الفروق بين متوسطات معامل الصعوبة بعد تحويلها إلى قيم دلتا المناظرة وفقا لنمط الاختبار والجنس والتفاعل بينهما

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة ف	الدلالة الاحصائية
نمط الاختبار	.381	2	.191	.012	.988
جنس المستجيب	.247	1	.247	.015	.902
التفاعل	.503	2	.251	.015	.985
الخطأ	3121.233	192	16.256		
المجموع	3122.364	197			

يتضح من جدول (7) عدم وجود فروق ذات دلالة إحصائية بين متوسطات معاملات الصعوبة ترجع لنمط الاختبارات الثلاث، فقد كانت قيمة ف المحسوبة صغيرة جداً (0.012) وهي قيمة غير دالة إحصائيا عند مستوي (0.05). وتعد هذه النتيجة متوقعة إذا ما راجعنا إلى قيم متوسطات معاملات الصعوبة لأنماط الاختبارات الثلاثة (الاختيار من متعدد، والصواب والخطأ والإكمال) كما هو في ملحق رقم(2-1) حيث كانت قيمة متوسط معاملات الصعوبة لنمط الاختيار من متعدد هو 12.94 وكان متوسط معاملات الصعوبة لنمط الصواب والخطأ هو 13,04 وكان متوسط معاملات الصعوبة لنمط الإكمال هو 12,96 وهي قيم متقاربة جداً مما أدى لعدم دلالة الفروق بين تلك المتوسطات.

كما يتضح من جدول (7) عدم وجود فروق ذات دلالة إحصائية بين متوسطات معاملات الصعوبة ترجع لجنس المستجيب، فقد كانت قيمة ف المحسوبة صغيرة جداً (0,015) وهي قيمة غير دالة إحصائياً عند مستوى (0,05). وتعد هذه النتيجة متوقعة إذا ما راجعنا إلى قيم متوسطات معاملات الصعوبة وفقا لجنس المستجيب (ذكور، إناث) كما هو موضّح في ملحق رقم(2-1) حيث كانت قيمة متوسط معاملات الصعوبة للذكور هو 12.95 وكان متوسط معاملات الصعوبة للإناث هو 13,02 وهي قيم متقاربة جداً مما أدى لعدم دلالة الفروق بين تلك المتوسطات.

ويتضح كذلك من جدول (7) عدم دلالة التفاعل بين نمط الاختبار وجنس المستجيب فقد كانت قيمة ف المحسوبة صغيرة جداً (0,015) وهي قيمة غير دالة إحصائياً عند مستوى (0,05). وتعد هذه النتيجة متوقعة إذا ما راجعنا إلى قيم متوسطات معاملات الصعوبة وفقا لجنس المستجيب (ذكور، إناث) لدى كل نمط من أنماط الاختبار كماهو في ملحق رقم (2-1) حيث كانت قيمة متوسط معاملات الصعوبة للذكور في نمط الاختيار من متعدد هو 13.01 وكان متوسط معاملات الصعوبة للإناث في نفس النمط هو 12,88 وهي قيم متقاربة جداً، كذلك كانت قيمة

متوسط معاملات الصعوبة للذكور في نمط الصواب والخطأ هو 13.12 وكان متوسط معاملات الصعوبة للإناث في نفس النمط هو 12,97 وهي قيم متقاربة جداً ، كذلك كانت قيمة متوسط معاملات الصعوبة للذكور في نمط الإكمال هو 12.93 وكان متوسط معاملات الصعوبة للإناث في نفس النمط هو 13,00 وهي قيم متقاربة جداً أيضاً مما أدى لعدم دلالة الفروق بين تلك المتوسطات. وبذلك يمكن القول بتحقق فرض الدراسة الأول.

مناقشة نتائج اختبار الفرض الأول :

توصلت نتائج الدراسة إلى تحقق فرض الدراسة الأول والذي ينص على أنه " لا يوجد أثر دال إحصائيا للتفاعل بين نمط الاختبار (اختيار من متعدد، صواب وخطأ، إكمال) وجنس المستجيب (ذكور، إناث) على معامل صعوبة الفقرة ". أي أن متوسط معاملات الصعوبة لم يتغير بتغير نمط الاختبار وجنس المستجيب. و بالرغم من ذلك يلاحظ الباحث أن هناك تباين في معاملات صعوبة الفقرات على مستوى أنماط الاختبارات فلم تكن أسهل مفردات اختبار الاختيار من متعدد هي الأسهل بالنسبة للنمطين الآخرين للاختبار، كذلك لم تكن أصعب مفردات اختبار الاختيار من متعدد هي الأصعب بالنسبة للنمطين الآخرين للاختبار ولذلك فإن الترتيب النسبي للمفردات من حيث معاملات الصعوبة على مستوى الأنماط الثلاث للمفردات لم يكن متساوي، وعلى نفس المنوال فقد اختلف الترتيب النسبي للمفردات وفقا للجنس فلم تكن أسهل المفردات لدى الذكور هي الأسهل لدى الإناث في كل نمط من أنماط الاختبار وكذلك لم تكن أصعب المفردات لدى الذكور هي الأصعب لدى الإناث في كل نمط من أنماط الاختبار.

-وتتفق نتائج الدراسة الحالية مع دراسة فريسبي وسويني (1982) والتي توصلت إلى عدم توفر دلالة حاسمة إزاء صعوبة أي من نوعي الاختبار على الأخر، وعند سؤال الطلبة عن درجة صعوبة الاختبارين أجاب (47%) منهم بأن الاختبار من نوع صح-خطأ أصعب و(22%) أجاب أن الاختبار من نوع الاختيار من متعدد هو الأصعب وأن (31%) أجاب بأن للاختبارين نفس الدرجة من الصعوبة. وتتفق أيضا مع نتائج دراسة بيلر وجافاني (2000) والتي توصلت لأن الفروق بين الجنسين لا تكفي لتفسير أداء الطلاب في مادة الرياضيات. ومع دراسة سيمنز وجلوبني (2001) والتي توصلت عدم وجود فروق بين الجنسين في الصعوبة في ثلاثة أنماط من نوع الاختيار من متعدد.

-وتختلف نتائج الدراسة الحالية مع دراسة الكحلوت (2002) والتي توصلت لأن متوسط معاملات الصعوبة لاختيار من متعدد أكبر من متوسط معاملات الصعوبة لاختبار التكميل، ودراسة أمريزق (1992) التي توصلت لوجود فروق ذات دلالة إحصائية بين متوسط الصعوبة بين اختبار الاختيار المتعدد و الصح والخطأ المتعدد لصالح اختبار الاختيار المتعدد ووجود فرق ذات دلالة إحصائية بين متوسط الصعوبة بين اختبار و الصح والخطأ المتعدد و الصح والخطأ لصالح اختبار الصح والخطأ. ودراسة دبك (1998) والتي توصلت لوجود فروق ذات دلالة إحصائية في

معاملات الصعوبة تعزى لشكل الفقرات بين النموذجين ذي الفقرات من نوع (نمط) التكميل وذي الأربعة بدائل لصالح نموذج التكميل، وبين النموذجين ذي الفقرات من نوع(نمط) التكميل وذي الثلاث بدائل لصالح نموج التكميل. ودراسة أندرسون وَ هايرز (1991) والتي توصلت إلى أن اختبار التكميل أصعب من اختبار الاختيار من متعدد. ودراسة أريكسون (1988) والتي توصلت لاختبار الصواب والخطأ المتعدد له أعلى متوسط صعوبة ثم اختبار الاختيار من متعدد. ودراسة اوسترهوف وَ جلاسناب (1972) والتي توصلت لأن أسئلة الاختيار من المتعدد أسهل من أسئلة الصواب و الخطأ. ودراسة ويلسون وَ زانج (1998) والتي توصلت لأن الفرق بين متوسطي الذكور والإناث في المفردات ذات المتطلبات المعرفية الأعقد (الأصعب) أكبر بكثير من الفرق بين المتوسطين في المفردات ذات المتطلبات المعرفية البسيطة (الأسهل نسبيا). ودراسة كيركدسي وَ وأنتونيو (2009) والتي توصلت إلى أن أداء الطلاب كان أفضل في المفردات الصعبة في حين كان أداء الطالبات أفضل في المفردات السهلة. ودراسة بيلنسكي وَدافيسون (2001) والتي توصلت إلى أن المفردات السهلة كانت أسهل لدى الإناث منها لدى الذكور في حين أن المفردات الصعبة كانت أصعب لدى الإناث منها لدى الذكور. وكذلك مع دراسة ديمارس (2000) والتي توصلت لوجود تفاعل بين نمط الاختبار وجنس المستجيب فقد كانت درجات الطلاب أعلى من درجات الطالبات على أسئلة الاختيار من متعدد، في حين كانت درجات الطالبات أعلى من الطلاب على أسئلة الإجابات المنشأة. ودراسة أندرسون (1998) التي توصلت إلى أن الأسئلة الصعبة تبدو أصعب لدى الطالبات وكذلك الأسئلة السهلة تبدو أسهل لدى الطالبات.

ثانياً: اختبار الفرض الثاني للدراسة:

ينص فرض الدراسة الثاني على أنه " لا يوجد أثر دال إحصائيا للتفاعل بين نمط الاختبار (اختيار من متعدد، صواب وخطأ، إكمال) وجنس المستجيب (ذكور، إناث) على معامل تمييز الفقرة ".

قبل استخدام تحليل التباين قام الباحث بالتحقق من افتراضات تحليل التباين: الاعتدالية وتجانس التباين. وقد قام الباحث بالتحقق من افتراض الاعتدالية باستخدام اختبار شابيرو-ولك والذي كانت قيمته 0.062 وهي قيمة غير دالة إحصائيا عند مستوي 0.05. كما قام الباحث بالتحقق من عدم انتهاك افتراض تجانس التباين باستخدام اختبار ليفين لاختبار تساوي تباينات المجتمعات، وقد كانت قيمة اختبار ليفين غير دالة إحصائياً حيث كانت قيمة ف = 1.637 (5، 192) وهي قيمة غير دالة إحصائياً (الاحتمال = 0.152).

وللتحقق من صحة هذا الفرض قام الباحث بتحويل معاملات التمييز للفقرات إلى درجات فيشر المعيارية المناظرة والتي تقع على مقياس فئوي وهو ما يتيح حساب متوسطات معاملات التمييز وتحليل البيانات باستخدام تحليل التباين. وقد قام الباحث بعرض معاملات تمييز جميع فقرات الاختبارات الثلاثة (اختيار من متعدد، صواب وخطأ، إكمال) بعد تحويلها إلى درجات فيشر المعيارية المناظرة طبقاً لجنس المستجيب (ذكور، إناث) معاملات التمييز كما هو في ملحق رقم (2-2) ثم قام الباحث

بحساب دلالة الفروق بين متوسطات معامل التمييز بعد تحويلها إلى درجات فيشر المعيارية المناظرة وفقا لنمط الاختبار والجنس والتفاعل بينهما في جدول (8).

يتضح من معاملات التمييز كما هو في ملحق (2-2) تباين معاملات التمييز للفقرات، حيث تراوحت معاملات التمييز للإناث في نمط أسئلة الاختيار من متعدد بعد تحويلها لدرجات فيشر بين 0,15 للمفردة 6 إلى 0,49 للمفردة 20 بمتوسط 0,33 وانحراف معياري 0,11 في حين تراوحت معاملات التمييز للذكور في نمط أسئلة الاختيار من متعدد بعد تحويلها لدرجات فيشر بين 0,19 للمفردة 1 إلى 0,67 للمفردة 29 بمتوسط 0,41 وانحراف معياري 0,13. كما تراوحت معاملات التمييز للإناث في نمط أسئلة الصواب والخطأ بعد تحويلها لدرجات فيشر بين 0,15 للمفردات 2، 12، 15، 29، 30، 31 إلى 0,42 للمفردة 32 بمتوسط 0,22 وانحراف معياري 0,09 في حين تراوحت معاملات التمييز للذكور في نمط أسئلة الصواب والخطأ بعد تحويلها لدرجات فيشر بين 0,15 للمفردات7، 9، 13، 16، 26، 27، 29 إلى 0,40 للمفردة 11 بمتوسط 0,21 وانحراف معياري 0,07. وأخيراً تراوحت معاملات التمييز للإناث في نمط أسئلة الإكمال بعد تحويلها لدرجات فيشر بين 0,21 للمفردة 7 إلى 0,80 للمفردة 26 بمتوسط 0,52 وانحراف معياري 0,16 في حين تراوحت معاملات التمييز للذكور في نمط أسئلة الإكمال بعد تحويلها لدرجات فيشر بين 0,22 للمفردة 13 إلى 0,88 للمفردة 5 بمتوسط 0,55 وانحراف معياري 0.18.

و من معاملات التمييز في ملحق رقم (2-2) يلاحظ الباحث أنه بالرغم من تقارب متوسطات معاملات التمييز للفقرات بعد تحويلها لدرجات فيشر وقيم انحرافاتها المعيارية بين الجنسين وبين أنماط الاختبارات الثلاث إلا أن هناك تباين في معاملات تمييز الفقرات على مستوى أنماط الاختبارات فلم تكن أكبر مفردات اختبار الاختيار من متعدد تمييزاً هي الأكبر تمييزاً بالنسبة للنمطين الآخرين للاختبار، كذلك لم تكن أقل مفردات اختبار الاختيار من متعدد تمييزاً هي الأقل تمييزاً بالنسبة للنمطين الآخرين للاختبار ولذلك فإن الترتيب النسبي للمفردات من حيث معاملات التمييز على مستوي الأنماط الثلاث للمفردات لم يكن متساوي. وعلى نفس المنوال فقد اختلف الترتيب النسبي للمفردات وفقا للجنس فلم تكن أقل المفردات تمييزاً لدى الذكور هي الأقل تمييزاً لدى الإناث في كل نمط من أنماط الاختبار وكذلك لم تكن أكبر المفردات تمييزاً لدى الذكور هي الأكبر تمييزاً لدى الإناث في كل نمط من أنماط الاختبار.

ملحق رقم (2-2) :

جدول معاملات التمييز لمفردات الاختبارات الثلاث (اختيار من متعدد، صواب وخطأ، إكمال) وفقا لمتغير الجنس للعينة

رقم المفردة	الاختيار من متعدد		الصواب والخطأ		الإكمال	
	ذكور	إناث	ذكور	إناث	ذكور	إناث
1	0.19	0.15	0.16	0.17	0.34	0.33
2	0.63	0.28	0.40	0.15	0.58	0.69
3	0.45	0.33	0.22	0.41	0.36	0.46
4	0.58	0.27	0.28	0.37	0.62	0.57
5	0.51	0.33	0.16	0.16	0.88	0.57
6	0.42	0.15	0.16	0.28	0.51	0.42
7	0.28	0.23	0.15	0.31	0.55	0.21
8	0.32	0.16	0.19	0.34	0.33	0.65
9	0.35	0.38	0.15	0.15	0.79	0.43
10	0.44	0.46	0.20	0.19	0.55	0.38
11	0.66	0.41	0.17	0.32	0.24	0.34
12	0.35	0.20	0.16	0.15	0.45	0.22
13	0.48	0.41	0.15	0.16	0.22	0.29
14	0.35	0.49	0.20	0.22	0.40	0.50
15	0.38	0.40	0.29	0.15	0.69	0.66
16	0.31	0.42	0.15	0.21	0.37	0.49
17	0.31	0.41	0.40	0.38	0.64	0.74
18	0.36	0.39	0.20	0.21	0.58	0.63
19	0.61	0.45	0.15	0.16	0.45	0.61
20	0.38	0.49	0.29	0.29	0.43	0.68
21	0.22	0.22	0.25	0.31	0.36	0.42
22	0.38	0.17	0.16	0.16	0.56	0.50
23	0.48	0.25	0.34	0.16	0.55	0.49
24	0.60	0.39	0.21	0.16	0.64	0.76
25	0.20	0.48	0.16	0.19	0.49	0.56
26	0.45	0.35	0.15	0.27	0.82	0.80
27	0.31	0.34	0.15	0.16	0.75	0.68
28	0.20	0.32	0.17	0.19	0.49	0.30
29	0.67	0.19	0.15	0.15	0.87	0.64
30	0.37	0.37	0.18	0.15	0.38	0.48
31	0.38	0.40	0.20	0.15	0.66	0.68
32	0.36	0.23	0.36	0.42	0.66	0.50
33	0.42	0.20	0.16	0.16	0.84	0.65
المتوسط	0.41	0.33	0.21	0.22	0.55	0.52
الانحراف المعياري	0.13	0.11	0.07	0.09	0.18	0.16
عدد الطلاب	75	75	78	77	73	68

دلالة الفروق بين المتوسطات

للتحقق من دلالة الفروق بين المتوسطات الفروق بين متوسطات معامل التمييز بعد تحويلها إلى درجات فيشر المناظرة وفقا لنمط الاختبار والجنس والتفاعل بينهما، قام الباحث باختبار تحليل التباين الثنائي لنمط الاختبار وجنس المستجيب والتفاعل بينهما وتم تلخيص نتائج تحليل التباين الثنائي في جدول (8).

جدول (8)

دلالة الفروق بين المتوسطات الفروق بين متوسطات معامل التمييز بعد تحويلها إلى درجات فيشر المناظرة وفقا لنمط الاختبار والجنس والتفاعل بينهم

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة ف	الدلالة الاحصائية	الدلالة العملية
نمط الاختبار	3.373	2	1.686	99.17	0,001	0.515
جنس المستجيب	.041	1	.041	2.497	.116	
التفاعل	.080	2	.040	2.421	.092	
الخطأ	3.181	192	.017			
المجموع	6.676	197				

يتضح من جدول (8) وجود فروق ذات دلالة إحصائية بين متوسطات معاملات التمييز ترجع لنمط الاختبارات الثلاث، فقد كانت قيمة ف المحسوبة كبيرة (99,17) وهي قيمة دالة إحصائياً عند مستوى (0.05). وتعد هذه النتيجة متوقعة إذا ما راجعنا إلى قيم متوسطات معاملات التمييز لأنماط الاختبارات الثلاثة (الاختيار من متعدد، والصواب والخطأ والإكمال) كما هو في ملحق رقم (2-2) حيث كانت قيمة متوسط معاملات التمييز لنمط الاختيار من متعدد هو 0.37 وكان متوسط معاملات التمييز لنمط الصواب والخطأ أقل بكثير وهو 0,22 وكان متوسط معاملات التمييز لنمط الإكمال هو أعلاهم وهو 0,54 وهي قيم متباينة مما أدى لدلالة الفروق بين تلك المتوسطات. كما كانت قيمة الدلالة العملية لهذا الفرق الإحصائي الدال متمثل في مؤشر مربع ايتا مرتفع القيمة وسجل 0.515 كما يتضح من جدول (8) وهي قيمة مرتفعة وفقا لتقديرات كوهين. ولتحديد اتجاهات الفروق ذات الدلالة الإحصائية فقد تم استخدام طريقة شيفيه للمقارنات البعدية لنتائج تحليل التباين الدالة إحصائياً، وجدول (9) يبين نتائج المقارنات البعدية الثنائية بين متوسطات معاملات التمييز لأنماط الاختبار الثلاثة.

جدول (9) :

نتائج المقارنات البعدية الثنائية بين متوسطات معاملات التمييز لأنماط الاختبار الثلاثة باستخدام طريقة شيفيه

المقارنة		الفرق بين المتوسطين	الخطأ المعياري	الدلالة
الاختيار من متعدد	الصواب والخطأ	0.15	0.02	0,001
الاختيار من متعدد	الإكمال	-0.17	0.02	0,001
الصواب والخطأ	الإكمال	-0.32	0.02	0,001

ويتضح من جدول (9) وجود فروق دالة إحصائياً عند مستوى (0.05) بين متوسطات معاملات التمييز للفقرات ترجع إلى اختلاف نمط الاختبار من اختيار من متعدد إلى صواب وخطأ لصالح نمط الاختيار من متعدد. ويدل ذلك على أن مفردات نمط الاختيار من متعدد كانت أكثر تمييزاً من مفردات نمط الصواب والخطأ في المتوسط. كما يتضح من جدول (9) وجود فروق دالة إحصائياً عند مستوى (0.05) بين متوسطات معاملات التمييز للفقرات ترجع إلى اختلاف نمط الاختبار من اختيار من متعدد إلى الإكمال لصالح نمط الإكمال. ويدل ذلك على أن مفردات نمط الإكمال كانت أكثر تمييزاً من مفردات نمط الاختيار من متعدد في المتوسط. ويتضح أيضاً من جدول (9) وجود فروق دالة إحصائياً عند مستوى (0.05) بين متوسطات معاملات التمييز للفقرات ترجع إلى اختلاف نمط الاختبار من الصواب والخطأ إلى الإكمال لصالح نمط الإكمال. ويدل ذلك على أن مفردات نمط الإكمال كانت أكثر تمييزاً من مفردات نمط الصواب والخطأ في المتوسط.

وبالرجوع الى جدول (9) يلاحظ الباحث عدم وجود فروق ذات دلالة إحصائية بين متوسطات معاملات التمييز ترجع لجنس المستجيب، فقد كانت قيمة ف المحسوبة صغيرة (2,497) وهي قيمة غير دالة إحصائياً عند مستوى (0,05). وتعد هذه النتيجة متوقعة إذا ما راجعنا إلى قيم متوسطات معاملات التمييز وفقا لجنس المستجيب (ذكور، إناث) كما هو موضّح في ملحق (2-2) حيث كانت قيمة متوسط معاملات التمييز للذكور هو 0.39 وكان متوسط معاملات التمييز للإناث هو 0,36 وهي قيم متقاربة جداً مما أدى لعدم دلالة الفروق بين تلك المتوسطات.

ويتضح كذلك من جدول (9) عدم دلالة التفاعل بين نمط الاختبار وجنس المستجيب فقد كانت قيمة ف المحسوبة صغيرة جداً (2,421) وهي قيمة غير دالة إحصائياً عند مستوى (0,05). وتعد هذه النتيجة متوقعة إذا ما راجعنا إلى قيم متوسطات معاملات التمييز وفقا لجنس المستجيب (ذكور، إناث) لدى كل نمط من أنماط الاختبار كما هو في ملحق رقم (2 -2) حيث كانت قيمة متوسط معاملات التمييز للذكور في نمط الاختيار من متعدد هو 0.41 وكان متوسط معاملات التمييز للإناث في نفس النمط هو 0,33 وهي قيم متقاربة ، كذلك كانت قيمة متوسط معاملات التمييز للذكور في نمط الصواب والخطأ هو 0.21 وكان متوسط معاملات التمييز للإناث في نفس النمط هو 0,22 وهي قيم متقاربة جداً، كذلك كانت قيمة

متوسط معاملات التمييز للذكور في نمط الإكمال هو 0.55 وكان متوسط معاملات التمييز للإناث في نفس النمط هو 0,52 وهي قيم متقاربة جداً أيضاً مما أدى لعدم دلالة الفروق بين تلك المتوسطات. وبذلك يمكن القول بتحقق فرض الدراسة الثاني، حيث جاءت نتائج التفاعل غير دالة.

مناقشة نتائج اختبار الفرض الثاني:

توصلت نتائج الدراسة إلى تحقق فرض الدراسة الثاني والذي ينص علي أنه " لا يوجد أثر دال إحصائياً للتفاعل بين نمط الاختبار (اختيار من متعدد، صواب وخطأ، إكمال) وجنس المستجيب (ذكور، إناث) على معامل تمييز الفقرة ". أي أن متوسط معاملات التمييز قد تغير بتغير نمط الاختبار ولكنه لم يتغير بتغير النوع (جنس المستجيب). وقد لاحظ الباحث أن هناك تباين في معاملات تمييز الفقرات على مستوي أنماط الاختبارات حيث وجود فروق دالة إحصائياً بين متوسطات معاملات التمييز للفقرات ترجع إلى اختلاف نمط الاختبار من اختيار من متعدد إلى صواب وخطأ لصالح نمط الاختيار من متعدد، كما توجد فروق دالة إحصائياً بين متوسطات معاملات التمييز للفقرات ترجع إلى اختلاف نمط الاختبار من اختيار من متعدد إلى الإكمال لصالح نمط الإكمال وأيضاً وجود فروق دالة إحصائياً بين متوسطات معاملات التمييز للفقرات ترجع إلى اختلاف نمط الاختبار من الصواب والخطأ إلى الإكمال لصالح نمط الإكمال. وقد اختلف الترتيب النسبي للمفردات وفقا للجنس فلم تكن أكثر المفردات تمييزا لدى الذكور هي الأكثر تمييزا لدى الإناث في كل نمط من أنماط الاختبار وكذلك لم تكن أقل المفردات تمييزا لدى الذكور هي الأقل تمييزا لدى الإناث في كل نمط من أنماط الاختبار.

- وتتفق نتائج الدراسة الحالية مع دراسة فريسبي وسويني (1982) والتي توصلت إلى عدم توفر دلالة حاسمة إزاء تمييز أي من نوعي الاختبار على الآخر. وتتفق أيضا مع نتائج دراسة بيلر وجافاني (2000) والتي توصلت لأن الفروق بين الجنسين لا تكفي لتفسير أداء الطلاب في مادة الرياضيات. ومع دراسة سيمنز وَ جلوبني (2001) والتي توصلت لعدم وجود فروق بين الجنسين في التمييز في ثلاثة أنماط من نوع الاختيار من متعدد.

- وتختلف نتائج الدراسة الحالية مع دراسة الكحلوت (2002) والتي توصلت لأن متوسط معاملات التمييز لاختيار من متعدد أكبر من متوسط معاملات التمييز لاختبار التكميل، ودراسة أمريزق (1992) التي توصلت لوجود فروق ذات دلالة إحصائية بين متوسط التمييز بين اختبار الاختيار المتعدد و الصح والخطأ المتعدد لصالح اختبار الاختيار المتعدد ووجود فرق ذات دلالة إحصائية بين متوسط التمييز بين اختبار والصح والخطأ المتعدد و الصح والخطأ لصالح اختبار الصح والخطأ. ودراسة دبك (1998) والتي توصلت لوجود فروق ذات دلالة إحصائية في معاملات التمييز تعزى لشكل الفقرات بين النموذجين ذي الفقرات من نوع (نمط) التكميل وذي الأربعة بدائل لصالح نموذج التكميل، وبين النموذجين ذي الفقرات من نوع(نمط) التكميل

وذي الثلاث بدائل لصالح نموج التكميل. ودراسة أريكسون (1988) والتي توصلت لاختبار الصواب والخطأ المتعدد له أعلى متوسط تمييز ثم اختبار الاختيار من متعدد. ودراسة ويلسون وَ زانج (1998) والتي توصلت لأن الفرق بين متوسطي الذكور والإناث في المفردات ذات المتطلبات المعرفية الأعقد (الأصعب) أكبر بكثير من الفرق بين المتوسطين في المفردات ذات المتطلبات المعرفية البسيطة (الأسهل نسبيا). ودراسة بيلنسكي وَ دافيسون (2001) والتي توصلت لفروق دالة بين الإناث والذكور في معاملات التمييز. وكذلك مع دراسة ديمارس (2000) والتي توصلت لوجود تفاعل بين نمط الاختبار وجنس المستجيب فقد كانت درجات الطلاب أعلى من درجات الطالبات على أسئلة الاختيار من متعدد، في حين كانت درجات الطالبات أعلى من الطلاب على أسئلة الإجابات المنشأة.

ثالثاً: اختبار الفرض الثالث للدراسة:

ينص فرض الدراسة الثالث على أنه " لا يوجد أثر دال إحصائيا للتفاعل بين نمط الاختبار (اختيار من متعدد، صواب وخطأ، إكمال) وجنس المستجيب (ذكور، إناث) على معامل ثبات الفقرات ".

وللتحقق من صحة هذا الفرض قام الباحث بحساب معاملات الثبات بطريقة ألفا كرونباخ لكل نمط من أنماط الاختبار لدى كل من الذكور والإناث. ثم قام الباحث بحساب دلالة الفروق بين متوسطات كل زوج من معاملات الثبات للذكور والإناث (متغير الجنس) لكل نمط من أنماط الاختبار الثلاثة وللأنماط الثلاثة (نمط الاختبار) لدى كل من الذكور والإناث باستخدام الإحصائي (M) وعرضت النتائج في جدولي 10، 11.

جدول (10) :
دلالة الفروق بين معاملات الثبات بطريقة ألفا للذكور والإناث لاختبارات الدراسة الثلاث (اختيار من متعدد، صواب وخطأ، إكمال)

الاختبار	ذكور	إناث	الاحصاءM	الدلالة
اختيار من متعدد	0,868	0,810	2,35	0.125
صواب وخطأ	0,584	0,634	0.30	0.584
إكمال	0,917	0,914	0.02	0.888

يتضح من جدول (10) عدم وجود فروق دالة إحصائياً عند مستوى (0.05) بين الذكور والإناث في معاملات الثبات بطريقة ألفا في أي من أنماط الاختبار الثلاث، حيث تشير النتائج المعروضة في جدول (10) إلى تقارب معاملات الثبات بطريقة ألفا بين الذكور والإناث في نمط الاختيار من متعدد حيث كانت قيمة ثبات درجات الذكور 0.868 وكانت قيمة ثبات درجات الإناث 0.810 وهي قيم متقاربة. كما كانت قيم معاملات الثبات بطريقة ألفا بين الذكور والإناث في نمط الصواب والخطأ متقاربة حيث كانت قيمة ثبات درجات الذكور 0.584 وكانت قيمة ثبات درجات الإناث 0.634 وهي قيم متقاربة. وأيضا كانت قيم معاملات الثبات بطريقة ألفا بين الذكور والإناث

في نمط الإكمال متقاربة حيث كانت قيمة ثبات درجات الذكور 0.917 وكانت قيمة ثبات درجات الإناث 0.914 وهي قيم متقاربة مما أدى لعدم دلالة الفروق بين تلك المعاملات. وبذلك فإنه يمكن القول بعدم وجود فروق دالة إحصائياً بين معاملات الثبات ترجع لجنس المستجيب.

جدول (11) :

دلالة الفروق بين معاملات الثبات بطريقة ألفا لاختبارات الدراسة الثلاث (اختيار من متعدد، صواب وخطأ، إكمال) لكل من الذكور والإناث

الجنس	اختيار من متعدد	صواب وخطأ	الاحصاءMi	الدلالة	اختيار من متعدد	إكمال	الاحصاءMi	الدلالة	صواب وخطأ	إكمال	الاحصاءMi	الدلالة
ذكور	0,868	0,584	22.55	أقل من 0.001	0,868	0,917	3.61	0.057	0,584	0,917	40,89	أقل من 0.001
إناث	0,810	0,634	7,70	0.006	0,810	0,914	10.78	0.001	0,634	0,914	34.80	أقل من 0.001
الكلية	0.840	0.620	11.22	0.001	0.840	0.915	12.52	أقل من 0.001	0.620	0.915	36.52	أقل من 0.001

يتضح من جدول (11) وجود فروق دالة إحصائياً بين معاملات الثبات بطريقة ألفا بين نمط الاختيار من متعدد والصواب والخطأ لصالح نمط الاختيار من متعدد لدى كل من الذكور والإناث، حيث كانت قيمة ثبات درجات نمط الاختيار من متعدد للذكور 0.868 وكانت قيمة ثبات درجات نمط الصواب والخطأ للذكور 0.584 وكان هذا الفرق دال إحصائياً عند مستوى (0.05)، مما يشير لأن ثبات درجات الذكور في نمط الاختيار من متعدد كان أفضل من ثباتها في نمط الصواب والخطأ. الأمر نفسه تكرر لدى الإناث حيث كانت قيمة ثبات درجات نمط الاختيار من متعدد للإناث 0.810 وكانت قيمة ثبات درجات نمط الصواب والخطأ للإناث 0.634 وكان هذا الفرق دال إحصائياً عند مستوى (0.05)، مما يشير لأن ثبات درجات الإناث في نمط الاختيار من متعدد كان أفضل من ثباتها في نمط الصواب والخطأ.

كما يتضح من جدول (11) وجود فروق دالة إحصائياً بين معاملات الثبات بطريقة ألفا بين نمط الاختيار من متعدد والإكمال لصالح نمط الإكمال لدى الإناث وعدم وجود نفس هذا الفرق بين معاملات الثبات بطريقة ألفا بين نمط الاختيار من متعدد والإكمال لدى الذكور، حيث كانت قيمة ثبات درجات نمط الاختيار من متعدد للذكور 0.868 وكانت قيمة ثبات درجات نمط الإكمال للذكور 0.917 وكان الفرق غير دال إحصائياً عند مستوى (0.05) ولكنه غير دال مما يشير لأن ثبات درجات الذكور في نمط الاختيار من متعدد والإكمال لا يختلفان إحصائياً عند

مستوى (0.05). الأمر اختلف لدى الإناث حيث كانت قيمة ثبات درجات نمط الاختيار من متعدد للإناث 0.810 وكانت قيمة ثبات درجات نمط الإكمال للإناث 0.914 وكان هذا الفرق دال إحصائياً عند مستوى (0.05)، مما يشير لأن ثبات درجات الإناث في نمط الإكمال كان أفضل من ثباتها في نمط الاختيار من متعدد. وتشير تلك النتائج لدلالة التفاعل بين نمط الاختبار (اختيار من متعدد، إكمال) والجنس (ذكور، إناث) عند مستوى (0.05) لصالح الإناث في نمط الإكمال.

كذلك يتضح من جدول (11) وجود فروق دالة إحصائياً بين معاملات الثبات بطريقة ألفا بين نمط الصواب والخطأ والإكمال لصالح نمط الإكمال لدى كل من الذكور والإناث، حيث كانت قيمة ثبات درجات نمط الصواب والخطأ للذكور 0.584 وكانت قيمة ثبات درجات نمط الإكمال للذكور 0.917 وكان هذا الفرق دال إحصائياً عند مستوى (0.05)، مما يشير لأن ثبات درجات الذكور في نمط الإكمال كان أفضل من ثباتها في نمط الصواب والخطأ. الأمر نفسه تكرر لدى الإناث حيث كانت قيمة ثبات درجات نمط الصواب والخطأ للإناث 0.634 وكانت قيمة ثبات درجات نمط الإكمال للإناث 0.914 وكان هذا الفرق دال إحصائياً عند مستوى (0.05)، مما يشير لأن ثبات درجات الإناث في نمط الإكمال كان أفضل من ثباتها في نمط الصواب والخطأ.

وتشير النتائج المعروضة في جدولي 10، 11 لرفض الفرض الثالث جزئيا حيث جاءت نتائج الفروق بين الجنسين في معاملات الثبات بطريقة ألفا غير دالة إحصائيا بينما كانت نتائج الفروق الراجعة لنمط الاختبار (اختيار من متعدد، صواب وخطأ، إكمال) في معاملات الثبات بطريقة ألفا دالة إحصائياً، وأخيراً كانت نتائج الفروق الراجعة للتفاعل بين نمط الاختبار (اختيار من متعدد، إكمال) والجنس (ذكور، إناث) في معاملات الثبات بطريقة ألفا دالة إحصائياً لصالح الإناث في نمط الإكمال.

مناقشة نتائج اختبار الفرض الثالث:

توصلت نتائج الدراسة إلى وجود فروق دالة إحصائياً بين معاملات الثبات بطريقة ألفا بين نمط الاختيار من متعدد والصواب والخطأ لصالح نمط الاختيار من متعدد لدى كل من الذكور والإناث، حيث كان ثبات درجات الذكور في نمط الاختيار من متعدد أفضل من ثباتها في نمط الصواب والخطأ. الأمر نفسه تكرر لدى الإناث حيث كانت قيمة ثبات درجات الإناث في نمط الاختيار من متعدد أفضل من ثباتها في نمط الصواب والخطأ. ووجود فروق دالة إحصائياً بين معاملات الثبات بطريقة ألفا بين نمط الاختيار من متعدد والإكمال لصالح نمط الإكمال لدى الإناث وعدم وجود نفس هذا الفرق بين معاملات الثبات بطريقة ألفا بين نمط الاختيار من متعدد والإكمال لدى الذكور، حيث كان ثبات درجات الذكور في نمط الاختيار من متعدد والإكمال لا يختلفان إحصائيا. والأمر اختلف لدى الإناث حيث كان ثبات درجات الإناث في نمط الإكمال أفضل من ثباتها في نمط الاختيار من متعدد. وتشير تلك النتائج لدلالة التفاعل بين نمط الاختبار (اختيار من متعدد، إكمال) والجنس (ذكور، إناث) لصالح الإناث في نمط الاختيار من متعدد. كما توصلت الدراسة لوجود فروق

دالة إحصائياً بين معاملات الثبات بطريقة ألفا بين نمط الصواب والخطأ والإكمال لصالح نمط الإكمال لدى كل من الذكور والإناث، حيث كانت ثبات درجات الذكور في نمط الإكمال أفضل من ثباتها في نمط الصواب والخطأ. الأمر نفسه تكرر لدى الإناث حيث كان ثبات درجات الإناث في نمط الإكمال أفضل من ثباتها في نمط الصواب والخطأ. وهذا يعني رفض الفرض الثالث جزئيا حيث جاءت نتائج الفروق بين الجنسين في معاملات الثبات بطريقة ألفا غير دالة إحصائياً بينما كانت نتائج الفروق الراجعة لنمط الاختبار (اختيار من متعدد، صواب وخطأ، إكمال) في معاملات الثبات بطريقة ألفا دالة إحصائياً، وأخيرا كانت نتائج الفروق الراجعة للتفاعل بين نمط الاختبار (اختيار من متعدد، إكمال) والجنس (ذكور، إناث) في معاملات الثبات بطريقة ألفا دالة إحصائياً لصالح الإناث في نمط الإكمال.

وتتفق نتائج الدراسة الحالية مع دراسة الكحلوت (2002) والتي أظهرت نتائجها الدراسة أن معاملي ثبات اختبار التكميل المحسوبين بطريقة الفا كرونباخ وطريقة إعادة الاختبار أكبر من معاملي ثبات اختيار الاختيار من متعدد. ودراسة أمريزق (1992) والتي توصلت لفروق ذات دلالات إحصائية بين معاملي ثبات الاختبار لصالح اختبار الاختيار من متعدد تم اختبار الصح والخطأ واختبار الصح والخطأ المتعدد. ودراسة العلي (1989) والتي توصلت لوجود فروق بين معاملات ثبات الاختبار ذات دلالة إحصائية في مستوى الأداء الكلي، ومستوى التذكر و كانت هذه الفروق لصالح الإجابة القصيرة في مستوى الأداء الكلي، و لصالح الاختبار من نوع المزيج بين النوعين معا، في مستوى التذكر. ودراسة فيرسبي وايبل (1972) والتي توصلت إلي أن هناك فروق ذات دلالة إحصائية في معامل الثبات (KR20) لفقرات اختبار الصح والخطأ وفقرات اختبار الاختيار من متعدد لصالح أسئلة الاختيار من متعدد. ودراسة فريسبي وسويني (1982) والتي توصلت لأن معامل ثبات فقرات الاختبار من نوع الصح والخطأ أعلى منه في الاختبار من نوع الاختيار من متعدد. ودراسة أندرسون وهايرز (1991) والتي توصلت إلى أن معامل ثبات اختبار التكميل أعلى من معامل ثبات اختبار الاختيار من متعدد. و دراسة أوسترهوف وجلاسناب (1972) والتي توصلت إلى أن أسئلة الصواب و الخطأ (خاصة الجمل الصحيحة) أقل ثبات من أسئلة الاختيار من المتعدد.

وتختلف نتائج الدراسة الحالية مع دراسة اريكسون (1988) والتي أشارت نتائجها إلى أن معامل ثبات ألفا كان متقاربا جميع الاختبارات (الصواب والخطأ المتعدد والإجابة القصيرة والاختيار من متعدد). ودراسة إبيل روبرت (1971) والتي توصلت لأن أسئلة الصواب والخطأ لها ثبات مثل أسئلة الاختيار من المتعدد، وانه لا يوجد اختلاف بينهما في القياس. ودراسة تاسديمر (2010) والتي لم تظهر نتائجها وجود فروق ذات دلالة إحصائية بين متوسطات اختبار الاختيار من متعدد والصواب والخطأ على أساس الجنس. وتختلف أيضا مع نتائج دراسة بيلر وجافاني (2000) والتي توصلت لأن الفروق بين الجنسين لا تكفي لتفسير أداء الطلاب في مادة

الرياضيات. وتختلف مع نتائج دراسة أندرسون (1998) والتي توصلت لعدم وجود فروق دالة إحصائياً في ثبات وصدق الاختبار تعزى للجنس.

توصيات الدراسة:

في ضوء النتائج التي توصلت لها الدراسة الحالية توصي الدراسة بالآتي:

- عند اختيار نمط الاختبار يجب أن يضع المعلم في اعتباره أولا مناسبة نمط الاختبار لهدف ومحتوي المادة التعليمية، حيث لم تثبت الدراسة الحالية أثر لنمط الاختبار أو جنس المستجيب أو التفاعل بينهما على صعوبة أو تمييز الفقرة.

- عند إعداد الفقرات الاختبارية يفضل أن يلجأ المعلم لفقرات الإكمال قبل أن يفكر في إعداد الفقرات بنمط الصواب والخطأ يليه نمط الاختيار من متعدد كلما أمكن ذلك حيث توصلت الدراسة لوجود أثر دال لنمط السؤال على الثبات لصالح أسئلة الإكمال وأسئلة الاختيار من متعدد، خاصة عندما يكون المستجيبين من الإناث.

مقترحات الدراسة:

- إجراء دراسة مماثلة للدراسة الحالية لدراسة أثر التفاعل بين نمط فقرات الاختبار

  وجنس المستجيب على بعض الخصائص السيكومترية للفقرات وثبات الاختبار في مواد أخرى كالعلوم والدراسات الاجتماعية.

- إجراء دراسة مماثلة للدراسة الحالية لدراسة أثر التفاعل بين نمط فقرات الاختبار وجنس المستجيب على بعض الخصائص السيكومترية للفقرات وثبات الاختبار باستخدام نظرية الاستجابة للمفردة.

- إجراء دراسة مماثلة للدراسة الحالية لدراسة أثر التفاعل بين نمط فقرات الاختبار وجنس المستجيب على بعض الخصائص السيكومترية للفقرات وثبات الاختبار باستخدام أنماط أخرى من الأسئلة المفتوحة النهاية وذات الاستجابة المنشأة.

- إجراء دراسة لفحص أثر التفاعل بين نمط فقرات الاختبار وجنس المستجيب على بعض الخصائص السيكومترية للفقرات وصدق الاختبار في مقابل ثبات الاختبار الذي اهتمت به الدراسة الحالية.

- إجراء دراسة لفحص أثر التفاعل بين نمط فقرات الاختبار والصف الدراسي أو المرحلة الدراسية على بعض الخصائص السيكومترية للفقرات وثبات الاختبار في مقابل وجنس المستجيب الذي اهتمت به الدراسة الحالية.

المراجع

المراجع العربية:

أبولبدة، سبع محمد(2008م). مبادي القياس النفسي والتقييم التربوي. عمّان: دار الفكر.

الزاملي،علي وَ الصارمي،عبدالله وَ كاظم، علي(2009م). مفاهيم وتطبيقات في التقويم والقياس التربوي. عمّان: دار حنين للنشر والتوزيع .

الصمادي ،عبدالله ؛الدرابيع ،ماهر(2004م). القياس النفسي والتربوي بين النظرية والتطبيق. الكرك: مركز يزيد للخدمات الطلابية.

الطريري، عبدالرحمن بن سليمان(1997م). القياس النفسي والتربوي.الرياض: مكتبة الرشيد للنشر والتوزيع .

النبهان،موسى (2004م).أساسيات القياس في العلوم السلوكية. عمّان: دارالشروق للنشر والتوزيع.

العلي ،محمد(1989م) . المقارنة بين أثر اختبار من متعدد ، واختبار الإجابة القصيرة ، واختبار المزيج من النوعين من حيث الخصائص السيكومترية للاختبار (الصدق و الثبات) ، رسالة ماجستير غير منشورة ، جامعة اليرموك ،إربد، الأردن.

أبو عالم ، رجاء (1993).علم النفس التربوي. الكويت : دار القلم

امريزق ، خالد(1992م) .المقارنة بين اختبار الاختبار من متعدد ، واختبار الصح و الخطأ ،من حيث الخصائص السيكومترية وشكل منحنى التوزيع للأداء على الاختبار ،رسالة ماجستير غير منشورة ، جامعة الاردنية ،اربد، الأردن.

دبك ،أحمد (1998م). أثر شكل الفقرات و عدد البدائل في الخصائص السيكومترية للاختبار وفقراته، رسالة ماجستير غير منشورة ، جامعة اليرموك ،إربد، الأردن .

سليمان، شاهر خالد أحمد (2001م). أثر بنية فقرات الصح والخطأ والتكميل على الخصائص السيكومترية للاختبار وفقراته، رسالة دكتوراه غير منشورة، جامعة الخرطوم.

عبدالهادي،نبيل(2001م).القياس والتقويم التروي واستخدامه في التدريس الصفي (ط2). عمّان : دار وائل للطباعة و النشر.

علام ، صلاح الدين محمود (2002م). القياس والتقويم التربوي والنفسي أساسياته وتطبيقاته وتوجهاته المعاصرة. القاهرة: دار الفكر العربي.

عطوي، جودت عزت(2007م). أساليب البحث العلمي (ط2). عمّان: دار الثقافة للنشر والتوزيع.

ملحم،سامي(2002م). القياس والتقويم في التربية وعلم النفس. عمّان: دار المسيرة للنشر والتوزيع والطباعة.

مراد،صلاح أحمد(2000م). الأساليب الإحصائية في العلوم النفسية و التربوية والاجتماعية . القاهرة : مكتبة الانجلو المصرية .

المراجع العربية مترجمة:

Abu Libdeh, S. M., (2008 AD). *Principles of psychological measurement and educational evaluation*. Amman: Dar Al-Fikr.

Al-Zamili, A., & Al-Sarmi, A., & Kazem, A., (2009). *Concepts and applications in educational assessment and measurement*. Amman: Dar Hanin for Publishing and Distribution.

Al-Smadi, A., & Al-Darabe', M., (2004 AD). *Psychological and educational measurement between theory and practice*. Karak: Yazid Center for Student Services.

Al-Tariri, A. S., (1997 AD). *Psychological and educational measurement. Riyadh*: Al-Rasheed Library for Publishing and Distribution.

Al-Nabhan, M., (2004 AD). The *basics of measurement in the behavioral sciences*. Amman: Dar Al-Shorouk for Publishing and Distribution.

Al-Ali, M., (1989 AD). *Comparison between the effect of a multiple test, a short answer test, and a mixture test of the two types in terms of the psychometric properties of the test (honesty and reliability)*, an unpublished master's thesis, Yarmouk University, Irbid, Jordan.

Abu Alem, R., (1993). *Educational Psychology*. Kuwait: Dar Al-Qalam

Amreziq, K., (1992). *Comparison between the multiple test test and the true and false test, in terms of psychometric properties and the shape of the distribution curve for performance on the test*, unpublished master's thesis, University of Jordan, Irbid, Jordan.

Dabek, A., (1998 AD). *The effect of paragraph shape and number of alternatives on the psychometric characteristics of the test and its paragraphs*, unpublished master's thesis, Yarmouk University, Irbid, Jordan.

Suleiman, S. K., (2001 AD). *The effect of the structure of true, false and complementary clauses on the psychometric properties of the test and its clauses*, unpublished Ph.D. thesis, University of Khartoum.

Abd al-Hadi, N., (2001 AD). *Prudent measurement and evaluation and its use in classroom teaching (2nd ed.)*. Amman: Wael House for Printing and Publishing.

Allam, S. El., (2002). *Educational and psychological measurement and evaluation: its basics, applications, and contemporary directives*. Cairo: Arab Thought House.

Atwi, J. E., (2007). *Scientific Research Methods* (I 2). Amman: House of Culture for Publishing and Distribution.

Melhem, S., (2002). *Measurement and evaluation in education and psychology.* Amman: Dar Al Masirah for publishing, distribution and printing.

Murad, S. A., (2000 AD). *Statistical methods in psychological, educational and social sciences*. Cairo: Anglo-Egyptian Library.

المراجع الأجنبية:

Anderson, P. (1998). Respondent gender by response format interaction: Test validity and reliability issues. *Paper Presented at the Annual Meeting of the American Educational Research Association*, San Diego, CA, April 13-17.

Anderson,P. & Hyers,A.(1991).Quantitative Comparisons of Difficulty, Discrimination and Reliability of Machine-Scored Completion Items and Tests in Contrast with Statistics from Comparable Multiple Choice Questions.(ED349 319)

Bielinski, J. & Davison, M.L. (2001). A Sex Difference by Item Difficulty Interaction in Multiple-choice Mathematics Items Administered to National Probability Samples. *Journal of Educational Measurement, 38(1), 51-77.*

Erickson,I.L.(1988). The effect of item type and test format on level of difficulty and mean item discrimination.Dissertation Abstract international 499(6),14-37.

Ediger,M.(2000).The Principal and Evaluation of Student Achievement .*Journal of Instructional Psychology,* 27(3), 155-161.

Ebel,R.L.(1971).The Comparative Effectiveness of true-false Multiple-Choice Achievement Tests. (ED 050148)

Frisbie,D.A. & Ebel,R.L.(1972).Comparison of reliabilities and validities of true-false and Multiple-Choice Tests.*Journal of Educational Measurement*, 10(4), 297-304.

Gadalla, T. (1999). Multiple-choice versus Constructed-response Tests in the Assessment of Mathematics Computation Skills.*Paper Presented at the Annual Meeting of the American Educational Research Association, Monteral*, CA, April 19-23.

HAKSTIAN,A.RALPH & E.WHALEN,THOMAS.(1976) A k-sample significance test for independent alphacoefficients.PSYCHOMETRIKA-VOL.41,NO.2

Kyriakides, L. & Antoniou, P. (2009). Gender differences in mathematics achievement: an investigation ofgender differences by item difficulty interactions. *Educational Research and Evaluation,* 15 (3), 223-242.

Oosterhof,A.C.&Glasnapp,D.R.(1972).Comparative Reliabilities of Multiple-Choice and true-false formats.*Journal of Experimental Education*, 42 (1),62-64.

Oosterhof,A.C.& Coats, P.K. (1981). Comparison of Difficulties and Reliabilities of Math-Completion and Multiple-Choice Item Formats. *Paper Presented at the Annual Meeting of the American Educational Research Association,* Los Angeles, CA, April 13-17.

Simens, S. & Globni, N. (2001). Gender differences in three multiple choice formats in Mathematics testing: Effects of item difficulty and discrimination. *Mathematics Education*, 12 (3), 248-259.

Tasdemir, M.(2010). A Comparison of Multiple-Choice Tests and True-False Tests Used in Evaluating Student Progress. *Journal of Instructional Psychology,* 37 (3), 258-266.

Wilson, L.D. & Zhang, L. (1998). A cognitive analysis of gender differences on constructed-response and multiple choice assessments in Mathematics. *Paper Presented at the Annual Meeting of the American Educational Research Association*, San Diego, CA, April 13-17.