

الفصل الخامس نتائج الدراسة وتفسيرها

- اختبار صحة الفروض.
- مناقشة نتائج الفروض.
- التطبيقات التربوية للدراسة.
- توصيات وبحوث مقترحة.

الفصل الخامس

نتائج الدراسة وتفسيرها

في ضوء مشكلة البحث والفروض التي قامت عليها والأساليب الإحصائية المستخدمة تعرض الباحثة في هذا الفصل لنتائج الدراسة وتفسيرها على النحو التالي:

أولاً: اختبار الفروض:

اختبار الفرض الأول ونصه كالتالي: "يختلف تقدير قيمة معامل ثبات الاختبار باختلاف طول الاختبار (عدد مفرداته)".

وللتحقق من صحة هذا الفرض تم حساب الإحصاءات الخاصة بالاختبار وكذلك الإحصاءات الخاصة بمفردات الاختبار، وتضمنت هذه الإحصاءات المتوسطات الحسابية والانحرافات المعيارية وقيم معامل الثبات وذلك لكل من صور الاختبار (أ، د) حيث تألفت الصورة (أ) من (٧٠) مفردة والصورة (د) من (٤٠) مفردة وحتى يسهل اختبار هذا الفرض إحصائياً تم إعادة صياغته في صورة إحصائية كالتالي:

الفرض الصفري: "لا توجد فروق دالة إحصائية بين قيم معاملات الثبات عند اختلاف طول الاختبار (٤٠ مفردة، ٧٠ مفردة)".

وقد استخدمت الباحثة اختبار الدلالة الإحصائية للفروق بين معاملات الارتباط (٠)، على اعتبار أن معامل الثبات هو في حقيقته معامل ارتباط بين المقياس ونفسه.

وبتطبيق معادلة إحصائية الاختبار السابق على بيانات الدراسة الحالية، تم التوصل إلى النتائج التي يعرضها الجدول التالي:

جدول (١٠)

قيمة (z) لدلالة الفروق بين معاملات الثبات المحسوبة لصورتى الاختبار (أ، د) والتي تختلف فيما بينها في عدد المفردات (٤٠ مفردة، ٧٠ مفردة)

| صور الاختبار | المتوسط الحسابي | الانحراف المعياري | معامل الثبات (ألفا) | قيمة (z) المحسوبة | الدلالة الإحصائية |
|-------------------------------|-----------------|-------------------|---------------------|-------------------|---------------------------|
| (أ) تتألف من مفردة (٧٠) | ٣٣,٧٥ | ١٤,٦ | ٠,٩ | ١,٩٥ | دالة عند مستوى دلالة ٠,٠٥ |
| (د) تتألف من مفردة (٤٠) | ١٩,١٥ | ٤,٠٣ | ٠,٦٧ | | |

يتضح من الجدول السابق وجود فروق دالة إحصائية عند مستوى الدلالة (٠,٠٥) بين قيمتي معامل الثبات لصورتى الاختبار (أ، د)، وكانت الفروق نتيجة لزيادة طول الاختبار من (٤٠) مفردة إلى (٧٠) مفردة، وقد بلغت القيمة الجدولية (١,٦٥) عند مستوى الدلالة (٠,٠٥)، والاختبار من طرف واحد، وبذلك يمكن رفض الفرض الصفري وقبول الفرض البديل (فرض البحث) عند مستوى (٠,٠٥).

اختبار الفرض الثاني ونصه كالتالي:

" يختلف تقدير قيمة معامل الثبات باختلاف موقع المفردة من الاختبار (ترتيب عرض المفردات داخل الاختبار)".

ولاختبار الفرض، تم حساب الإحصاءات الخاصة بصورتى الاختبار (أ، ب)، وتضمنت المتوسطات والانحرافات المعيارية وقيم معاملات الثبات (ألفا)، وتم تطبيق اختبار الدلالة الإحصائية للفروق بين معاملات الارتباط، بعد إعادة صياغة فرض البحث في الصورة الإحصائية التالية:

الفرض الصفري: "لا توجد فروق دالة إحصائية بين قيم معاملات الثبات نتيجة تغير موقع المفردة من الاختبار (ترتيب عرض مفردات الاختبار)".

$$(٢٢ = ١٢)$$

وقد أسفر اختبار الفرض إحصائياً عن النتائج الموضحة بالجدول التالي:

جدول (١١)

قيمة (Z) لدلالة الفروق بين قيم معاملات الثبات المحسوبة لصورتَي الاختبار (أ، ب) والتي تختلف فيما بينها في ترتيب عرض المفردات (من السهل إلى الصعب - من الصعب إلى السهل)

| صور الاختبار | ترتيب المفردات | المتوسط الحسابي | الانحراف المعياري | معامل الثبات (ألفا) | قيمة (Z) المحسوبة | الدلالة الإحصائية |
|--------------------------|--------------------|-----------------|-------------------|---------------------|-------------------|-------------------------------|
| (أ) عدد المفردات (٧٠) | من السهل إلى الصعب | ٣٣,٧٥ | ١٤,٦ | ٠,٩ | ١,٧١ | دالة عند مستوى الدلالة (٠,٠٥) |
| (ب) عدد المفردات (٧٠) | من الصعب إلى السهل | ٢١,٦ | ٧,٣ | ٠,٧٥ | ... | |

يتضح من الجدول السابق وجود فروق دالة إحصائياً بين قيم معاملات الثبات المحسوبة لصورتَي الاختبار (أ، ب) نتيجة لاختلاف ترتيب عرض مفردات الاختبار من حيث مستوى الصعوبة (من السهل إلى الصعب والعكس) وكانت هذه الفروق نتيجة لترتيب المفردات من السهل إلى الصعب. وهذا يعني قبول الفرض البديل (فرض البحث) ورفض الفرض الصفري عند مستوى الدلالة (٠,٠٥).

اختبار الفرض الثالث ونصه كالتالي:

"يختلف تقدير قيمة معامل ثبات الاختبار باختلاف عدد بدائل مفردات الاختبار من متعدد (٣ بدائل - ٤ بدائل)".

ولاختبار الفرض، تم حساب الإحصاءات الخاصة بصورتَي الاختبار (أ، ج) وتضمنت المتوسطات الحسابية والانحرافات المعيارية وقيم معاملات الثبات، حيث تألفت الصورة الأولى من (٧٠) مفردة بأربعة بدائل لكل مفردة، والصورة الثانية من (٧٠) مفردة أيضاً بثلاثة بدائل لكل مفردة، وقد تم حساب قوة وفاعلية المشتتات قبل إجراء التطبيق النهائي لكلا صورتين^(٩)، وتم حذف المشتتات الأقل فاعلية من جميع مفردات الاختبار بهدف إعداد

الصورة (ج) وتمهيداً لاختبار الفرض إحصائياً تم إعادة صياغته في الصورة الإحصائية التالية:

الفرض الصفري: "لا توجد فروق دالة إحصائية بين قيم معاملات الثبات نتيجة لاختلاف عدد بدائل مفردات الاختبار (٣ بدائل - ٤ بدائل للمفردة)".

$$(r = 1)$$

وقد نتج عن تطبيق اختبار الدلالة الإحصائية للفروق بين معاملات الارتباط على بيانات صورتي الاختبار (أ، ج)، النتائج الموضحة بالجدول التالي:

جدول (١٢)

قيمة (Z) لدلالة الفروق بين قيم معاملات الثبات المحسوبة لصورتي الاختبار (أ، ج) والتي تختلف في عدد بدائل مفرداتها (٣-٤) بدائل للمفردة

| صورة الاختبار | عدد بدائل المفردة | المتوسط الحسابي | الانحراف المعياري | معامل الثبات (ألفا) | قيمة (Z) المحسوبة | الدلالة الإحصائية |
|--------------------------|-------------------|-----------------|-------------------|---------------------|-------------------|-------------------------------|
| (أ) عدد المفردات (٧٠) | أربعة بدائل | ٣٣,٧٥ | ١٤,٦ | ٠,٩ | ٢,٠٣ | دالة عند مستوى الدلالة (٠,٠٥) |
| (ج) عدد المفردات (٧٠) | ثلاثة بدائل | ٢١ | ٧,٢٨ | ٠,٧ | | |

يتضح من الجدول السابق وجود فروق دالة إحصائية بين قيم معاملات الثبات مع التغيير في عدد بدائل مفردات الاختبار لكل من الصورتين (أ، ج) وهذه الفروق نتيجة استخدام أربعة بدائل للمفردة، وهذا يعني قبول الفرض البديل (فرض البحث) ورفض الفرض الصفري عند مستوى الدلالة (٠,٠٥).

اختبار الفرض الرابع ونصه كالتالي:

"يختلف تقدير قيمة معامل ثبات الاختبار باختلاف حجم عينة المختبرين".

ولاختبار الفرض، قامت الباحثة باشتقاق عينات عشوائية مختلفة الحجم باستخدام الحاسب الآلي، وتم حساب الإحصاءات الخاصة بالاختبار وكذلك الإحصاءات الخاصة بالمفردات وذلك لكل عينة من هذه العينات، وتضمنت المتوسطات الحسابية والانحرافات

المعيارية ومعاملات الثبات، حيث تألفت العينة الأولى من (٣٠) طالب والعينة الثانية من (٥٠) طالب، وتمهيداً لاختبار الفرض إحصائياً تم إعادة صياغته في الصورة التالية:

الفرض الصفري: "لا توجد فروق دالة إحصائية بين قيم معاملات الثبات نتيجة اختلاف حجم عينة المختبرين".

وقد نتج عن تطبيق اختبار الدلالة الإحصائية للفروق بين معاملات الارتباط النتائج الموضحة بالجدول التالي:

جدول (١٣)

قيمة (Z) لدلالة الفروق بين معاملات الارتباط (معاملات الثبات) المحسوبة من درجات عينتين من الطلاب مختلفتين في عدد أفرادها (٥٠-٣٠) طالب

| الدلالة الإحصائية | (Z) المحسوبة | معامل الثبات | الانحراف المعياري | المتوسط الحسابي | حجم العينة |
|---------------------------|--------------|--------------|-------------------|-----------------|------------|
| غير دالة عند مستوى (٠,٠٥) | ١,١٩ | ٠,٨٨ | ١٤,٧ | ٣٣,٦ | ٣٠ طالب |
| | | ٠,٩٤ | ١٦,٦ | ٣٢,٦ | ٥٠ طالب |

يتضح من الجدول السابق عدم وجود فروق دالة إحصائية بين قيم معاملات الثبات، مع التغير في حجم عينة المختبرين وهذا يعني رفض الفرض البديل (فرض البحث) وقبول الفرض الصفري عند مستوى الدلالة (٠,٠٥) وتختلف هذه النتيجة مع توقع الباحثة والإطار النظري للدراسة.

ثانياً: مناقشة نتائج الفروض:

تقوم الباحثة في هذا الجزء من الدراسة بمناقشة وتفسير النتائج المستمدة من التحليل الإحصائي والاختبار الإحصائي لفروض الدراسة وذلك على النحو التالي:

مناقشة نتائج الفرض الأول:

أوضحت نتائج اختبار الفرض إحصائياً أن قيمة معامل ثبات الاختبار زادت مع زيادة عدد مفردات الاختبار بفروق دالة إحصائية عند مستوى الدلالة (٠,٠٥). وتتفق هذه النتيجة مع التراث السيكولوجي والإطار النظري للدراسة، وما أكدت عليه دراسات كل من Merrill

1993, Wagner 1990, Ebel, 1969, Fitzpatrick 2001, Leon & et al., 1995
وترى الباحثة أن نتائج هذه الدراسات جميعها تؤكد على التأثير الإيجابي لزيادة عدد مفردات
أداة القياس على دقة نتائج القياس وبالتالي على قيمة معامل الثبات.

فقد أوضحت دراسة Ebel, 1969 والتي هدفت إلى تقدير معامل الثبات المتوقع كدالة
في طول الاختبار وعدد البدائل أو الاختيارات لكل مفردة، وجود فروق دالة إحصائياً في قيم
معاملات الثبات نتيجة لزيادة عدد مفردات الاختبار حيث ارتفع معامل الثبات من (٠,٤٨) إلى
(٠,٨٦) عندما زاد طول الاختبار من (٥٠) إلى (٢٠٠) مفردة.

وقد اتفقت أيضاً نتائج دراسة Merrill, 1993 مع نتائج الدراسة الحالية فيما يتعلق
بدراسة أثر طول الاختبار على قيمة معامل ثباته، حيث هدفت الدراسة إلى اختبار العلاقة بين
خصائص المقياس السيكمترية (الصدق والثبات) وطول المقياس بالنسبة لمقاييس الشخصية،
وقد أظهرت نتائج الدراسة أن معاملات ثبات المقياس بصفة عامة ارتبطت إيجابياً بطول
المقياس (عدد مفرداته).

وتتفق النتيجة التي توصلت إليها الباحثة الحالية مع ما أكدت عليه دراسة Leon
& et al., 1995 والتي هدفت إلى اختبار أثر عدد مفردات المقياس (طول المقياس) على
مقدار التشبث داخل المجموعات (تباين الخطأ) وحجم الأثر بين المجموعات وكذلك ثبات أداة
القياس. فقد أوضحت نتائج الدراسة أنه بزيادة عدد مفردات المقاييس الفرعية، اتجه التشبث
داخل المجموعات (تباين الخطأ) إلى الانخفاض، وارتفعت قيم معامل الثبات والقوة الإحصائية
للاختبار، فقد زادت قيمة معامل الثبات من (٠,٥) بالنسبة للمقياس الفرعي الذي يتألف من
(٤) مفردات إلى (٠,٨٨) للمقياس الكلي الذي يتضمن (٢٤) مفردة.

وقد لاحظت الباحثة اتفاق هذه النتيجة مع ما أكد عليه كل من رجاء أبو علام ١٩٨٧،
٢٠٠١، صلاح الدين محمود علام ٢٠٠١، سعد عبدالرحمن ١٩٨٣، فقد اتفقوا على أن
المفردات التي يتألف منها الاختبار تمثل عينة من المفردات التي درسها الطلاب وأنه بزيادة
عدد هذه المفردات يزداد تمثيل هذه العينة للمجتمع الأصلي (الموضوعات التي تهدف لقياسها)
وبالتالي يزداد انتشار درجات الطلاب وترتفع قيمة معامل الثبات، وأن المفهوم الأساسي الذي
تقدمه لنا معادلة سبيرمان براون هو أن العامل الأول الذي يؤثر على قيمة معامل ثبات
الاختبار هو طوله (عدد مفرداته) والمنطق وراء ذلك هو أن العدد الأكبر من المفردات يؤدي
دائماً إلى قياس عينة أكبر من السلوك مما يعطي فرصة أكبر لانتشار الدرجات مما يؤدي إلى
انخفاض قيمة تباين الخطأ وبالتالي تقترب الدرجات الملاحظة للطلاب من درجاتهم الحقيقية

وهذا بلا شك يؤثر إيجابياً على قيمة معامل الثبات. وبالتالي فإن هذه النتيجة تتفق مع الدراسات السابقة والإطار النظري للدراسة وما أكدته المتخصصون في هذا المجال.

مناقشة نتائج الفرض الثاني:

أوضحت نتائج اختبار الفرض إحصائياً وجود فروق دالة إحصائياً في قيم معاملات الثبات مع تغير أسلوب عرض وترتيب مفردات الاختبار (من السهل إلى الصعب - من الصعب إلى السهل) وكانت هذه الفروق دالة نتيجة لترتيب مفردات الاختبار من السهل إلى الصعب عند مستوى الدلالة (٠,٠٥).

وتتفق النتيجة التي تم التوصل إليها مع الإطار النظري للدراسة ومع ما أكدت عليه بعض الدراسات التي اهتمت باختبار أثر تسلسل وترتيب مفردات الاختبار على خصائصه السيكومترية ومنها معامل الثبات وكذلك على خصائص المفردة (معاملات الصعوبة - معاملات التمييز)، ولكنها تختلف مع نتائج بعض الدراسات الأخرى في هذا الشأن.

فقد اتفقت النتيجة التي توصلت إليها الباحثة مع بعض نتائج دراسة Fleugher & et al., 1968، والتي هدفت إلى اختبار أثر إعادة ترتيب مفردات الاختبار على الأداء الاختباري والإحصاءات المرتبطة بالمفردة (الصعوبة - الارتباط الجزئي)، حيث توصلت نتائج الدراسة إلى وجود فروق معنوية جديرة بالملاحظة في مستويات الصعوبة نتيجة إعادة ترتيب مفردات الاختبار، فقد ارتفعت تقديرات معاملات الصعوبة بالنسبة للترتيب المعياري (من السهل إلى الصعب) خاصة بالنسبة للمفردات التي ظهرت في نهاية الاختبار. وهذه النتيجة تتفق مع ما توصلت إليه الباحثة، حيث إن ارتفاع تقديرات معاملات الصعوبة يؤدي بلا شك إلى انخفاض تباين درجات الطلاب وبالتالي تنخفض قيم معاملات التمييز مما يؤثر سلباً على قيمة معامل الثبات.

اتفقت أيضاً النتيجة الحالية مع ما أسفرت عنه نتائج دراسة Cromack & Sax, 1966 فقد هدفت الدراسة إلى اختبار أثر ترتيب وتقديم مفردات الاختبار إلى التلاميذ في أربعة صور تجريبية مختلفة على الأداء الاختباري للطلاب. حيث توصلت الدراسة إلى ارتفاع في متوسط درجات الطلاب على صورة الاختبار (E) والتي تم ترتيب مفرداتها من السهل إلى الصعب بالمقارنة بالصورة (H) والتي رتب مفرداتها من الصعب إلى السهل بينما لم توجد فروق دالة في متوسطات صور الاختبار (M, R).

وترى الباحثة أن الارتفاع في متوسط درجات الطلاب له تأثيره الإيجابي على الخصائص السيكومترية للاختبار ومنها معامل الثبات، لأن التحسن في مستوى أداء الطلاب يعني اعتدالية صعوبة الأسئلة وهذا يؤثر إيجابياً في تباين درجات الطلاب وبالتالي تتحسن قيمة معامل الثبات.

وقد اتفقت مع النتيجة السابقة ما أوضحته دراسة Hambleton & Ross 1974، والتي هدفت إلى دراسة أثر ترتيب وتغير موقع المفردات ضمن الاختبار على الأداء الاختباري للطلاب، وتوصلت الدراسة إلى وجود فروق دالة إحصائية في متوسطات درجات الطلاب على صور الاختبار التي اختلفت في أسلوب ترتيب مفرداتها نتيجة لترتيب المفردات من السهل إلى الصعب، وتفسر الباحثة هذه النتيجة بأن ترتيب المفردات ضمن الاختبار من السهل إلى الصعب يزيد من دافعية الطلاب أثناء الإجابة على الاختبار، وهذا بدوره يؤدي إلى التحسن في مستوى أدائهم الاختباري، وبالتالي تتحسن قيم معاملات صعوبة المفردات ومعاملات تمييزها مما يؤثر إيجابياً على قيمة معامل الثبات.

وقد اختلفت النتيجة التي توصلت إليها الباحثة مع ما أسفرت عنه نتائج بعض الدراسات التي اهتمت باختبار أثر متغير موقع المفردة من الاختبار (ترتيب المفردات داخل الاختبار) على خصائص مفردات الاختبار (الصعوبة - قوة التمييز)، وقيمة معامل الثبات. فقد توصلت دراسة Brenner, 1964 إلى عدم وجود فروق دالة إحصائية بين صور الاختبار المختلفة الترتيب في قيم معاملات الصعوبة ومعاملات الثبات، وترى الباحثة أن اختلاف هذه النتيجة مع النتيجة الحالية، ربما يرجع إلى اختلاف طبيعة الاختبارات التي تم تطبيقها والاختلاف في عدد المفردات التي يتضمنها الاختبار، حيث تألف كل اختبار من (٤٠) مفردة بالمقارنة بالاختبار الحالي (٧٠) مفردة، بالإضافة إلى اختلاف الأساليب التي اعتمد عليها الباحث في ترتيب المفردات، حيث تم استخدام الأسلوب العشوائي في عملية الترتيب.

اختلفت أيضاً النتيجة الحالية مع نتائج دراسة كل من Smouse & Marso, 1970 و Munz, 1968 حيث هدفت كل دراسة إلى اختبار أثر موقع المفردة من الاختبار وكذلك متغير قلق الاختبار على الأداء الاختباري للطلاب، وأسفرت النتائج عن عدم وجود تأثير دال إحصائياً لمتغير موقع وترتيب المفردات داخل الاختبار على الأداء الاختباري للطلاب، وترى الباحثة أن هذه النتيجة قد ترجع إلى طبيعة المتغيرات التي تم دراستها وكذلك طبيعة العينات التي تم تطبيق الاختبارات عليها وأحجام هذه العينات، بالإضافة إلى اختلاف طبيعة

الاختبارات التي تم تطبيقها، حيث كانت هذه الاختبارات تقيس القدرات اللغوية بينما الاختبار الحالي هو اختبار تحصيلي.

اتفقت معهم أيضاً نتائج دراسات William & Monk, 1970، Husck, 1978، حيث أوضحت نتائج هذه الدراسات عدم وجود فروق دالة إحصائياً بين قيم معاملات الصعوبة ومعاملات الثبات مع التغير في أساليب ترتيب مفردات الاختبار.

وترى الباحثة أن هذه النتائج لا يجب أن تعمم على جميع المواقف الاختبارية، لأن اتفاق النتيجة الحالية مع نتائج بعض الدراسات واختلافها مع البعض الآخر، يعني الحاجة إلى المزيد من الدراسات التي تهتم باختبار أثر التغير في موقع المفردات من الاختبار وترتيبها من حيث مستوى صعوبتها على خصائص الاختبار السيكومترية، وأن النتيجة الحالية وهي وجود تأثير دال لترتيب المفردات ضمن الاختبار من السهل إلى الصعب على قيمة معامل الثبات، هي نتيجة منطقية لأن توقع الطلاب لترتيب المفردات بهذا الأسلوب يزيد من دافعيتهم للإجابة ولتكملة الاختبار، وهذا بدوره يحسن من أدائهم الاختباري وبالتالي تتحسن الخصائص السيكومترية للمفردات (معاملات الصعوبة - معاملات التمييز)، وهذا بلا شك يؤثر إيجابياً على قيمة معامل ثبات الاختبار.

مناقشة نتائج الفرض الثالث:

أظهرت نتائج اختبار الفرض إحصائياً وجود فروق دالة في قيم معاملات الثبات مع تغير عدد البدائل لمفردات الاختبار (٣-٤) بدائل للمفردة، عند مستوى الدلالة (٠,٠٥)، نتيجة لاستخدام أربعة بدائل للمفردة.

وقد اتفقت هذه النتيجة مع نتائج بعض الدراسات التي اهتمت باختبار أثر عدد بدائل مفردات الاختبار من متعدد على الخصائص السيكومترية للاختبار (الصدق والثبات) وكذلك خصائص المفردات (معاملات الصعوبة - معاملات التمييز)، وتناقضت أيضاً مع بعضها الآخر.

فقد اتفقت النتيجة الحالية مع نتائج دراسات كل من Green & et al., 1982، Marshal, Budescu & Nevo, 1985 حيث أكدت هذه الدراسات على فاعلية استخدام أربعة بدائل للمفردة في تحسين الخصائص السيكومترية للمفردة (معاملات الصعوبة - معاملات التمييز) وبالتالي ارتفاع قيمة معامل ثبات الاختبار، وتفسر الباحثة هذه النتيجة بأن الانخفاض في عدد بدائل المفردة عن أربعة بدائل قد يزيد من فرص التخمين في الإجابة أمام بعض الطلاب، وبصفة خاصة في حالة انخفاض جاذبية هذه البدائل (قوة المشتتات) وهذا بدوره يؤدي إلى انخفاض الثبات في درجات الطلاب فتتخفف قيم معاملات التمييز لهذه

المفردات، وبالتالي تتخفف قيمة معامل الثبات، وهنا يكون من الأفضل زيادة عدد البدائل لكل مفردة عن ثلاثة بدائل حتى نقلل من فرص التخمين في الإجابة وبالتالي تتحسن قيم معاملات التمييز ومعاملات الثبات.

ومن جانب آخر تناقضت النتيجة التي توصلت إليها الباحثة مع نتائج دراسات كل من Trevison & et al., Trevison, 1990, Straton & Catts, 1980, Sax, 1980, James Bruno, Rogers & Harley, 1999, Hancock & et al., 1993, 1991, 1995.

فقد توصلت دراسة Sax, 1980 والتي هدفت إلى المقارنة بين معاملات الثبات والصدق الخاصة بأسئلة الاختيار من متعدد ذات البديلين وأسئلة الصواب والخطأ إلى عدم وجود فروق دالة إحصائية في قيم معاملات الثبات بين كلا النوعين، وترى الباحثة أن اختلاف هذه النتيجة مع النتيجة الحالية، قد يرجع إلى محدودية أعداد البدائل التي تمت المقارنة بينها وكذلك تحويل أسئلة الاختيار من متعدد إلى أسئلة صواب وخطأ، بينما اهتمت الدراسة الحالية بأسئلة الاختيار من متعدد ذات البدائل (٤/٣)، وعلى العكس من ذلك أوضحت نتائج دراسة Hancock & et al., 1993 وجود فروق دالة إحصائية بين قيم معاملات الثبات لأسئلة الاختيار من متعدد ذات البديلين بالمقارنة بأسئلة الصواب والخطأ وترى الباحثة أن هذه النتيجة قد ترجع إلى الصعوبة النسبية لأسئلة الاختيار من متعدد بالمقارنة بأسئلة الصواب والخطأ وخاصة عند ارتفاع قوة المشتتات وكذلك انخفاض نسبة التخمين في الإجابة بالمقارنة بأسئلة الصواب والخطأ.

اختلفت أيضاً مع النتيجة الحالية نتائج دراسات كل من Trevison & et al., 1991, Trevison, 1990 فقد هدفت كلا الدراستين إلى تحديد العدد الأمثل لبدائل المفردة، الذي يمكنه أن يرفع من ثبات وصدق الاختبار وفقاً لقدرات التلاميذ التحصيلية وتم استخدام أعداد مختلفة من البدائل (٥ / ٤ / ٣) عبر المجموعات متفاوتة القدرة (منخفضة - متوسطة - مرتفعة) وقد أسفرت النتائج عن ارتفاع في قيمة معاملات تمييز المفردات بانخفاض عدد بدائل المفردة إلى ثلاث بدائل وخاصة بالنسبة للمجموعة مرتفعة القدرة وقد أكدت النتائج أيضاً على أن العدد الأمثل من البدائل عبر المجموعات متفاوتة القدرة ثلاثة بدائل.

وقد يرجع اختلاف هذه النتائج مع النتيجة الحالية إلى اختلاف خصائص العينات وطبيعة المادة الدراسية وأعداد البدائل التي تمت المقارنة بينها (٥ / ٤ / ٣) بدائل للمفردة بالإضافة إلى تقسيم أفراد العينة إلى ثلاثة مجموعات متفاوتة القدرة التحصيلية (مرتفعة -

متوسطة - منخفضة) وذلك بالمقارنة بعينة الدراسة الحالية والتي لم يتبع معها مثل هذا التصنيف وتفسر الباحثة أيضاً ارتفاع معاملات تمييز المفردات بانخفاض عدد البدائل إلى ثلاث وخاصة مع المجموعة مرتفعة القدرة بأن هؤلاء الطلاب تتخضع لديهم نسبة التخمين في الإجابة بالمقارنة بالمجموعة منخفضة القدرة التحصيلية وانخفاض نسبة التخمين يحسن من قيم معاملات تمييز المفردات وبالتالي ينعكس أثر ذلك إيجابياً على قيمة معامل ثبات الاختبار.

وترى الباحثة أن إعداد اختبار الاختيار من متعدد بأربعة بدائل للمفردة مع مراعاة جودة وفاعلية هذه البدائل يعطى الفرصة للطلاب للتفكير جيداً في اختيار الإجابة الصحيحة مما يقلل من درجة تخمينهم للإجابة الصحيحة وبالتالي تتحسن نسبة التباين في درجاتهم مما يؤثر إيجابياً على قيم معاملات تمييز المفردات ومعامل ثبات الاختبار

وبصفة عامة فإن أفضل استخدام عدد معين من البدائل لمفردات اختبار الاختيار من متعدد والذي يمكنه أن يسهم في تحسين الخصائص السيكومترية للاختبار تحكمه مجموعة من الضوابط وتتمثل في:

١- قوة وفاعلية المشتتات (البدائل الخاطئة): فإذا تميزت مفردات الاختبار بارتفاع قوة مشتتاتها فإن انخفاض عدد بدائل المفردات إلى اثنين أو ثلاثة قد لا يؤثر سلباً على قيمة معامل الثبات، بل العكس قد يكون أفضل من استخدام أربعة أو خمسة بدائل للمفردة مع انخفاض جاذبية المشتتات، فقد أوضحت نتائج دراسة فتحى الزيات ١٩٨٩، وجود تأثير دال لفاعلية المشتتات في أسئلة الاختيار من متعدد على خصائص المفردات (الصعوبة - التمييز) وخصائص الاختبار السيكومترية (الصدق والثبات).

٢- المرحلة التعليمية: فالتلاميذ في المراحل التعليمية الأولى يفضل أن نطبق عليهم اختبارات الاختيار من متعدد بعدد منخفض من البدائل حتى لا يزيد من صعوبة الموقف الاختباري عليهم، وذلك بالمقارنة بالطلاب في مراحل التعليم الأعلى حيث إن زيادة عدد بدائل المفردة إلى أربعة بدائل مع زيادة فاعلية المشتتات لهذه المفردات، يمكنه أن يسهم في إظهار الفروق الفردية في مستوياتهم التحصيلية وبالتالي تحسين الخصائص السيكومترية لأدوات القياس التي تطبق عليهم.

مناقشة نتائج الفرض الرابع:

أوضحت نتائج اختبار الفرض إحصائياً عدم وجود فروق دالة إحصائياً بين قيم معاملات الثبات مع تغير حجم عينة المختبرين عند مستوى الدلالة (٠,٠٥). وتختلف هذه النتيجة مع توقع الباحثة والإطار النظري للدراسة وما أكدت عليه نتائج الدراسات السابقة التي اهتمت باختبار أثر حجم العينة على دقة نتائج القياس والقيمة العددية لمعامل الثبات.

فقد اختلفت النتيجة الحالية مع نتائج دراسات كل من Feldt & Robert, 1998، Fan & Yin, 2001 حيث هدفت الدراسة الأولى إلى توجيه الباحثين في مواقف القياس المختلفة إلى كيفية تحديد حجم العينة المناسب والذي يجب استخدامه للمقارنة بين قيمتين من قيم معاملات ألفا للثبات وأوضحت النتائج أنه من غير المرغوب استخدام أحجام من العينات أقل من (٣٠) فرداً لأن ذلك يضعف من إمكانية اكتشاف فروق دالة بين قيم معاملات ألفا للثبات، وأن الحد الأدنى لمستوى قوة الاختبار الإحصائي للمقارنة بين قيم معاملات ألفا للثبات هو (٠,٥) وإذا رغب الباحث في زيادة هذا المستوى يجب عليه زيادة حجم العينة المستخدمة لأن قوة الاختبار دالة في حجم العينة، واستهدفت الدراسة الثانية اختبار مدى تأثير خصائص عينة المختبرين على قيمة معامل الثبات وتباين درجات الخطأ، واعتمد الباحثان على استخدام أحجام كبيرة من العينات حيث تألفت المجموعة الأولى من (٥٠٠٠) طالب والثانية من (١٠٠٠٠) طالب وأسفرت النتائج عن أن أخطاء القياس اتجهت إلى الانخفاض مع الزيادة في حجم العينة ومستوى قدراتها، وذلك بالمقارنة بالمجموعة منخفضة الحجم والقدرة التحصيلية، وقد يرجع الاختلاف بين هذه النتائج والنتيجة التي توصلت إليها الدراسة الحالية إلى اختلاف خصائص العينات وأحجامها حيث كانت المقارنة في هذه الدراسات تتم بين أحجام كبيرة من العينات بالإضافة إلى اختلاف أدوات القياس المستخدمة بالمقارنة بالدراسة الحالية.

وترى الباحثة أن عدم وجود فروق دالة إحصائياً بين قيم معاملات الثبات (ألفا) بالرغم من زيادة حجم عينة الطلاب من (٣٠) إلى (٥٠) طالب، قد يرجع إلى محدودية أحجام العينات التي تمت المقارنة بينها، حيث إن تعدد متغيرات الدراسة وتقسيم العينة الكلية التي تتألف من (١١٠) طالب وطالبة إلى أربعة مجموعات وفقاً لهذه المتغيرات لم يعطى الفرصة للباحثة لزيادة أحجام العينات التي يتم المقارنة بينها بدرجة كافية، وخاصة وأن الاختبار هو اختبار تحصيلي وموجه لفئة معينة من الطلاب، مما يتطلب إجراء المزيد من الدراسات التي تهتم باختبار أثر حجم العينة على دقة وثبات القياس.

وقد ترجع أيضاً عدم دلالة الفروق بين قيم معاملات الثبات المحسوبة من درجات عينة الطلاب مع زيادة عدد أفرادها إلى تقارب المستوى التحصيلي لهؤلاء الطلاب في المادة التي يغطيها الاختبار حتى مع زيادة أعدادهم مما أدى إلى انخفاض نسبة الفروق الفردية أو نسبة التباين الملاحظ في درجاتهم، مما نتج عنه وجود فروق طفيفة بين قيم معاملات الثبات المحسوبة من درجات هؤلاء الطلاب، فعلى الرغم من زيادة قيمة معامل الثبات من (٠,٨٨) إلى (٠,٩٤) عند زيادة حجم عينة الطلاب من (٣٠) إلى (٥٠) طالب، إلا أن هذه الزيادة لم تكن ذات دلالة إحصائية.

وبصفة عامة تتفق الباحثة مع ما أوضحه فاروق عبد الفتاح موسى (١٩٩٠) حيث يؤكد على أن زيادة عدد أفراد العينة بدرجة كافية تحقق الاستقرار في نتائج القياس، لأن ذلك يضمن للباحث أو القائم بالقياس تمثيل العناصر غير الشائعة في مجتمع الدراسة، مما يؤثر إيجابياً على تباين درجات أفراد العينة وبالتالي على دقة وثبات أداة القياس.

مناقشة عامة لنتائج الدراسة:

في ضوء ما عرضته الباحثة في الإطار النظري وما تم التوصل إليه من نتائج يمكن استخلاص مجموعة من النقاط التي تمثل خصائص أو سمات للاختبار ذو الثبات المرتفع، وهذه النقاط يمكن أن تغيد الباحثين في مواقف القياس المختلفة في الحصول على أفضل تقدير ممكن لثبات درجات المقاييس الخاصة بهم، وتتضمن ما يلي:

١- الاختبار مرتفع الثبات يجب أن تكون مفرداته من نوع الاختيار من متعدد، حيث إن هذا النوع من المفردات يسمح بتغطية جزء كبير من محتوى مادة الاختبار، وبالتالي يؤثر إيجابياً على صدق وثبات أداة القياس، ولكن مع التركيز على جودة وفاعلية بدائل الإجابة المستخدمة مع تحليل فاعلية المشتتات الخاصة بها.

٢- الاختبار مرتفع الثبات يجب أن يكون عدد بدائل الإجابة لمفرداته في حدود أربعة بدائل، حيث إن هذا العدد يقلل من فرص التخمين أمام بعض الطلاب وخاصة إذا تميزت هذه البدائل بقوة المشتتات الخاصة بها، مما ينعكس أثره إيجابياً على قيم معاملات تمييز المفردات وبالتالي على ثبات درجات الاختبار.

٣- الاختبار مرتفع الثبات إذا كان اختبار تحصيلي من نوع الاختيار من متعدد فيجب أن يتضمن عدداً كافياً ومناسباً من المفردات، كما أوضحت الدراسة الحالية، لأن انخفاض

عدد مفردات الاختبار يؤثر سلباً على صدق وثبات أداة القياس، حيث إن العدد المنخفض من المفردات لا يساعد على تمثيل المحتوى تمثيل جيد وبالتالي ينخفض صدق الأداة، وكذلك ينخفض تباين درجات الطلاب مما يؤثر سلباً على ثبات هذه الدرجات.

٤- الاختبار مرتفع الثبات يجب أن يتم تطبيقه على عينة كبيرة الحجم تمثل التباين في مجتمع الدراسة تمثيلاً جيداً، حيث تساعد العينة الكبيرة على زيادة الثقة في النتائج المستمدة منها، وكذلك زيادة الثقة في دقة واستقرار هذه النتائج.

التطبيقات التربوية للدراسة:

إن النتائج المستمدة من الدراسة الحالية يمكن أن توجه عمل التربويين والقائمين بإعداد الاختبارات والمعلمين والباحثين في مجال القياس النفسي والمجال التربوي والذين يهتمون بالخصائص السيكومترية للمفردات الاختبارية (الصعوبة - التمييز - قوة المشتتات)، وفيما يلي توضيح ذلك:

١- في ضوء ما عرضته الباحثة في الإطار النظري وما توصلت إليه من نتائج توصي بضرورة الاهتمام باستخدام أسئلة الاختبار من متعدد عند إعداد الاختبارات التحصيلية، لأن هذا النوع من الاختبارات إذا راعى الباحث أو المعلم الدقة في إعداده وفي انتقاء المشتتات الجيدة لمفرداته، فإن الدرجات المستمدة منه تكون صادقة ومعبرة عن المستويات الحقيقية للطلاب مما ينعكس أثره إيجابياً على خصائص المفردات (الصعوبة - التمييز) وخصائص الاختبار (الصدق - الثبات)، ولتحقيق الموضوعية المنشودة في عملية التقويم، يجب أيضاً على المعلم والقائم بالقياس الاهتمام باستخدام الأساليب الحديثة في عملية تقييم أداء الطلاب والتي تعتمد على استخدام البورتفوليو، والتقييم الأصيل الذي يركز على توثيق أداء التلميذ وإنتاجه وكذلك قدرته على حل المشكلات..

٢- بينت الدراسة الحالية أن الزيادة في طول الاختبار التحصيلي (عدد مفرداته) لها تأثير دال إحصائياً على خصائص الاختبار السيكومترية وخاصة معامل الثبات، لذلك يجب على المعلمين والقائمين بإعداد الاختبارات التحصيلية الاهتمام بزيادة أعداد المفردات التي يتألف منها الاختبار، لأن هذه المفردات ما هي إلا عينة من المحتوى الذي درسه الطلاب وكلما زاد حجم عينة المفردات زاد تمثيلها لمحتوى المقرر وبالتالي يزداد التباين في الدرجات الملاحظة للطلاب ويقترّب التباين الحقيقي من التباين الملاحظ مما يحسن من قيمة معامل ثبات الاختبار.

ويجب أن يراعى المعلم أو القائم بإعداد الاختبار أن استخدام عدداً كافياً ومناسباً من المفردات التي تمثل المحتوى تمثيل جيد وتساعد على تحسين الخصائص السيكومترية لأداة القياس (الصدق - الثبات) يتوقف على بعض المتغيرات، والتي من أهمها زمن الاختبار ونوعية المفردات المستخدمة (استدعاء - تعرف)، فزيادة الزمن المخصص لإجابة الاختبار يتيح الفرصة أمام معد الاختبار لاستخدام عدد كافٍ وجيد من المفردات والعكس صحيح وذلك بشرط أن يكون هذا الزمن في الحدود المعقولة حتى لا يشعر الطلاب بالضيق والملل من الاختبار وبالتالي ينعكس ذلك سلبياً على درجاتهم من الاختبار. كذلك فإن نوعية المفردات التي يتألف منها الاختبار تؤثر على العدد الذي يمكن استخدامه من هذه المفردات، فأسئلة الاستدعاء ومن بينها مفردات المقال لا تعطى فرصة مناسبة لمعد الاختبار لتغطية جزء أكبر من المحتوى، لأن هذا النوع من المفردات يتطلب وقتاً أطول من الطالب في التفكير فيه والإجابة عليه، لذلك فإن الاختبارات من هذا النوع تكون عدد مفرداتها محدوداً، وذلك بالمقارنة بأسئلة التعرف، والتي من بينها مفردات الاختيار من متعدد ومفردات الصواب والخطأ، فهذا النوع من المفردات يسمح بتغطية جزء أكبر من محتوى مادة الاختبار، وبالتالي يتيح لمعد الاختبار استخدام عدد جيد من المفردات، ولكن مع التركيز على شروط وقواعد إعدادها وبصفة خاصة جودة وفاعلية البدائل المستخدمة في الإجابة حتى لا يلجأ بعض الطلاب إلى التخمين الذي يؤثر سلبياً على ثبات درجات الاختبار. والمقصود هنا بجودة البدائل أو المشتتات هو قدرتها على جعل الطالب يفكر جيداً قبل اختيار الإجابة الصحيحة.

ويجب أيضاً على معد الاختبار أن يهتم بإجراء الدراسة الأولية لهذا الاختبار والتي تفيده جيداً في معرفة زمن الإجابة المناسب لعدد المفردات التي يتألف منها الاختبار، بالإضافة إلى فائدتها الأساسية في تقويم الخصائص الإحصائية لأداة القياس (الصدق - الثبات).

٣- الاهتمام بمتغير موقع المفردة من الاختبار (ترتيب عرض المفردات ضمن الاختبار) لما له من تأثير دال إحصائياً على قيمة معامل ثبات الاختبار، وخاصة أسلوب ترتيب المفردات من السهل إلى الصعب فقد ثبت أن هذا الأسلوب له تأثير دال في زيادة دافعية الطلاب للإجابة على الاختبار، بالمقارنة بالأسلوب العكسي وهو ترتيب مفردات الاختبار من الصعب إلى السهل أو الترتيب العشوائي والذي يلجأ إليه بعض المعلمين بهدف إعداد صور مختلفة لنفس الاختبار تجنباً لمحاولات الغش من بعض الطلاب، دون إدراك أثر ذلك على دافعية الطلاب ومستوى القلق لديهم وخاصة إذا واجهتهم الأسئلة الصعبة في بداية الاختبار، مما يؤثر سلبياً على درجاتهم في الاختبار وبالتالي على ثبات هذه الدرجات. لذلك يجب أن يراعى البدء بالأسئلة السهلة يليها الأسئلة الأقل سهولة تدريجياً

حتى نزيد من دافعية الطلاب للإجابة ولتكملة الاختبار وبالتالي تتحسن درجات هؤلاء الطلاب وترتفع قيمة معامل ثبات الاختبار. وإذا أراد المعلم تجنب محاولات الغش من بعض الطلاب، فيمكنه أن يقوم بإعداد صور مختلفة لنفس الاختبار تختلف فيما بينها من حيث مواقع بدائل الإجابة وليس مواقع المفردات ذاتها، مع الأخذ في الاعتبار التساوي في توزيع البديل الصحيح على جميع مفردات الاختبار، فعلى سبيل المثال إذا كان الاختبار يتألف من (١٠٠) مفردة من نوع الاختيار من متعدد وكان عدد البدائل لكل مفردة (٤) بدائل، فيجب على القائم بإعداد الاختبار توزيع البدائل الصحيحة بالتساوي على هذه البدائل الأربعة بأسلوب عشوائي، فإذا كانت بدائل الإجابة هي (أ، ب، ج، د)، تكون (٢٥) مفردة إجابتها (أ)، (٢٥) مفردة إجابتها (ب) وهكذا، ويجب أن يتم هذا التوزيع بأسلوب عشوائي وليس بالتسلسل، حيث يلجأ بعض الطلاب إلى تخمين الإجابة الصحيحة مما يؤثر سلبياً على ثبات درجات الاختبار.

٤- مراعاة استخدام أعداد مناسبة من بدائل الإجابة لمفردات الاختيار من متعدد لما له من تأثير دال إحصائياً على قيمة معامل ثبات الاختبار فقد أوضحت نتائج الدراسة فاعلية استخدام أربعة بدائل للمفردة وأثره الإيجابي على قيمة معامل الثبات بالمقارنة باستخدام ثلاثة بدائل للمفردة، مع التركيز على حساب قوة المشتتات (البدائل الخاطئة) لجميع المفردات، حيث إن انخفاض قوة المشتتات بالرغم من زيادة عدد البدائل يؤدي إلى زيادة نسبة التخمين في الإجابة مما يؤثر سلبياً على قيم معاملات تمييز المفردات وبالتالي على قيمة معامل ثبات الاختبار.

٥- التركيز على تعديل مفردات الاختبار في ضوء قوة المشتتات لما لها من تأثير إيجابي في تحسين قيم معاملات التمييز نتيجة ارتفاع تباين المفردات وبالتالي زيادة الفروق الفردية في درجات الطلاب وتحسين ثبات هذه الدرجات.

٦- يجب على الباحثين والقائمين بإجراءات القياس عدم التركيز فقط عند تطبيق أداة القياس على حجم عينة المختبرين أو المفحوصين دون مراعاة الدقة في اختيارها وجودة تمثيلها لمجتمع الدراسة، فالعينة كبيرة الحجم التي تمثل المجتمع بصدق تعطى نتائج أكثر دقة واستقراراً حيث تضمن للقائم بالقياس تمثيل العناصر المختلفة في مجتمع الدراسة، وبالتالي تزيد الثقة في النتائج المستمدة منها.

٧- يجب على الباحثين والمعلمين والمهتمين بإعداد الاختبارات التحصيلية عدم الإقتصار فقط عند تحليل مفردات الاختبار على الطرق التقليدية التي تركز فقط في حساب قيم

معاملات الصعوبة والتمييز وفاعلية المشتتات على المجموعتين الطرفيتين (العليا والدنيا) من الطلاب وإهمال إجابات باقي الطلاب، وضرورة الاهتمام بتطبيق الأساليب الإحصائية التي تأخذ في اعتبارها إجابات جميع الطلاب، مثل حساب معامل ارتباط المفردة بالدرجة الكلية على الاختبار Corrected Item Total Correlation والذي يعتبر مؤشر جيد لقوة تمييز المفردات، وكذلك يمكن باستخدام وحدة التكرارات ببرنامج SPSS تحديد تكرارات اختيار بدائل الإجابة لجميع المفردات حتى يسهل حساب فاعلية المشتتات لهذه المفردات وهذه الإجراءات في تحليل مفردات الاختبار يمكن تنفيذها بسهولة وفي أقل وقت بمساعدة البرنامج الإحصائي SPSS، مما يعطي فرصة أكبر لتعديل المفردات غير الجيدة والتي تؤثر بلا شك على دقة نتائج القياس وعلى قيمة معامل ثبات الاختبار.

توصيات وبحوث مقترحة:

أولاً: توصيات مستخلصة من الدراسة:

- دراسة أثر عدد البدائل في أسئلة الاختيار من متعدد (٣ / ٤ / ٥) بدائل على صدق الاختبار.
- دراسة أثر ترتيب مفردات الاختبار (من السهل إلى الصعب) من الصعب إلى السهل - الترتيب العشوائي) على قيم معاملات التمييز ومعاملات الصعوبة للمفردات.
- دراسة أثر عدد المفردات في اختبار الاختيار من متعدد على صدق الاختبار.
- اختبار أثر ترتيب المفردات من حيث مستوى الصعوبة على مستوى قلق الاختبار والدرجات التحصيلية للطلاب.
- دراسة ثبات أسئلة الاختيار من متعدد كدالة في عدد بدائل المفردة وقدرات التلاميذ التحصيلية.
- العلاقة بين اختلاف مستوى القدرة التحصيلية (منخفضة - متوسطة - مرتفعة) واختلاف حجم عينة المختبرين على الثبات.

- العلاقة بين اختلاف مستوى القدرة التحصيلية واختلاف ترتيب المفردات من حيث مستوى الصعوبة على ثبات الاختبار.

ثانياً: توصيات بدراسات وبحوث:

- دراسة أثر قوة المشتتات في أسئلة الاختيار من متعدد على خصائص المفردة (الصعوبة - التمييز) والخصائص السيكومترية للاختبار (الصدق - الثبات).
- أثر ظروف تطبيق الاختبار على معامل ثبات الاختبار.
- برنامج مقترح لخفض قلق الاختبار وأثره على خصائصه السيكومترية (الصدق - الثبات).
- تأثير التصحيح من أثر التخمين في أسئلة الاختيار من متعدد وفاعلية المشتتات على ثبات الدرجات.
- دراسة مقارنة بين نظرية الاستجابة للمفردة والنظرية التقليدية في تحليل مفردات اختبارات الاختيار من متعدد التحصيلية.
- دراسة أثر اختلاف المستويات المعرفية للمفردات على خصائصها السيكومترية (الصعوبة - التمييز)، وعلى ثبات درجات الاختبار.
- أثر ترتيب مفردات الاختبار من حيث صعوبتها المعرفية (تذكر - فهم - تطبيق) على ثبات درجات الطلاب.
- دراسة أثر استخدام التغذية الراجعة على ثبات درجات الاختبار.