المفاضلة بين أسئلة التكميل وأسئلة الاختيار من متعدد لقياس التذكر المعرفي في ضوء الخصائص السيكومترية

أ. د. كامل ثامر الكبيسي
د. خمائل مهدي صالح سلمان

ملخص البحث:

يهدف البحث إلى المفاضلة بين أسئلة الاختيار من متعدد التي تمثل التعرف وأسئلة التكميل التي تمثل الاستدعاء في قياس التذكر المعرفي في ضوء الخصائص السيكومترية للاختبار المتمثلة بالصدق والثبات ومؤشر الحساسية، ولفقراته المتمثلة بمعاملات صعوبة الفقرة وصدقها وقدرتها على التمييز، ولتحقيق ذلك أعدّ الباحثان اختباراً تحصيلياً لمادة القياس والتقويم للصف الرابع في معاهد المعلمين والمعلمات بنمطين، أحدهما بنمط الاختيار من متعدد والثاني بنمط التكميل، وقاما بتطبيق النمطين على عينة مكونة من (٢٠٠) طالباً وطالبة بأسلوب القياسات المتكررة، وبعد تحليل الإجابات وحساب الخصائص السيكومترية لكل نمط وإجراء المفاضلة بينهما في ضوء هذه الخصائص على وفق معيار للمفاضلة أعدّه الباحثان، أظهرت النتائج أفضلية أسئلة التكميل في قياس التذكر المعرفي على أسئلة الاختيار من متعدد، إذ حصل على (٦) درجات من مجموع درجات المفاضلة البالغ (١٨) درجة وبنسبة (٣٣.٣٣%) في حين حصلت أسئلة الاختيار من متعدد على درجة واحدة من مجموع درجات المفاضلة وبنسبة (٥.٥٦%)، وعليه من الأفضل في قياس التذكر في ضوء تصنيف بلوم استعمال أسئلة التكميل.

المقدمة:

لقد شهد علم القياس والتقويم التربوي تغيرات جوهرية في مفاهيمه ومبادئه وفي منهجياته وتقنياته خلال العقود القليلة الماضية، أدت إلى تطورات واضحة في أسس وقواعد بناء أدوات القياس المختلفة، وذلك لما للقياس والتقويم من أهمية بالغة في صنع القرارات التربوية، والتعرف على فاعلية البرامج التربوية والمناهج الدراسية والوقوف على جوانب القوة والضعف في التحصيل المعرفي للطلبة (علام، ٢٠٠٧: ص١٩)، الذي يعد الارتقاء به من أهم أهداف العملية التعليمية والتعلمية حتى أن "بلوم وآخرون" (Bloom, et. al: 1981) وضعوا النمو المعرفي في مقدمة مجالات الأهداف التربوية، وصنفوه إلى ست فئات رئيسة على شكل هرم، قاعدته المعرفة أو التذكر، وظلت المعرفة والتذكر أيضاً في قاعدة هرم تصنيف "بلوم" المعدل، إذ طور هذا التصنيف من مستويات المعرفة والفهم والتطبيق والتحليل والتقويم إلى مستويات المعرفة أو التذكر، والفهم، والتطبيق، والتحليل، والتقويم، والإبداع (Stiggins, 1994: P. 40) كما جاء التعرف (التذكر) في قاعدة تصنيف "وارد ووارد" لمستويات النمو المعرفي التي تكونت من التعرف والاستنتاج والتحليل والتقويم وحل المشكلات، ولا يختلف التعرف في هذا التصنيف عما هو عليه في التذكر أو المعرفة في تصنيف "بلوم" إذ يتعرف فيه الطالب أو يسترجع المعلومات التي حفظها بجهاز الذاكرة بصيغتها الأصلية نفسها (Ward & Ward, 1999: PP.35-38).

وقد جاءت المعرفة أو التذكر في قاعدة هذه التصنيفات للأهداف التربوية لمستويات النمو المعرفي، لكون المعرفة أساس عملية التعلم وقاعدته وبدونها لا يمكن للمتعلم أن يصل إلى المستويات المعرفية العليا، ولا غنى للطالب عن التذكر، سواء

في حياته الدراسية أو في حياته العملية، حتى أصبح التذكر مرادفاً للتعلم في بعض المواقف، وفي اكتساب المعارف والحقائق والمفاهيم والمصطلحات .

(عدس وتوق، ١٩٨٤: ص٤٣-٤٤).

من هذا المنطلق نلاحظ أن التذكر (المعرفة) حظي بحيز كبير من الأهداف السلوكية لتدريس معظم المواد الدراسية، ونال قدراً كبيراً من أسئلة الاختبارات التحصيلية، سواء أُعدت من المعلمين أو من الباحثين، وقد توصلت الدراسات التي أجريت في هذا المجال إلى أن المعلومات أو المعارف التي يتعلمها الطالب تُخزن في جهاز الذاكرة من خلال الترميز (Encoding) الذي يسهل عملية التخزين (Storage) في جهاز الذاكرة (يوسف، ٢٠١١: ص١١١)، ويتم استرجاعها (Retrieval) وقت الحاجة من جهاز الذاكرة بأحدى وسيلتين أو طريقتين هما:

١. **التعرف Recognition** : وهو قدرة الشخص على استرجاع المعلومات بدون عناء أو جهد كبير، بسبب الاستعانة بما يعرض عليه منها أثناء التعرف، أي تقدم له مجموعة اختيارات أو بدائل، ويُطلب منه التعرف على أحدها المطلوب منه تذكره، وخير مثال على ذلك هي أسئلة الاختيار من متعدد التي تقتصر على قياس مستوى التذكر (العتوم، ٢٠٠٤: ص٢٩١).

٢. **الاستدعاء Recall**: وهو القدرة على استحضار المعلومات المخزونة في الذاكرة،، وتستدعى عادة عن طريق الاستعانة بأشياء ترتبط بها تسمى التلميحات أو الهاديات Cues (يوسف، ٢٠١١: ص١١٢).

ويرى الباحثان إن أسئلة التكميل خير دليل على ذلك، لأن الطالب يستدعي فيها المعومات بدلاً من التعرف عليها، كما هو في أسئلة الاختيار من متعدد التي تعتمد على التعرف (علام، ٢٠١٠: ص١١٣).

بيد أن هذا لا يعني أن التذكر المعرفي لا يقاس إلا بأسئلة الاختيار من متعدد أو بأسئلة التكميل، بل أن جميع أنماط الأسئلة يمكن أن تقيس التذكر ، ولكن الأسئلة الموضوعية، وكما تسمى بأسئلة الإجابات المحددة البنية التي تعتمد على الاستجابة المختارة، هي الأكثر استخداماً في قياس التذكر، لاسيما أسئلة الاختيار من متعدد التي تعد من أفضل أنماط الأسئلة الموضوعية، لما تمتاز به من قلة التخمين وموضوعية التصحيح وسهولته وشمولها للمادة، كما يمكن استخدام أسئلة التكميل في قياس التذكر التي تنتمي إلى أسئلة الإجابات المقيدة على الرغم من أنها تعتمد على الإجابة المنشأة، لما تتميز به من سهولة الاستخدام والإعداد، وانعدام التخمين وقلة الذاتية في التصحيح، وشموليتها للمادة، قياساً بالأسئلة المقالية وحتى بأسئلة الإجابات القصيرة (Lukhele, et. al., 1994: PP.234-240).

ويبدو أن أسئلة الاختيار من متعدد أسهل للطالب في الإجابة من أسئلة التكميل لكون أسئلة الاختيار من متعدد تعتمد على التعرف الذي يعد أسهل من الاستدعاء الذي تعتمد عليه أسئلة التكميل، لأن صعوبة الاستدعاء تكمن في القدرة على استحضار الماضي، في حين أن التعرف يعتمد على تذكر شيء هو معروض أمامه (الهاشمي، ١٩٩٢: ٢١٣)، ولكن لا يمكن أن نقول أن أسئلة الاختيار من متعدد هي أفضل من أسئلة التكميل في قياس التذكر، لأن المفاضلة لا تعتمد على سهولة الإجابة بل على قدرة الأسئلة في قياس ما أعدت لقياسه بدقة وبأقل ما يمكن

من أخطاء (عودة، ١٩٩٨: ص٣٣٥)، لذلك كلما كان المقياس قادراً على قياس الخصيصة التي أعد لقياسها بأقل ما يمكن من أخطاء، كان أفضل من غيره الذي يقيس الخصيصة نفسها ولكن بخطأ أكبر، ولقد أجمع علماء القياس على أن هناك خصائص سيكومترية للاختبار أو المقياس ولفقراته تؤشر قدرته على القياس بأقل ما يمكن من أخطاء، ولعل من أهم الخصائص السيكومترية للاختبار هي الصدق والثبات ومؤشر الحساسية في قياس العلاقة بين الخاصية والأداء عليها، أما أهم الخصائص السيكومترية التي ينبغي أن تتوافر بدرجة جيدة في فقرات الاختبار هي معاملات التميز والصعوبة والصدق (سلمان، ٢٠٠٧: ص١٦).

لذلك ظهر جدل كبير حول الكيفية التي تعد بها أسئلة الاختبارات التحصيلية، إذ هناك ثلاثة أنماط رئيسة لبناء أو صياغة فقرات الاختبار ألتحصيلي وهي نمط أسئلة الاستجابة المختار (Selected Response Items) التي تشمل الأسئلة الموضوعية بأنواعها الثلاثة: الاختيار من متعدد، والصواب والخطأ، والمزاوجة، ونمط أسئلة الاستجابة المنشأة (Constructed Response Items) التي تتوزع إلى نمطين هما نمط الإجابات المقيدة مثل أسئلة التكميل وأسئلة الإجابات القصيرة، ونمط الإجابات الحرة، والتي تمثلها الأسئلة المقالية.

(Lukhele, et.al. 1994, pp: 24- 24)]

ويبدو أن هناك اختلافاً في المتطلبات المعرفية بين فقرات أسئلة الاستجابات المختارة التي تمثلها في هذا البحث أسئلة الاختيار من متعدد، وأسئلة الاستجابة المنشأة المقيدة التي تمثلها في هذا البحث أسئلة التكميل، نتيجة لاختلاف مدى حرية الطالب في تكوين الإجابة بين أسلوبي استرجاع المعرفة من جهاز الذاكرة، بينما

أشارت دراسة (كندي ولستاد، ١٩٩٧) إلى أن الاختبارات التي فقراتها من نمط الاستجابة المنشأة تقيس ما تقيسه الفقرات التي من نمط الاستجابة المختارة.

.(Kennedy & Walstad, 1997, pp: 459- 362)

بينما يذكر (تروب وماكروري، ١٩٩٢)، أن بعض الفقرات من نمط الاستجابة المنشأة تقيس قدرات تختلف عن تلك التي تقيسها أسئلة الاستجابة المختارة ولكن عندما تكون أسئلة الاستجابات المنشأة محددة جداً تتطلب بضع كلمات فأن الفروق تكون غير كبيرة أو غير هامة.

.(Traub& Macrury, 1992, p 48)

لهذا نجد أن "بردجمان" ، ١٩٩٢، يؤكد الحاجة إلى وجود دليل علمي قاطع يفيد أن النمطين يقيسان عمليات معرفية متشابهة أو مختلفة.

.(Bridgeman, 1992, p.268)

بينما يعزو "تروب، ١٩٩٣" الاختلاف بين ما تقيسه الأنماط المختلفة للفقرات إلى اختلاف المحتوى الذي تقيسه، أو إلى اختلاف شكلها، أو الاختلاف في كليهما، لذلك أوصى بضرورة اعتماد الباحثين فقرات متشابهة في الشكل وفي المحتوى، التي تتطلب استخدام الجذر أو الأرومة stem نفسها في النمطين والتي تسمى بعائلة الفقرات (Item family)، لذلك أدى هذا الاختلاف إلى اختلاف الدراسات في تحديد أفضلية نمط معين للأسئلة في قياس المستويات المعرفية.

.(Traub, 1993, pp. 30- 35)

مشكلة البحث:

من خلال ما جاء في مقدمة هذا البحث، فضلاً عما لمسه الباحثان في أن هناك بعض المعلمين وفي مختلف المراحل الدراسية يميلون إلى استخدام أسئلة الاختيار من متعدد في قياس التذكر، لكونها من أفضل أنواع الأسئلة الموضوعية، في حيث يميل البعض إلى استخدام أسئلة التكميل لكونها لا تتعرض للتخمين وسهلة الأعداد، وأقل ذاتية من الأنماط الأخرى لأسئلة الاستجابات المنشأة، ولكن كما يذكر " بروجمان، ١٩٩٢" ضرورة وجود دليل علمي قاطع لتفضيل أحد هذين النمطين على الأخر (Bridgeman, 1992, p. 268).

بيد أن المفاضلة في قياس التذكر المعرفي قد تكمن في قدرة الاختبار أو الأسئلة في قياس ما أعد لقياسه بأقل ما يمكن من أخطاء تؤشر عليها الخصائص السيكومترية للاختبار ولفقراته، وليس من خلال الدرجات التي يحصل عليها الطالب، وعليه يمكن تحديد مشكلة البحث الحالي من خلال الإجابة عن السؤال الأتي :"**أيهما أفضل في قياس التذكر المعرفي، أسئلة التكميل أو أسئلة الاختيار من متعدد وذلك من خلال المفاضلة بينهما في الخصائص السيكومترية للاختبار ولفقراته؟**"

أهمية البحث: أن أهمية أي بحث عادة تتجلى في مقدار ما يضيفه إلى المعرفة من الناحيتين النظرية والتطبيقية، وعليه فأن الأهمية النظرية للبحث الحالي فيمكن أن تنبثق من أهمية:

١. الاختبارات التحصيلية: إذ تلعب دوراً رئيساً في عملية التقويم وعليها تعتمد معظم القرارات التربوية المختلفة، وبها نكمم مقدار ما يتعلمه الطلبة، ونحدد نواحي القوة والضعف في تحصيلهم المعرفي (Thorndike, et. at, 1991, p. 38).

٢. التذكر: يعد التذكر عملية عقلية اختيارية منظورة، فهو عملية تخزين لما تعلمه الفرد واسترجاعه عند الحاجة (العياصرة، ٢٠١١: ٤٢٢)، لذا فهو أساس عمليات

النمو المعرفي، إذ بدون التذكر لا يمكن للفرد أن يتواصل في حياته اليومية ولا في حياته الدراسية، لذلك وضع في قاعدة الهرم في معظم تصنيفات مستويات النمو المعرفي.

٣. الخصائص السيكومترية للاختبار ولفقراته: إذ بدونها لا يمكن أن نعد أداة قياسية أو اختباريه قادرة على قياس الغرض الذي أعدت من اجله، لذلك كلما كانت هذه الخصائص بدرجات عالية قلت أخطاء القياس (سلمان، ٢٠٠٧، ص١٧).

أما الأهمية التطبيقية للبحث الحلي فيمكن أن تتحدد بما يأتي:

١. أمكانية الإفادة من نتائج هذا البحث في توجيه المعلمين والباحثين نحو استخدام النمط المفضل في الاختبارات التحصيلية التي ترمي إلى قياس التذكر المعرفي.

٢. المساهمة مع الدراسات السابقة في حسم الجدل حول أفضلية أي نمط من أنماط الأسئلة الأختبارية، ولاسيما بين أسئلة الاستجابة المختارة، وأسئلة الاستجابة المنشأة.

هدف البحث:

يهدف البحث الحالي إلى المفاضلة بين أسئلة التكميل وأسئلة الاختيار من متعدد لقياس التذكر المعرفي، في ضوء الخصائص السيكومترية للاختبار المتمثلة بالصدق والثبات ومؤشر الحساسية، ولفقراته المتمثلة بمعاملات التمييز والصعوبة والصدق، ولتحقيق هذا الهدف وضع الباحثان الفرضيات الصفرية الآتية:

١. لا يوجد فرق ذو دلالة إحصائية عند مستوى (٠.٠٥) في معاملي صدق الاختبار بين أسئلة التكميل وأسئلة الاختيار من متعدد.

٢. لا يوجد فرق ذو دلالة إحصائية عند مستوى (٠.٠٥) في معاملي ثبات الاختبار بين أسئلة التكميل وأسئلة الاختيار من متعدد.

٣. لا يوجد اختلاف في مستوى دلالة مؤشر حساسية الاختبار بين أسئلة التكميل وأسئلة الاختيار من متعدد.
٤. لا يوجد فرق ذو دلالة إحصائية عند مستوى (٠.٠٥) في متوسطي معاملات تمييز فقرات تبعا لنوع فقرات الاختبار بين أسئلة التكميل وأسئلة الاختيار من متعدد.
٥. لا يوجد فرق ذو دلالة إحصائية عند مستوى (٠.٠٥) في متوسطي معاملات صدق فقرات الاختبار بين أسئلة التكميل وأسئلة الاختيار من متعدد.
٦. أ- لا يوجد فرق ذو دلالة إحصائية عند مستوى (٠.٠٥) بين متوسطي معاملات صعوبة فقرات الاختبار بأسئلة التكميل والمتوسط المعياري* لمعاملات الصعوبة البالغ (٠.٥٠).

ب- لا يوجد فرق ذو دلالة إحصائية عند مستوى (٠.٠٥) بين متوسطي معاملات صعوبة فقرات الاختبار بأسئلة الاختيار من متعدد والمتوسط المعياري لمعاملات الصعوبة البالغ (٠.٥٠).

حدود البحث:

على الرغم من إن طبيعة هذا البحث وأهدافه لا تتأثر كثيراً بنوع المادة الدراسية أو المرحلة الدراسية، إلا أن الباحثان اقتصرا في إجراءات بحثهما على مادة

* المتوسط المعياري لمعاملات الصعوبة هو (٠.٥٠) إذ يشير علماء القياس أن أفضل معامل صعوبة لفقرات الاختبارات المعيارية المرجح هو (٠.٥٠).

القياس والتقويم لطالبات الصف الرابع بمعاهد إعداد المعلمات في محافظة بغداد للعام الدراسي ٢٠١٠/٢٠١١.

تحديد المصطلحات:

لقد ارتأى الباحثان أن يعرّفا المصطلحات الأساسية للبحث إجرائياً وكالآتي:

١. المفاضلة Priority:

هي تفضيل أسئلة التكميل على أسئلة الصواب والخطأ أو بالعكس في قياس التذكر المعرفي في تصنيف (بلوم) في ضوء مجموع درجات المفاضلة بينهما للخصائص السيكومترية للاختبار ولفقراته على وفق معيار المفاضلة المعتمد في هذا البحث.

٢. فقرات التكميل:

تسمى أحياناً إكمال الفراغات، وهي عبارة عن عبارات أو جمل تعبر عن مضمون معين في المادة الدراسية، وفيها كلمة أو أكثر ناقصة لا يتم معنى العبارة إلا بوجودها، أي لا تصبح العبارة تامة المعنى إلا بإكمال الفراغ بكلمة أو أكثر.

٣. فقرات الاختيار من متعدد:

هي الأسئلة التي يتكون فيها السؤال من عبارة على شكل سؤال أو عبارة تقريرية تسمى مقدمة، وهذه العبارة ناقصة المعنى لا يتم إلا بأحد بدائل الإجابة عنها، وعادة يكون بديل واحد من هذه البدائل صحيحاً والأخرى خاطئة، وقد اعتمدا الباحثان أربعة بدائل للإجابة أحدها صحيحاً والبدائل الثلاثة الأخرى خاطئة وبعبارات تقريرية.

٤. الخصائص السيكومترية Sychometric Properties:

هي مؤشرات أو دلائل قياسية تؤشر جودة أو دقة الاختبار أو فقراته في قياس ما أُعد لقياسه بأقل ما يمكن من أخطاء (Dick & Hagert, 1971, P.13)، وهذه المؤشرات في هذا البحث هي:

أ- الصدق Validity:

وهو معامل الصدق التلازمي لكل من أسئلة التكميل، وأسئلة الاختيار من متعدد مع محك خارجي وهو درجات الامتحان النهائي لمادة القياس والتقويم الذي يجريه المدرس أو المدرسة للطلبة في نهاية العام الدراسي ٢٠١٠/٢٠١١.

ب- الثبات Reliability:

وهو معامل التجانس بين فقرات أسئلة التكميل مرة، وبين أسئلة الاختيار من متعدد مرة أخرى، في قياس ما يفترض أن يقيسه كل نمط وهو التذكر المعرفي للمادة الدراسية، وذلك باستخدام معادلة هويت Hoyt من نتائج تحليل التباين الثنائي بدون تفاعل (Fox, 1969, P. 249).

ج- مؤشر الحساسية Sensitivity:

قدرة المقياس على قياس العلاقة بين الخاصية والأداء عليها ويحسب من نتائج تحليل التباين الثنائي "بدون تفاعل" للدرجات إذ تعتمد معادلة حسابه على متوسط التباين بين الأفراد ومتوسط تباين الخطأ وتختبر دلالته الإحصائية في ضوء مستويات الدلالة للتوزيع الطبيعي (عبد الرحمن، ١٩٩٨، ص١٧٣).

د- معامل تميز الفقرة Discrimination Coefficient:

وهو قدرة الفقرة على التمييز بين ذوي المستويات العليا والدنيا في المجموع الكلي لدرجات أسئلة التكميل مرة، ولدرجات أسئلة الاختيار من متعدد مرة أخرى،

ويحسب من خلال (عدد الإجابات الصحيحة في المجموعة العليا – عدد الإجابات الصحيحة في المجموعة الدنيا) مقسوماً على عدد أفراد أحدى المجموعتين.

هـ- معامل صعوبة الفقرة:

هو عبارة عن ناتج قسمة عدد الإجابات الخاطئة عن الفقرة عند أفراد العينة على عدد أفراد العينة، والذي يفضل عادة أن يكون (0.50).

و- معامل صدق الفقرة Item's Validity Coefficient:

وهو معامل ارتباط الفقرة بالدرجة الكلية للاختبار، إذ تشير "انستازي" إلى أن ارتباط درجة الفقرة بمحك خارجي أو داخلي هو مؤشر على صدقها، وحينما لا يتوفر محك خارجي فإن أفضل محك داخلي هو الدرجة الكلية للاختبار .

.(Anastasi, 1988, P. 206)

دراسات سابقة:

١. أجري "العلي، ١٩٨٩" دراسة استهدفت المقارنة في الصدق التلازمي للاختبار في ثلاثة أنماط من الاختبارات هي اختبار أسئلة من نوع الاختيار من متعدد، واختبار أسئلة من نوع الإجابات القصيرة، أما الاختبار الثالث فكانت أسئلة مزيج من النوعين السابقين، وقد شملت المقارنة مستوى الأداء على الاختبار الكلي وعلى مستويات التذكر والاستيعاب والتطبيق كل على انفراد تكونت العينة من (١٩٧) طالباً من طلبة الصف الأول الثانوي في الأردن، ثم تقسيمها إلى ثلاث مجموعات، كل مجموعة طبق عليها نمط من الأنماط الثلاثة للاختبار وأظهرت النتائج وجود فروق بدلالة إحصائية عند مستوى (0.5) في معامل الصدق التلازمي للاختبار الكلي، ولصالح الاختبار المزيج من نمطين، أسئلة الاختيار من متعدد وأسئلة الإجابات القصيرة، وكانت الفروق لصالح الإجابات القصيرة في مستوى التذكر بينما لم يكن الفرق بين الأنماط

الثلاثة للاختبار بدلالة إحصائية عند مستوى (٠.٠٥) في الصدق التلازمي في كل من مستوى الاستيعاب ومستوى التطبيق .

(العلي، ١٩٨٩).

٢. وقامت "بادجر، ١٩٩٠" (Badger) بدراسة استهدفت المقارنة بين استخدام فقرات الاختيار من متعدد وأسئلة الإجابات القصيرة، أجريت على عينة من طلبة الصف الثاني عشر في ولاية "مساشوستش" الأمريكية. حللت الدراسة درجات اختبارات أسئلة العلوم والرياضيات، تكونت اختبارات الرياضيات من (٣٨) فقرة اختبار من متعدد و (١٢) فقرة منشأة من نوع الإجابات القصيرة، وتكونت اختبارات العلوم من (٣٦١) فقرة من نوع الاختيار من متعدد، و(١١) فقرة منشأة من نوع الإجابات القصيرة، وأظهرت النتائج أن الطلبة الذين كان أدائهم على الاختيار من متعدد جيداً، كان جيداً أيضاً على نمط الإجابات القصيرة، ولكن أدائهم على النمطين لم يكن متكافئاً.

٣. واجري "بردجمان Bridgeman,1992" دراسة استهدفت المقارنة بين نمطي الاختيار من متعدد والتكميل، تكونت العينة من (٦٧٣٥) طالباً وطالبة اختيرت من ست ولايات أمريكية تقدموا لامتحان (GRE- Q) إذ قسمت العينة إلى قسمين، طبق على القسم الأول منها الامتحان المذكور، مضاف إليه (١٤) فقرة من نمط الاختيار من متعدد تقيس بعض الخوارزميات الرياضية ومهارة التقدير، وطبق على القسم الثاني الامتحان نفسه أيضا، مضاف إليه (١٤) فقرة، وهي الفقرات نفسها التي أضيفت للقسم الأول ولكن من غير بدائل أي حولت إلى أسئلة تكميل كانت الأرومة نفسها في أسئلة الاختيار من متعدد وأسئلة التكميل. وأظهرت النتائج أن نسبة النجاح

في أسئلة الاختيار من متعدد أعلى مما هي عليه في أسئلة التكميل ولكن النمطان يرتبان الطلبة على وفق الدرجة الكلية بالطريقة نفسها .

.(Bridgeman, 1992. 253- 271)

٤. وقام "بارنت وناجي Barnett& Nagy, 1995" بدراسة استهدفت المقارنة بين أسئلة الاختيار من متعدد وأسئلة الاستجابة المنشأة، وهي الإجابة القصيرة، وقد أجريت الدراسة على (٢٦١) طالباً وطالبة بامتحان الكيمياء لطلبة البكالوريوس في الولايات المتحدة الأمريكية، وكان النمطان لهما جذر أو أرومة الفقرة نفسها، وأظهرت النتائج وجود فروق معنوية في أنماط الأخطاء وفي استراتيجيات حل السؤال لصالح أسئلة الاختيار من متعدد.

. (Barnett& Nagy, 1995, pp. 18- 35)

٥. وفي عام ١٩٩٨ قام "دبك" بدراسة استهدفت معرفة أثر شكل الفقرات وعدد بدائل الاختيار من متعدد في صدق الاختبار، أعدت الدراسة اختباراً تحصيليا في مادة الفيزياء لطلبة الصف الثاني ثانوي يتكون من (٤٠) فقرة من نوع الاختبار من متعدد بأربعة بدائل وبثلاثة بدائل، ومثلها من نوع التكميل، طبقت الأنماط الثلاثة للاختبار على عينة مكونة من (٣٠٠) طالباً. وحللت النتائج فأظهرت وجود فرق بدلالة إحصائية عند مستوى (٠.٠٥) في معامل الصدق المرتبط بمحك، لصالح اختبار التكميل (دبك، ١٩٩٨).

٦. وفي عام ٢٠٠٠ قامت "الحموري" بدراسة استهدفت الكشف عن أثر اختلاف نمط الاختبار حسب نوعية الاستجابة المتطلبة من فقراته، منشأة أو مختارة في الصدق العاملي والتنبؤي، تكونت عينة الدراسة من (٢٢٤) طالباً و (٢٣٠) طالبة اختيرت من

مدارس منطقة عمان / الأردن من طلبة الصف السادس الأساسي. واعدت الدراسة اختبارين تحصيليين في مادة الرياضيات، احدهما من نوع الاختيار من متعدد، والأخر من نوع التكميل ويتكون كل منهما من (٢٠)فقرة، ولكل فقرة من الاختبارين لها جذر أو الأرومة stem ذاتها، لها بدائل في أسئلة الاختيار من متعدد وبدون بدائل في أسئلة التكميل وتوصلت النتائج إلى أن متوسط درجات نمط الاختيار من متعدد أكبر من متوسط درجات نمط أسئلة التكميل، وأن النمطين لهما قدرة على التنبؤ بالتحصيل تكاد تكون متساوية، كما أظهرت الدراسة أن النمطين يتشبعان على عوامل مختلفة في عددها، ويفسران نسباً مختلفة من التباين.

الاستنتاج من الدراسات السابقة:

يتضح مما تقدم من دراسات سابقة ما يأتي:

١. لم تعتمد هذه الدراسات في المفاضلة بين أنماط الأسئلة الأختبارية الخصائص السيكومترية للاختبار او لفقراته. لذلك تناول البحث الحلي هذه الخصائص.

٢. لم تحدد هذه الدراسات المستويات المعرفية التي تناولتها في الاختبار التحصيلي عدا دراسة العلي، ١٩٨٩، فأنها اعتمدت المستويات الثلاثة الأولى من تصنيف بلوم، وهي التذكر والفهم والتطبيق. أما البحث الحالي فقد أقتصر على مستوى التذكر كي يتمكن من ضبط تأثير المحتوى والهدف.

٣. تباينت الدراسات السابقة في تفضيل أسئلة الاختيار من متعدد أو أسئلة الإجابات المنشأة وبخاصة التكميل، ولذا فأن البحث الحالي سيساهم في تفضيل أحدهما أو عدمه كما في هذه الدراسات.

٤. معظم هذه الدراسات اعتمدت أسئلة الاختيار من متعدد كاستجابة، مختارة وأسئلة التكميل كاستجابة منشأة في المقارنة وهذا ما اعتمد عليه البحث الحالي أيضا.

إجراءات البحث

- مجتمع البحث وعينته:

على الرغم من أن طبيعة البحث الحالي وأهدافه تسمح بإجرائه على أية مرحلة دراسية وعلى أية مادة دراسية أيضاً، لأنه يتطلب اختباراً تحصيلياً بنمطين في أية مادة دراسية، إلا أن الباحثين تسهيلاً وتحديداً لإجراءات البحث اختارا طلبة الصف الرابع في معاهد إعداد المعلمين والمعلمات بمحافظة بغداد ومادة القياس والتقويم التي تدرس لطلبة هذا الصف ميداناً لهذا البحث*، لذا فإن مجتمع البحث الحالي يتكون من طلبة الصفوف الرابعة في معاهد المعلمين والمعلمات في بغداد، أما عينة البحث فتكونت من (٢٠٠) طالباً وطالبة موزعة بالتساوي على الجنسين، اختيرت عشوائياً من طلبة الصف الرابع في معهدين للبنين ومعهدين للبنات التي اختيرت عشوائياً من معاهد المعلمين والمعلمات في محافظة بغداد البالغ عددها (١٤) معهداً وبواقع (٥٠) طالباً أو طالبة من كل معهد، وبحسب الجنس.

- أداة البحث:

إن هدف البحث يتطلب إعداد اختبار تحصيلي بمادة القياس والتقويم لطلبة الصف الرابع بمعاهد المعلمين والمعلمات، ولقياس مستوى التذكر فقط من مستويات تصنيف بلوم للمجال المعرفي وبنمطين، نمط بأسئلة التكميل والآخر بأسئلة الاختيار من متعدد، لذلك قاما الباحثان بالإجراءات الآتية:

* قد يعود هذا الاختيار إلى كون الباحثة الثانية تعمل في معاهد إعداد المعلمين والمعلمات وتدرس هذه المادة التي هي اختصاصها مما تسهل إجراءات بحثها.

١. حللت مادة القياس والتقويم التي تدرس لطلبة الصف الرابع واشتقت منها (٣٥) هدفاً سلوكياً في مجال المعرفة (التذكر) غطت جميع الموضوعات الأساسية للمادة.
٢. قدمت هذه الأهداف إلى (٥) محكمين مختصين بالقياس والتقويم، فأكدوا صلاحيتها، بعد استبعاد ثلاثة أهداف ودمج هدفين آخرين، مما أصبح عدد الأهداف السلوكية (٣٠) هدفاً سلوكياً.
٣. أُعدت فقرات لقياس هذه الأهداف بصيغتين، واحدة بصيغة التكميل (إكمال الفراغات) Completion Items، التي تنتمي إلى نمط الفقرات ذات الاستجابة المنشأة Constructed Response، والأخرى بصيغة الاختيار من متعدد Multiple – Choice Items التي تنتمي إلى نمط الفقرات ذات الاستجابة المختارة Selected Response Items ، وكانت فقرتا التكميل والاختيار من متعدد تقيسان محتوى واحد أي تنتميان إلى عائلة واحدة (Item Family)، إذ يفضل في المقارنة بين هذين النمطين من الأسئلة أن تكون العبارة التقديمية (الأرومة) Stem نفسها للنمطين (Haub, 1993, PP.30-31). كي يتم تحييد عامل المحتوى في التأثير على نتائج الاستجابة، ويبقى تأثير الشكل إذ أن عدم استخدام عائلة واحدة للفقرات قد يعد أحد الأسباب التي يمكن أن تعزى إليها النتائج المتباينة. (حموري، ٢٠٠٠، ص٢).

وقد راعا الباحثان شروط صياغة فقرات أسئلة التكميل وأسئلة الاختيار من متعدد، وتكونت أسئلة الاختيار من متعدد من مقدمة Stem تتبعها (٤) إجابات مقترحة (بدائل) Alternatives وتشمل البدائل إجابة واحدة صحيحة والبدائل الثلاثة

الأخرى خاطئة، وهي المشتتات أو المموهات Distractors، أما أسئلة التكميل فكانت جميعها عبارات ناقصة (فراغات) وضعت في نهاية العبارة وكل فراغ يتكون من كلمة أو كلمتين، ماعدا فقرة واحدة تتطلب (٣) كلمات مترابطة، وذلك بحسب عدد كلمات بدائل الاختيار من متعدد، الملحق (١) يوضح فقرات الاختيار من متعدد، والملحق (٢) يوضح فقرات أسئلة التكميل.

٤. للتثبت من صلاحية الفقرات كما تبدو ظاهرياً في قياس ما أعدت لقياسه قُدم الاختباران إلى (٥) محكمين متخصصين في القياس والتقويم، وطلب منهم الباحثان فحص الفقرات منطقياً وتقدير مدى صلاحيتها لقياس الأهداف التذكرية (المعرفية) التي أعدت لقياسها، فاتضح من ملاحظاتهم إن جميع الفقرات صالحة لقياس الأهداف التي أعدت لقياسها، وبموافقة جميع المحكمين، وبذلك فقد تحقق الباحثان من الصدق الظاهري Face Validity، إذ يشير "إيبل" إلى أن أفضل من يقوم بالتحقق من الصدق الظاهري للاختبار هم الخبراء المتخصصين في المجال. (Ebel, 1972, P. 555).

٥. **طريقة التصحيح وحساب الدرجات:** تعطى لكل اختيار صحيح في كل فقرة من فقرات أسئلة الاختيار من متعدد درجة واحدة، وتعطى درجة صفر للاختيار الخاطئ، وبذلك فإن أعلى درجة كلية ممكنة هي (٣٠) درجة، واقل درجة ممكنة هي (صفر) وبمتوسط نظري للاختبار مقداره (١٥) درجة، أما بالنسبة لنمط أسئلة التكميل فتعطى لكل فراغ إجابته صحيحة درجة واحدة، وتعطى درجة صفر للإجابة الخاطئة، وبذلك تكون الدرجة الكلية العليا (٣٠) درجة، والدرجة الكلية الصغرى (صفر) وبمتوسط نظري للاختبار مقداره (١٥) درجة أيضاً.

- طريقة تنفيذ التجربة (تطبيق الاختبارين):

بعد انتهاء المدرسين والمدرسات من تدريس مادة القياس والتقويم وقبيل بدء الامتحان النهائي قام الباحثان بنفسهما بتطبيق الاختبارين على العينة نفسها بأسلوب القياسات المتكررة، لكون هذا التصميم يحد كثيراً من تأثير المتغيرات الدخيلة، وقد قسمت العينة إلى مجموعتين، قدمت للمجموعة الأولى اختبار أسئلة الاختيار من متعدد أولاً للإجابة عنه وبعد مرور حوالي أسبوعين قدم لها اختبار أسئلة التكميل، إذ قد تعد هذه الفترة كافية إلى حد ما في عدم تذكر الطلبة لإجاباتهم الأولى، وقدمت للمجموعة الثانية أولاً اختبار أسئلة التكميل وبعد مرور حوالي أسبوعين أيضاً قدم لها اختبار أسئلة الاختيار من متعدد، إذ يفضل في هذا التصميم تقديم المعالجات أو مستويات المتغير المستقل بترتيب مختلف بعد تقسيم أفراد العينة إلى مجموعات بحسب المعالجات أو المستويات كي لا تتأثر درجة كل معالجة بترتيب تقديمها على أفراد العينة، وبهذا يتعادل تأثير ترتيب تقديم المعالجات وما قد يصاحب ذلك من تعب أو ملل أو تدريب.

(Pedhazur, 1982, P. 533)، (إسماعيل، ١٩٨٦، ص٢٤٥).

- إجراءات حساب الخصائص السيكومترية:

بعد الانتهاء من الإجابة عن نمطي الاختبار وتصحيح كل منهما على انفراد حُسبت الخصائص السيكومترية للاختبار ولفقراته وكالآتي:

١. معامل صدق الاختبار: لحساب الصدق التلازمي لكل نمط اعتمد الباحثان درجات المدرسين والمدرسات لأفراد العينة في الامتحان النهائي بمادة القياس والتقويم محكاً لهذا الصدق، وحسب معامل ارتباط "بيرسون" Pearson بين

درجات الامتحان النهائي (المحك) ودرجات كل نمط من نمطي أسئلة الاختيار من متعدد وأسئلة التكميل، فكان معامل الصدق التلازمي لنمط التكميل (٠.٨٣) ولنمط الاختيار من متعدد (٠.٧٤).

٢. معامل ثبات الاختبار: حللت درجات كل نمط على حدة من نمطي أسئلة التكميل، وأسئلة الاختيار من متعدد، باستخدام تحليل التباين الثنائي (بدون تفاعل) ANOVA, Two – way فكانت نتائج تحليل التباين بدرجات كل نمط كما هي في الجدول (١).

الجدول (١)

نتائج تحليل التباين لدرجات نمطي أسئلة التكميل وأسئلة الاختيار من متعدد

النمط	مصدر التباين	مجموع المربعات S.S.	درجة الحرية D.F.	متوسط المربعات S.M.
أسئلة التكميل	بين الفقرات	٢٥.١٠٥٣	٢٩	٠.٨٦٥٧
	بين الأفراد	٢١٥.٨٣٥٤	١٩٩	١.٠٨٤٦
	الخطأ	١٢٤٤.٨٠٤٧	٥٧٧١	٠.٢١٥٧
	الكلي	١٤٨٥.٧٤٥٤	٥٩٩٩	
أسئلة الاختيار من متعدد	بين الفقرات	٢٢.٤١٠٣	٢٩	٠.٧٧٢٨
	بين الأفراد	١٦٩.٠٧٠٤	١٩٩	٠.٨٤٩٦

	الخطأ	١٢٩٨.٩٩٤٤	٥٧٧١	٠.٢٢٥٠٩
	الكلي	١٤٩٠.٤٧٥١	٥٩٩٩	

وبعد استخدام معادلة "هويت" من نتائج تحليل التباين (Fox, 1969, P. 249)، كان معامل ثبات أسئلة التكميل (٠.٨٠١) ومعامل ثبات أسئلة الاختيار من متعدد (٠.٧٣٥).

٣. مؤشر حساسية الاختبار: بعد استخدام معادلة مؤشر حساسية الاختبار من نتائج تحليل التباين الجدول (١) التي تعتمد على متوسط تباين الخطأ ومتوسط التباين بين الأفراد، وتختبر دلالته الإحصائية في ضوء مستويات الدلالة الإحصائية للتوزيع الطبيعي (عبد الرحمن، ١٩٩٨، ص١٧٣)، كانت القيمة الزائية (Z) لمؤشر حساسية الاختبار في نمط أسئلة التكميل هي (٢.٠٠٧) وفي نمط أسئلة الاختيار من متعدد (١.٦٦٦).

٤. معاملات تمييز الفقرات: حسب معامل تمييز كل فقرة من فقرات أسئلة الاختيار من متعدد وأسئلة التكميل باستخدام معادلة التمييز في حالة الإجابة عن الفقرة تكون متقطعة (صح، خطأ) (علام، ٢٠٠٧، ص٢٥٤)، فاتضح أن جميع الفقرات لها قدرة على التمييز، إذ كان أقل معامل تمييز في أسئلة التكميل (٠.٣٨) وفي أسئلة الاختيار من متعدد (٠.٣٠) في حين كان أعلى معامل تمييز في أسئلة التكميل (٠.٧٩) وفي أسئلة الاختيار من متعدد (٠.٧٥).

٥. معاملات صدق الفقرات: حسبت معاملات صدق الفقرات سواء في نمط أسئلة التكميل أو في أسئلة الاختيار من متعدد باستخدام ارتباط درجة الفقرة بالدرجة الكلية، إذ تشير "انستازي" Anastasi إلى أن ارتباط الفقرة بالدرجة الكلية

للاختبار مؤشر على صدقها (Anastasi, 1988, P. 206)، واستخدم الباحثان معامل ارتباط (بوينت بايسيريال) Point Biserial، لكون الدرجة على الفقرة متقطعة، وقد تراوحت معاملات صدق الفقرات في أسئلة التكميل ما بين (٠.٣٥- ٠.٧١)، وفي أسئلة الاختيار من متعدد ما بين (٠.٢٧- ٠.٦٣).

٦. معاملات صعوبة الفقرات: استخدم الباحثان في حساب معاملات صعوبة الفقرات سواء في أسئلة التكميل أو في أسئلة الاختيار من متعدد معادلة حساب صعوبة الفقرات عندما تكون الإجابة عنها متقطعة (صح، خطأ) والتي تعتمد على عدد الإجابات الخاطئة مقسوماً على عدد أفراد العينة (علام، ٢٠٠٧، ص٢٥١)، وقد تراوحت معاملات صعوبة أسئلة التكميل ما بين (٠.٣٤- ٠.٦١) في حين تراوحت معاملات صعوبة أسئلة الاختيار من متعدد ما بين (٠.٢٣-٠.٥٥).

معيار المفاضلة ودرجاتها:

لقد وضع الباحثان معيار للمفاضلة بين أسئلة التكميل وأسئلة الاختيار من متعدد وكالآتي:

١. إذا تفوق نمط على النمط الآخر في معامل صدق الاختبار وبفرق دال إحصائياً عند مستوى (٠.٠٠١) فتعطى له (٣) درجات للمفاضلة، أما إذا كان التفوق عند مستوى(٠.٠١) فتعطى له درجتان للمفاضلة، وأما إذا كان التفوق عند مستوى (٠.٠٥) فتعطى له درجة واحدة للمفاضلة، وعندما لا يكون الفرق بينهما في معامل الصدق عند مستوى (٠.٠٥) فتعطى درجة (صفر) لكل منهما.

٢. ينطبق المعيار (١) آنفاً على المفاضلة بين النمطين في معامل ثبات الاختبار.

٣. إذا كان مستوى دلالة مؤشر حساسية أي نمط عند مستوى (٠.٠٠١) فتعطى للنمط (٣) درجات للمفاضلة، أما إذا كان عند مستوى (٠.٠١) تعطى للنمط (درجتان) للمفاضلة، وإذا كان عند مستوى (٠.٠٥) فتعطى للنمط درجة (واحدة) للمفاضلة، وعندما لا يكون مؤشر الحساسية بدلالة إحصائية عند مستوى (٠.٠٥) فتعطى له درجة (صفر) للمفاضلة.

٤. ينطبق المعيار(١) آنفاً على المفاضلة بين النمطين في معاملات تمييز الفقرات.

٥. ينطبق المعيار(١) آنفاً على المفاضلة بين النمطين أيضاً في معاملات صدق الفقرات.

٦. إذا كان الفرق بين متوسط معاملات صعوبة فقرات النمط والمتوسط المعياري لمعاملات الصعوبة لم يكن بدلالة إحصائية عند مستوى (٠.٠٥) فتعطى للنمط (٣) درجات للمفاضلة، وإذا كان الفرق بدلالة إحصائية عند مستوى (٠.٠٥) فتعطى للنمط (درجتان) للمفاضلة، أما إذا كان الفرق بدلالة إحصائية عند مستوى (٠.٠١) فتعطى للنمط درجة (واحدة) للمفاضلة، أما إذا كان الفرق بدلالة إحصائية عند مستوى (٠.٠٠١) فتعطى للنمط درجة (صفر) للمفاضلة.

نتائج البحث

١. صدق الاختبار:

بلغ معامل الصدق التلازمي للاختبار بأسئلة التكميل (٠.٨٣) وبلغ هذا الصدق بأسئلة الاختيار من متعدد (٠.٧٤)، ولمعرفة دلالة الفرق بين هذين المعاملين، استخدم الباحثان اختبار (Z) لدلالة الفرق بين معاملي الارتباط (Glass & Stanley, 1970, P. 311) فأتضح إن الفرق كان بدلالة إحصائية عند مستوى (٠.٠٥) ولصالح أسئلة التكميل، والجدول (١) يوضح ذلك، مما تعطى أسئلة التكميل درجة واحدة في المفاضلة وتعطى أسئلة الاختيار من متعدد درجة (صفر) على وفق

المعيار المعتمد في هذا البحث، وعليه ترفض الفرضية الصفرية الأولى وتقبل الفرضية البديلة.

الجدول (١)

قيمة (Z) المحسوبة والجدولية لدلالة الفرق بين معاملي صدق أسئلة التكميل وأسئلة الاختيار من متعدد

نوع الأسئلة	معامل الصدق R	القيمة المعيارية ZR	قيمة Z		مستوى الدلالة
			المحسوبة	الجدولية	
التكميل	٠.٨٣	١.١٢٨	٢.٣٦٢	١.٩٦	دالة عند (٠.٠٥) لصالح أسئلة التكميل
الاختيار من متعدد	٠.٧٤	٠.٩٥٠			

٢. ثبات الاختبار:

كان معامل ثبات الاختبار بأسئلة التكميل (٠.٨٠١)، وكان بأسئلة الاختيار من متعدد (٠.٧٣٥)، ولمعرفة دلالة الفرق بين هذين المعاملين باستخدام اختبار (Z) لدلالة الفرق بين معامل الارتباط بوصف إن معامل الثبات هو في حقيقته معامل ارتباط الاختبار مع نفسه (سلمان، ٢٠٠٧، ص١١٠)، اتضح أن الفرق لم يكن بدلالة إحصائية عند مستوى (٠.٠٥) إذ كانت قيمة (Z) المحسوبة (١.٦٠٦) أصغر من القيمة الجدولية (١.٩٦)، والجدول (٢) يوضح ذلك، لذا يعطى كل نمط منهما درجة (صفر) للمفاضلة في معامل ثبات الاختبار، وعليه تقبل الفرضية الصفرية الثانية.

الجدول (٢)

قيمة (Z) المحسوبة والجدولية لدلالة الفرق بين معاملي ثبات أسئلة التكميل وأسئلة الاختيار من متعدد

نوع الأسئلة	معامل الثبات R	القيمة المعيارية ZR	قيمة Z		مستوى الدلالة
			المحسوبة	الجدولية	
التكميل	٠.٨٠١	١.١٠١٨	١٠٦٦	١.٩٦	لا يوجد فرق عند مستوى دلالة (٠.٠٥)
الاختيار من متعدد	٠.٧٣٥	٠.٩٤٠			

٣. حساسية الاختبار:

بلغ مؤشر حساسية الاختبار بأسئلة التكميل (٢.٠٠٧) وهو بدلالة إحصائية عند مستوى (٠.٠٥) لكونه أكبر من القيمة الزائية الجدولية (١.٩٦) لذا يعطى هذا النمط للمفاضلة في حساسية الاختبار درجة واحدة، في حين بلغ مؤشر حساسية الاختبار بأسئلة الاختيار من متعدد (١.٦٦٦) وهو ليس بدلالة إحصائية عند مستوى (٠.٠٥) لكونه أصغر من القيمة الزائية الجدولية (١.٩٦) مما يعطي هذا النمط درجة (صفر) في المفاضلة، وبما إن هناك اختلافاً في مستوى دلالة مؤشر حساسية الاختبار بين النمطين، لذا ترفض الفرضية الصفرية الثالثة من فرضيات البحث وتقبل الفرضية البديلة.

٤. معاملات تمييز الفقرات:

تراوحت معاملات تمييز فقرات أسئلة التكميل ما بين (٠.٧٩- ٠.٣٨) وبمتوسط مقداره (٠.٤٩٢)، وانحراف معياري (٠.٠٨٨٦) في حين تراوحت معاملات تمييز أسئلة الاختيار من متعدد ما بين (٠.٧٥- ٠.٣٠)، وبمتوسط مقداره (٠.٤٣٣)، وانحراف معياري (٠.١٦٢).

ولمعرفة دلالة الفرق بين هذين المتوسطين، استخدم الباحثان اختبار (ت) لعينتين مستقلتين t-test فاتضح أن الفرق كان بدلالة إحصائية عند مستوى (٠.٠٥) إذ كانت القيمة التائية المحسوبة (٢.٣٣٢) أكبر من القيمة الجدولية (١.٩٩٩)،

والجدول (٣) يوضح ذلك، وعليه ترفض الفرضية الصفرية الرابعة وتقبل البديلة، لذا تعطى أسئلة التكميل درجة واحدة المفاضلة، وتعطى أسئلة الاختيار من متعدد درجة (صفر) في المفاضلة.

الجدول (٣)

قيمة (ت) لدلالة الفرق في متوسط معاملات التمييز بين أسئلة التكميل وأسئلة الاختيار من متعدد

نوع الأسئلة	المتوسط الحسابي	الانحراف المعياري	قيمة (ت)		مستوى الدلالة
			المحسوبة	الجدولية	
التكميل	٠.٤٩٢	٠.٠٨٨٦	٢.٣٣٢	١.٩٩٩	دالة عند (٠.٠٥) لصالح أسئلة التكميل
الاختيار من متعدد	٠.٤٣٣	٠.١٠٦٢			

٥. معاملات صدق الفقرات:

تراوحت معاملات صدق فقرات أسئلة التكميل ما بين (٠.٧١ – ٠.٣٥)، وبمتوسط مقداره (٠.٥٠٢)، وانحراف معياري (٠.٠٨٦٧)، في حين تراوحت معاملات صدق فقرات الاختيار من متعدد ما بين (٠.٦٣– ٠.٢٧)، وبمتوسط مقداره (٠.٤٧٨)، وانحراف معياري (٠.٠٨١٧).

ولمعرفة دلالة الفرق بين هذين المتوسطين استخدم الباحثان اختبار (ت) t-test لعينتين مستقلتين، فاتضح أن الفرق لم يكن بدلالة إحصائية عند مستوى (٠.٠٥)، إذ كانت القيمة المحسوبة (١.١٠١) أصغر من القيمة الجدولية (١.٩٩٩) بدرجة حرية (٥٨)، والجدول (٤) يوضح ذلك، وهذه النتيجة تؤشر عدم وجود فرق في معاملات صدق الفقرات بين أسئلة التكميل وأسئلة الاختيار من متعدد، لذا تقبل الفرضية الصفرية الخامسة من فرضيات البحث، وعليه يعطى كل نمط درجة (صفر) للمفاضلة في صدق الفقرات.

الجدول (٤)

قيمة (ت) لدلالة الفرق في متوسط معاملات صدق الفقرات بين أسئلة التكميل وأسئلة الاختيار من متعدد

نوع الأسئلة	المتوسط الحسابي	الانحراف المعياري	قيمة (ت)		مستوى الدلالة
			المحسوبة	الجدولية	
التكميل	٠.٥٠٢	٠.٠٨٦٧	١.١٠١	١.٩٩٩	لا يوجد فرق عند (٠.٠٥)
الاختيار من متعدد	٠.٤٧٨	٠.٠٨١٧			

٦. معاملات صعوبة الفقرات:

تراوحت معاملات صعوبة فقرات أسئلة التكميل ما بين (٠.٦١-٠.٣٤) وبمتوسط حسابي مقداره (٠.٤٨٠) وانحراف معياري (٠.٠٧٧٦)، بينما تراوحت معاملات صعوبة الفقرات في أسئلة الاختيار من متعدد ما بين (٠.٥٥- ٠.٢٣)، وبمتوسط حسابي مقداره (٠.٤٥٧)، وانحراف معياري (٠.٧٨٦).

ويبدو أن أسئلة التكميل أكثر صعوبة من أسئلة الاختيار من متعدد، لأن متوسط معاملات صعوبتها أكبر من متوسط معاملات صعوبة أسئلة الاختيار من متعدد، وقد يعود ذلك إلى أن أسئلة التكميل تعتمد على الاستدعاء Recall، وهو أكثر صعوبة من التعرف Recognition الذي تعتمد عليه أسئلة الاختيار من متعدد (Halonen & Santrok, 1996, P. 516).

لأن صعوبة لاستدعاء تكمن في القدرة على استحضار الماضي، أما التعرف فهو تذكر أمور معروضة أمام المجيب (الهاشمي، ١٩٩٢، ص٩٣).

ولمعرفة دلالة الفرق بين متوسط معاملات صعوبة فقرات أسئلة التكميل وفقرات أسئلة الاختيار من متعدد كل على انفراد والمتوسط المعياري لمعاملات الصعوبة باستخدام اختبار (ت) لعينة واحدة(Glass & Stanley, 1970, P.293) اتضح ما يأتي:

أ- لم يكن الفرق بين متوسط معاملات صعوبة فقرات أسئلة التكميل ومتوسط الصعوبة المعياري (٠.٥٠) بدلالة إحصائية عند مستوى (٠.٠٥)، إذ كانت القيمة التائية المحسوبة أصغر من القيمة الجدولية. والجدول (٥) يوضح ذلك، لذا تقبل الفرضية الصفرية السادسة، وعليه تعطى أسئلة التكميل للمفاضلة في معاملات صعوبة الفقرات (٣) درجات.

ب- كان الفرق بين متوسط معاملات صعوبة فقرات أسئلة الاختيار من متعدد والمتوسط المعياري للصعوبة البالغ (٠.٥٠) بدلالة إحصائية عند مستوى (٠.٠١) إذ كانت القيمة التائية المحسوبة أكبر من القيمة الجدولية، والجدول (٥) يوضح ذلك، لذا ترفض الفرضية الصفرية وتقبل البديلة عند مستوى دلالة (٠.٠١). وعليه تعطى أسئلة الاختيار من متعدد للمفاضلة في معاملات صعوبة الفقرات درجة واحدة على وفق المعيار المعتمد في هذا البحث.

الجدول (٥)

قيمة (ت) لدلالة الفرق في متوسط صعوبة فقرات أسئلة التكميل وأسئلة الاختيار من متعدد والمتوسط العام للصعوبة (٠.٥)

نوع الأسئلة	المتوسط الحسابي	الانحراف المعياري	قيمة (ت)		مستوى الدلالة
			المحسوبة	الجدولية	
التكميل	٤٨٠	٠.٠٧٨٦	١.٤١١	٢.٠٤٥	غير دال عند مستوى (٠.٠٥)
الاختيار من متعدد	٤٥١	٠.٠٧٧٦	٣.٤١٥	٢.٧٥٦	دال عند مستوى (٠.٠١)

المفاضلة العامة بين نمطي الأسئلة:

في ضوء نتائج البحث المتعلقة بالمفاضلة بين أسئلة التكميل وأسئلة الاختيار من متعدد لقياس التذكر المعرفي في ضوء الخصائص السيكومترية للاختبار ولفقراته كانت درجات المفاضلة لهذه الخصائص كما هي في الجدول (٦).

الجدول (٦)

درجات المفاضلة بين أسئلة التكميل وأسئلة الاختيار من متعدد في الخصائص السيكومترية للاختبار ولفقراته

نوع الأسئلة	درجات المفاضلة للخصائص السيكومترية						المجموع	النسبة %
	صدق الاختبار	ثبات الاختبار	حساسية الاختبار	تمييز الفقرات	صدق الفقرات	صعوبة الفقرات		
التكميل	١	صفر	١	١	صفر	٣	٦	٣٣.٣٣
الاختيار من متعدد	صفر	صفر	صفر	صفر	صفر	١	١	٥.٥٦

ويتضح من الجدول (٦) إن درجة تفضيل أسئلة التكميل تساوي (٦) درجات من (١٨) درجة التي تمثل أعلى درجة للمفاضلة، وبنسبة (٣٣.٣٣)، في حين جاءت أسئلة الاختيار من متعدد بدرجة واحدة للمفاضلة وبنسبة (٥.٥٦) من الدرجة الكلية للمفاضلة، وقد يعود تفوق أسئلة التكميل على أسئلة الاختبار من متعدد بخصائصها السيكومترية في قياس التذكر المعرفي إلى عدة مبررات منها، أن أسئلة التكميل غير خاضعة للتخمين، كما هو الحال في أسئلة الاختيار من متعدد لأن التخمين أحد مسببات الخطأ المعياري للقياس مما ينعكس ذلك على دقة الخصائص السيكومترية للاختبار (Traub & Hambleton, 1972, P.740). إذ أشارت دراسة (Braswell & Jackson, 1995) إلى أن اختزال التخمين أو إلغائه يُنتج فقرات ذات تمييز أفضل وثبات جيد، وهذا ما توصلت إليه أيضاً دراسة (الحموري، ٢٠٠٠) إذ ذكرت أن عدم وجود احتمال للتخمين في أسئلة التكميل أدى إلى أن تكون هذه الأسئلة مؤشراً أدق على قدرة الطالب من أسئلة الاختبار من متعدد، التي يحتمل أن يكون فيها التخمين، والذي كان بنسبة (٢٥%) في هذا البحث.

وتوصلت دراسة (دوسي، ١٩٩٣) Dossey التي أجريت على نتائج امتحان التقويم الوطني للتقدم التربوي في الرياضيات عام ١٩٩٢ في أمريكا، أن مستويات الأداء على أسئلة الاختيار من متعدد كانت أعلى مما هي عليه في الأسئلة التي تتطلب استجابات منشأة، مما يعني هذا أنها أكثر سهولة وبالتالي لا تقيس القدرات الحقيقية للطلبة (Dossey, 1993, PP.18-20)، وهذا ما توصلت إليه أيضاً دراسة (لوكهلي وآخرون، ١٩٩٤) إذ وجدت أن نمط أسئلة الاستجابة المنشأة مثل أسئلة التكميل والأجوبة القصيرة، توفر معلومات عن القدرات التحصيلية للطلبة أكثر إلى حد ما من تلك التي توفرها أسئلة الاختيار من متعدد (Lukhele, & et.al., PP.238-

250)، ووجدت دراسة (العلي، ١٩٨٩) إن الصدق التلازمي للاختبار في مستوى التذكر وكان في أسئلة الإجابة القصيرة أفضل مما هو عليه في أسئلة الاختيار من متعدد، وتوصلت دراسة (دبك، ١٩٩٨) إلى أن معامل الصدق المرتبط بمحك كان أفضل في أسئلة التكميل مما هو عليه في أسئلة الاختيار من متعدد، وأظهرت نتائج دراسة (الحموري، ٢٠٠٠) إلى أن أسئلة التكميل تتسم بصدق تنبؤي وصدق عاملي أفضل إلى حد ما من أسئلة الاختيار من متعدد فضلاً عن كونها أكثر ثباتاً وتمييزاً.

الاستنتاجات

١. يمكن استخدام أسئلة التكميل وأسئلة الاختيار من متعدد لقياس التذكر المعرفي لدى الطلبة لكونهما حصلا على خصائص سيكومترية مقبولة سواء للاختبار أو لفقراته بشكل عام، إلا أن أسئلة التكميل كانت أفضل إلى حد ما في قياس التذكر المعرفي من أسئلة الاختيار من متعدد، لأن خصائصها السيكومترية التي تؤشر دقة الاختبار في قياس ما أعد لقياسه، كانت هي الأفضل في أسئلة التكميل.

٢. إن الإجابة عن أسئلة الاختيار من متعدد أكثر سهولة مما هي عليه في أسئلة التكميل، إذ حصل المجيبون فيها درجات أعلى من أسئلة التكميل.

٣. إن أسئلة التكميل لها قدرة على التمييز بين الطلبة في التذكر المعرفي أكثر مما هي عليه في أسئلة الاختيار من متعدد.

التوصيات والمقترحات:

في ضوء نتائج البحث واستنتاجاته يوصي الباحثان بما يأتي:

١. تدريب المدرسين والمدرسات على كيفية إعداد أسئلة الاختيار من متعدد وأسئلة التكميل في قياس التذكر المعرفي.

٢. توجيه المدرسين والمدرسات إلى استخدام أسئلة التكميل بدلاً من أسئلة الاختيار من متعدد في قياس التذكر المعرفي لكونها أكثر دقة وتتسم بخصائص سيكومترية أفضل من أسئلة الاختيار من متعدد.

٣. التأكيد على أهمية مادة القياس والتقويم بكليات التربية ومعاهد إعداد المعلمين وإعطائها أهمية أكبر مما هي عليه الآن، إذ ينبغي أن تأخذ ساعات دراسية أكثر وتدريب عملي مع الجانب النظري.
٤. يقترح الباحثان إجراء دراسات لاحقة مماثلة للبحث الحالي على مراحل دراسية أخرى وبمواد دراسية متنوعة، ودراسة أخرى للمفاضلة بين أسئلة الإجابات القصيرة وأسئلة التكميل أو المقالية أو الموضوعية في قياس مستويات تصنيف بلوم للمجال المعرفي.

الملحق (١)
فقرات اختبار الاختيار من متعدد بصيغته النهائية

عزيزتي الطالبة..

عزيزي الطالب..

لأغراض البحث العلمي ترجو الباحثة الإجابة عن الفقرات الآتية بكل دقة والتي هي عبارة عن أسئلة اختيار من متعدد، وذلك بوضع دائرة ◯ الحرف الدال على الإجابة الصحيحة من بين بدائل الإجابة الأربعة لكل فقرة، علماً أن الإجابة لا تستخدم للنجاح أو الرسوب بل لأغراض البحث العلمي.

مع جزيل الشكر والتقدير...

الباحثان

الأسئلة

ضعي دائرة ◯ حول حرف الإجابة التي تختاريها في ورقة الإجابة

١.	يمكن تلخيص مفهوم التقويم بأنه عملية توفير معلومات موضوعية لأجل: أ- تقويم الآخرين.ب- اختبار الطلبة.ج- إصدار الأحكام. د- تطوير التعليم.
٢.	تسمي عملية التكميم أو التعبير بلغة كمية أو حسابية عن الصفات أو الظواهر بـ: أ- التقويم. ب- التعلم. ج- القياس. د- التحصيل.
٣.	إن القياس الذي يكون مطلقاً ومباشراً يكون في الظواهر: أ- النفسية. ب- التربوية. ج- الاجتماعية. د- الطبيعية.
٤.	يرى (كرونباك) و (سكرفين) إن للتقويم دورين أساسيين، الأول دور تتبعي أو بنائي، والثاني دور: أ- كمي. ب- بدائي. ج- نهائي. د- قياسي.
٥.	تتكون العملية التعليمية من ثلاثة عناصر مهمة هي: أ- المعلم والطالب والمدرسة. ب- المعلم والطالب والمنهج. ج- المدير والمعلم والمنهج. د- الطالب والمعلم والكتاب.
٦.	يسمى تقويم الفرد لنفسه بالتقويم: أ- العلمي. ب- الكمي. ج- الذاتي. د- الموضوعي.
٧.	يسمى القصد أو الغاية التي تعمل التربية على الوصول إليها في سلوك الطلبة بالهدف: أ- التربوي. ب- السلوكي. ج- القصدي. د- العملي.
٨.	تشتق الأهداف التربوية العامة من: أ- فلسفة المدرسة. ب- الكتب والمدرسية.

	ج- فلسفة البلد. د- الأهداف السلوكية.
٩.	يسمى مجال الأهداف السلوكية الذي يهتم بالمهارات العقلية والمعلومات بالمجال: أ- المعرفي. ب- السلوكي. ج- الوجداني. د- النفسي.
١٠.	يسمى مجال الأهداف الذي يرمي على تنمية القدرات على استعمال المعارف وظيفياً في مواقف جديدة بمستوى: أ- الفهم. ب- التحليل. ج- التطبيق. د- التقويم.
١١.	الاتجاه نحو أي شيء قد يكون إيجابياً أو سلبياً أو: أ- عالياً. ب- واطئاً. ج- متغيراً. د- محايداً.
١٢.	تسمى الخبرات التربوية التي تقدمها المدرسة لتلامذتها سواء في داخلها أو خارجها بـ: أ- التعليم. ب- المنهج. ج- التربية. د- التقويم.
١٣.	يحتوي الكتاب المدرسي المادة الدراسية التي يحددها: أ- المعلم. ب- المنهج. ج- المدير. د- الكتاب.
١٤.	من أغراض الاختبارات هو قياس: أ- تحصيل التلاميذ. ب- كفاءة المعلمين. ج- قدرات الضعفاء. د- اتجاهات التلاميذ.
١٥.	يمكن أن تستخدم الاختبارات المدرسية في تقويم الأهداف: أ- المدرسية. ب- التربوية. ج- السلوكية. د- العقلية.
١٦.	الاختبارات التحصيلية نوعان، الأول هي التي يعدها المعلمون بأنفسهم والثاني هي الاختبارات: أ- الموضوعية. ب- المقننة. ج- الوزارية. د- المدرسية.
١٧.	تهدف الاختبارات الشفوية إلى التعرف على مدى إتقان التلاميذ للمادة بمعزل عن قدرات الكتابة و: أ- القراءة. ب- الحساب. ج- التعبير. د- الذكاء.

١٨.	تكون احتمالات الغش أقل في الاختبارات: أ- الموضوعية. ب- القصيرة. ج- التكميلية. د- المقالية.
١٩.	إن وجود أسئلة ترك في الامتحان قد تشجع الطالب على: أ- عدم الخوف من الامتحان. ب- عدم الإجابة عن جميع الأسئلة. ج- ترك قسم من المادة دون دراستها. د- قراءة المادة بدقة ورغبة.
٢٠.	ينبغي تحديد الإجابة النموذجية في الأسئلة: أ- الموضوعية. ب- المقالية. ج- الشفوية. د- التحريرية.
٢١.	أن أفضل الاختبارات التي يمكن أن تغطي أكبر كمية ممكنة من المادة الدراسية هي الاختبارات: أ- المقالية. ب- القصيرة. ج- الموضوعية. د- الشفوية.
٢٢.	يفضل جعل مكان الفراغ في الفقرة بأسئلة التكميل (إكمال الفراغات) في: أ- مقدمة الفقرة. ب- نهاية الفقرة. ج- وسط الفقرة. د- أي مكان في الفقرة.
٢٣.	ينبغي أن لا تقيس العبارات في أسئلة الصواب والخطأ أكثر من: أ- عبارة واحدة. ب- فكرة واحدة. ج- عبارتين. د- فكرتين.
٢٤.	من أهم شروط الاختبار المقنن الجيد هو أن يكون: أ- ثابتاً. ب- جيداً. ج- شاملاً. د- صادقاً.
٢٥.	يسمى المفهوم الذي يدل على أن الاختيار قادراً على قياس ما صمم لأجل قياسه بـ: أ- الصدق. ب- القياس. ج- الثبات. د- الموضوعية.
٢٦.	في اختبار الاختيار من المتعدد ينبغي أن تكون البدائل الخاطئة: أ- غامضة. ب- قليلة. ج- جذابة. د- واضحة.

٢٧.	إن الأسئلة التي تكون إجاباتها سهلة التصحيح هي الأسئلة: أ- المقالية. ب- الموضوعية. ج- القصيرة. د- العلمية.
٢٨.	يسمى المفهوم الذي يؤشر أن الاختيار لا تتغير درجاته في حالة تطبيقه مرتين على المجموعة نفسها بـ: أ- الصدق. ب- التقني. ج- الاختبار. د- الثبات.
٢٩.	إن مجال الأهداف السلوكية الذي يهتم بالمهارات واستعمالها وظيفياً بالمجال: أ- النفس- وظيفي. ب- النفس – حركي. ج- النفس – سلوكي. د- النفس – عقلي.
٣٠.	تسمى قدرة الفرد على أداء شيء معين بسرعة وإتقان وفهم وبأقل جهد وتكاليف بـ: أ- المهارة. ب- الفهم. ج- الاتقان. د- الدقة.

الملحق (٢)
فقرات اختبار التكميل بصيغته النهائية

عزيزتي الطالبة..
عزيزي الطالب..
تحية طيبة..

لأغراض البحث العلمي يرجو الباحثان معاونتكم في الإجابة عن الفقرات التالية التي هي عبارة عن أسئلة (التكميل أو إملاء الفراغات) بكل دقة وذلك بوضع ما يناسب الفقرة في الفراغ من كلمة أو أكثر بحسب طول الفراغ المحدد، علماً أن الإجابة لا تستخدم للنجاح والرسوب ولا لأي غرض سوى للبحث العلمي، لذا كلما تكون الإجابة دقيقة فإنكم قدمتم خدمة للبحث العلمي، ونرجو أن تكون الإجابة بسرعة قدر الإمكان.

مع جزيل الشكر ...

الباحثان

الفقرات

أكملي الفراغات بما يناسبها:

١.	يسمى القصد أو الغاية التي تعمل التربية على الوصول إليها في سلوك الطلبة بالهدف .....
٢.	يرى (كرونباك) و (سكرفين) إن للتقويم دورين أساسيين، الأول دور تتبعي أو بنائي، والثاني دور .....
٣.	تسمي عملية التكميم أو التعبير بلغة كمية أو حسابية عن الصفات أو الظواهر بـ.....
٤.	يمكن تلخيص مفهوم التقويم بأنه عملية توفير معلومات موضوعية لأجل.....
٥.	تتكون العملية التعليمية من ثلاثة عناصر مهمة هي..... و..... و .....
٦.	يسمى تقويم الفرد لنفسه بالتقويم.....
٧.	إن القياس الذي يكون مطلقاً ومباشراً يكون في الظواهر .....
٨.	تسمى الخبرات التربوية التي تقدمها المدرسة لتلامذتها سواء في داخلها أو خارجها بـ.....
٩.	الاتجاه نحو أي شيء قد يكون إيجابياً أو سلبياً أو .....
١٠.	يسمى مجال الأهداف الذي يرمي على تنمية القدرات على استعمال المعارف وظيفياً في مواقف جديدة بمستوى .....
١١.	تشتق الأهداف التربوية العامة من .....
١٢.	يسمى مجال الأهداف السلوكية الذي يهتم بالمهارات العقلية والمعلومات بالمجال .....
١٣.	يحتوي الكتاب المدرسي المادة الدراسية التي يحددها .....
١٤.	يمكن أن تستخدم الاختبارات المدرسية في تقويم الأهداف .....
١٥.	من أغراض الاختبارات هو قياس .....
١٦.	الاختبارات التحصيلية نوعان، الأول هي التي يعدها المعلمون بأنفسهم والثاني هي

	الاختبارات .....
١٧.	ينبغي تحديد الإجابة النموذجية في الأسئلة .....
١٨.	إن وجود أسئلة ترك في الامتحان قد تشجع الطالب على .....
١٩.	تكون احتمالات الغش أقل في الاختبارات .....
٢٠.	تهدف الاختبارات الشفوية إلى التعرف على مدى إتقان التلاميذ للمادة بمعزل عن قدرات الكتابة و.....
٢١.	أن أفضل الاختبارات التي يمكن أن تغطي أكبر كمية ممكنة من المادة الدراسية هي الاختبارات.....
٢٢.	يفضل جعل مكان الفراغ في الفقرة بأسئلة التكميل (إكمال الفراغات) في.....
٢٣.	ينبغي أن لا تقيس العبارات في أسئلة الصواب والخطأ أكثر من.....
٢٤.	من أهم شروط الاختبار المقنن الجيد هو أن يكون.....
٢٥.	يسمى المفهوم الذي يدل على أن الاختيار قادراً على قياس ما صمم لأجل قياسه بـ.....
٢٦.	تسمى قدرة الفرد على أداء شيء معين بسرعة وإتقان وفهم وبأقل جهد وتكاليف بـ .....
٢٧.	إن الأسئلة التي تكون إجاباتها سهلة التصحيح هي الأسئلة .....
٢٨.	في اختبار الاختيار من المتعدد ينبغي أن تكون البدائل الخاطئة.....
٢٩.	يسمى المفهوم الذي يؤشر أن الاختيار لا تتغير درجاته في حالة تطبيقه مرتين على المجموعة نفسها بـ ..... .
٣٠.	إن مجال الأهداف السلوكية الذي يهتم بالمهارات واستعمالها وظيفياً بالمجال.....

المصادر

– المصادر العربية:

١. إسماعيل، عزت سيد. (١٩٨٦): علم النفس التجريبي، الكويت، وكالة المطبوعات.

٢. الحموري، هند عبد المجيد. (٢٠٠٠): أثر اختلاف نمط الاختبار حسب نوعية الاستجابة المطلوبة منشأة أو مختارة في صدقه العاملي والتنبوي، مجلة كلية التربية، جامعة الإمارات العربية، العدد (١٦).

٣. ديك، أحمد رشيد. (١٩٩٨): أثر شكل الفقرات وعدد البدائل في الخصائص السيكومترية للاختبار وفقراته (رسالة ماجستير غير منشورة)، جامعة اليرموك، الأردن.

٤. سلمان، خمائل مهدي صالح. (٢٠٠٧): المفاضلة بين أنماط صياغة فقرات المواقف اللفظية في ضوء الخصائص السيكومترية للمقاييس النفسية (أطروحة دكتوراه غير منشورة)، جامعة بغداد، كلية التربية/ ابن رشد.

٥. عبد الرحمن، سعد. (١٩٩٨): القياس النفسي، الكويت، مكتبة الفلاح.

٦. عدس، عبد الرحمن ومحي الدين توق. (١٩٨٤): أساسيات علم النفس التربوي، الأردن، الجامعة الأردنية.

٧. علام، صلاح الدين. (٢٠١٠): القياس والتقويم التربوي في العملية التدريسية.

٨. العلي، محمد محمود. (١٩٨٩): المقارنة بين اختبارات الاختيار من عدة بدائل واختبارات الإجابة القصيرة واختبارات المزيج في النوعين معاً على بعض الخصائص السيكومترية للاختبار (رسالة ماجستير غير منشورة)، جامعة اليرموك، الأردن.

٩. عودة، أحمد سليمان. (١٩٩٨): القياس والتقويم في العملية التدريسية، عمان.

١٠. العياصرة، وليد رفيق. (٢٠١١): التعليم والتعلم وعلم النفس التربوي، عمان، دار أسامة للنشر والتوزيع.

١١. العتوم، عدنان. (٢٠٠٤): علم النفس المعرفي، النظرية والتطبيق، عمان، دار المسيرة.

١٢. الهاشمي، عبد الحميد محمد. (١٩٩٢): أصول علم النفس العام، جدة، دار الشروق.

١٣. يوسف، سليمان عبد الواحد. (٢٠١١): الفروق الفردية في العمليات العقلية المعرفية، عمان، دار المسيرة.

14. Anastasi, A. **Psychological Testing,** New York, Prentice-Hall.

15. Badger, E. (1990): Using different spectacles to look at student achievement implications for theory & pactice, **ERIC Document production service**, No. Eb. 5415761.

16. Barnett, F. D. & Nagy, P. (1995): A comparison of undergraduate test Response strategies for multiple- choice and constructed- Response Questions, Alberta Journal of Educational Research, (41) (1), 18-35.

17. Braswell, S. & Sackon, C. (1995): **An introduction of a new free – Response item type in mathematics**, paper presented at the Annual meeting of the national council on measurement id education, San Francisco, CA. April, 19-21.

18. Bridgeman, B. (1992): A comparison of quantitative questions in open- ended and multiple- choice formats, **Journal of Educational measurement**, 29 (3), pp. 253-271.

19. Bloom, B. et.al. (1981): **Evaluation to improve instruction**, N. Y. MC-Graw Hill.

20. Dick, W. & Hagert, Y. (1971): **Topicsin measurement**, New York, Mc-Graw-Hill.
21. Dossy, J. (1993): Can students do mathematical problem solving? Results from constructed- Response questions in NAEP'S 1992 National Assessment **Educational Testing service**.
22. Ebel, R. L. (1972): **Essentials of Educational Measu**rement, New Jersey, Prentice- Hall Inc.
23. Fox, D. T. (1969): **Statistical Analysis psychology and Education,** New York, Holt.
24. Glais, G. V. & Stanley, J. C. (1970): **Statistical methods in Educational and psychology**, New Jersey. Prentic- Hall, Inc.
25. Kennedy P. & Walstad, W. (1997): Combining Multiple-Choice and constructed- Response test scores: An Economist's view. **Applied measurement in education**, 10 (4), 359-376.
26. Lukhele, R. et. al. (1994): On the relative value of multiple-choice constructed- Response and Examiuee- Selected items on two achievement tests, **Journal of educational psychology**, (31), (3), pp. 224-250.
27. Pedhazur, E. J. (1982): **Multiple Regression in Behavioral Research**, New York: Holt, Kinehart and Winston.
28. Stiggins, R. 919940: **Student- Centered classroom Assessment**, N. Y. Macmilan.
29. Ward, A. & Ward, M. (1999): **Assessment in the classroom**, Wadswarth.
30. Traub, R. (1993): **On the equivalence of the traits assessed by multiple choice and constructed- response tests in R. S. Bennett & W. Ward (Sds) construction versus choice in cognitive measurement**: Issues in constructed response, performance testing, pp. 29-44.

31. Troub, R. & Macruray, K. (1992): **Multiple- choice vs. free-response in the testing of scholastic achievement**. In k. ingenkamp & R. gager (eds) tests and trends. Weinheim under Basel, Germany: Beltz veriag.