

دراسة مقارنة لأثر اختبارات الاختيار من متعدد واختبارات الإجابة القصيرة في الاحتفاظ بنواتج التعليم في المجال المعرفي

الدكتور . أحمد سليمان عودة
كلية التربية - جامعة اليرموك

الملخص

هدفت هذه الدراسة إلى معرفة أثر الاختبارات في احتفاظ طلاب الصف الأول الثانوي بمادة قواعد اللغة العربية، ومقارنة أثر نوعين من الاختبارات هما: اختبارات الاختيار من متعدد، واختبارات الإجابة القصيرة في هذا الاحتفاظ، على مستوى المعرفة، وعلى مستوى القدرات الذكائية (كل ما فوق المعرفة في تصنيف بلوم) وعلى المجال المعرفي (المستويين معا).

تكونت عينة الدراسة من (٣٠١) طالب من طلاب الصف الأول الثانوي الأكاديمي، موزعين على ثلاث عينات، بواقع ثلاث شعب في كل مدرسة، يدرسها المعلم نفسه. وقد قسمت العينة إلى ثلاث مجموعات هي، مجموعة الإجابة القصيرة وقد تعرضت إلى سلسلة من اختبارات الإجابة القصيرة، ومجموعة الاختيار من متعدد، وقد تعرضت إلى سلسلة من اختبارات الاختيار من متعدد، ومجموعة ضابطة لم تتعرض لأي نوع من الاختبارات الصفية.

أعد اختبار احتفاظ من صورتين، إجابة قصيرة، واختيار من متعدد، بأرومة متشابهة لكل سؤال في الاختبارين. بعد خمسة عشر يوما من انتهاء المعالجات (سلسلة الاختبارات) أعطي نصف طلاب كل مجموعة اختبار الاختيار من متعدد، وأعطى النصف الآخر اختبار الإجابة القصيرة. وكان الاختباران يقيسان التحصيل في مستوى المعرفة والقدرات الذكائية. وعدت علامة كل طالب على المستويين معا تحصيلًا في المجال المعرفي.

أظهرت نتائج تحليل التباين والمقارنات البعدية تفوق مجموعتي الإجابة القصيرة، والاختبار من متعدد على المجموعة الضابطة على الاختبارات الثلاث بشكل عام. كما أظهرت النتائج تفوق مجموعة الإجابة القصيرة على مجموعة الاختيار من متعدد، عندما قيس الاحتفاظ باختبار مكون من مزيج من

النوعين معا من الاختبارات ، وعلى المستويات الثلاثة . كما تفوقت عندما قيس الاحتفاظ باختبار الإجابة القصيرة على مستوى المعرفة ، وعلى مستوى المجال المعرفي . وتعادل أدائهما على اختبار الاحتفاظ من نوع الاختيار من متعدد ، وعلى المستويات الثلاثة . كما تساوى أدائهما على مستوى القدرات الذكائية من اختبار الإجابة القصيرة .

تشير هذه النتائج إلى أن الاختبارات يمكن أن تكون استراتيجية في التعليم ، كما تدل على قوة الأثر التعليمي لاختبارات الإجابة القصيرة مقارنة بالاختيار من متعدد في المواقف التقييمية التي يكون فيها المقوم بموضع الاختيار .

خلفية الدراسة

تستخدم الاختبارات الصفية لأغراض متعددة منها ، تقويم تحصيل الطلاب ، وتصنيفهم أو تشجيعهم وترفعهم من صف إلى آخر ، ولعل هذا هو الشائع في استخدام هذه الاختبارات . ولكن ألا يمكن أن يكون لهذه الاختبارات أغراض أخرى قد تتعدى في أهميتها الأغراض التقليدية . وبالتحديد ألا يمكن أن تؤثر في التعلم بشكل ما؟ وهل تختلف الاختبارات باختلاف أشكال الفقرات في قوة تأثيرها؟

هناك إجابات متباينة ليست كافية عن هذه التساؤلات فقد ذكر أدكنز (Adkins, 1974) أن الاختبار يشكل موقفا تعليميا ، وأن المعلم الذي يخفق في الاستفادة من هذا الموقف يكون قد ضيع واحدة من الطرق المهمة في التعلم . كما ذكر جرونلند (Gronlund, 1977) وبراون (Brown, 1983) أن الاختبارات تسهل عملية التعلم ، وتزيد احتفاظ الطالب بما يتعلمه ، وبالتالي توفر إمكانية انتقال أثر التعلم .

أجري العديد من الدراسات حول أثر تعدد الاختبارات على تخزين المادة المتعلمة والاحتفاظ بها ، مقارنة بالدراسة الزائدة أو إعطاء التمرينات عند قياس نواتج التعلم الإجمالية ، وأشارت نتائج بعضها مثل : ألن وآخرين (Allen, et. al, 1969) ، هوجان وكنتش (Hogan and Kintesch, 1971) ، لابورت وفوس (Laporte and Voss, 1975) ، جيه وجلاجار (Gay and Gallagher, 1976) ألن وآخرين (Allen, et. al., 1969) ، دكستل (Duchastel., 1979) ، دكستل ومانجستر (Duchastel., and Nungester, 1982) ، هالبن (Haplin and Halpin, 1982) إلى فروق دالة إحصائية لصالح تعدد الاختبارات ، ومفسرين ذلك بأن الاختبار محاولة تعلم إضافي ، وأنه يعمل على تثبيت نواتج التعلم بسبب المراجعة العقلية النشطة للمعلومات من الذاكرة في موقف الاختبار . إلا أن نتائج دراسات أخرى مثل : هالبن وهارنجتون (Haplin and Nungester, 1982)،

(Harington, 1982) أظهرت تعادل أثر الدراسة الزائدة مع أثر تعدد الاختبارات، إلا في الحالات التي تكون فيها الفقرات التي ترد في سلسلة الاختبارات مكررة في اختبار الاحتفاظ بنواتج التعلم.

وهكذا يظهر جليا أنه يمكن استخدام الاختبارات كاستراتيجية من استراتيجيات التعلم تعمل على تثبيت المادة المتعلمة وبالتالي الاحتفاظ بها. إلا أن الاختبارات أنواع: منها الاختبارات ذات الإجابة المصوغة، كاختبارات التكملة والإجابة القصيرة، واختبارات المقال المفتوحة والمحددة، ومنها الاختبارات ذات الإجابة المنتقاة، كاختبارات المطابقة، والاختيار من بدليلين، والاختيار من متعدد. ومع تعدد أنواع الاختبارات تعددت ميول المعلمين إلى استخدام نوع دون آخر على مختلف المستويات التعليمية، حيث تتحكم عدة عوامل في اختيار نوع دون آخر، ولعل أهمها: طبيعة المادة الدراسية، مستوى العبارات الهدافية، الغرض من التقويم، أعمار الطلاب ومستوياتهم العقلية، مدى تأثير الخصائص السيكموترية للاختبار بطريقة التصحيح. وهذا يعني أن اختيار نوع دون آخر ليس مزاجيا أو عشوائيا.

ترى فإذا توازن تأثير العوامل السابقة في نوعين من الاختبارات وأمكن إجراء تحويل متبادل للفقرات مع تأدية هذه الفقرات للغرض نفسه، وتحققت الدرجة نفسها من موضوعية التصحيح، فهل يتساوى أثر النوعين في احتفاظ الطلاب بنواتج التعلم نفسها؟

حاولت هذه الدراسة الإجابة عليه حيث قارنت بين اختبارات الإجابة القصيرة واختبارات الاختيار من متعدد من حيث أثرهما في احتفاظ الطلاب بنواتج التعلم. وقد اختير لهذا الغرض مادة قواعد اللغة العربية للصف الأول الثانوي. حيث تلعب القواعد دورا مهما في استخدامات اللغة، وتوفر هذه المادة إمكانية تحويل الفقرات من نوع إلى آخر، كما أن المستوى العمري للطلاب لا يستدعي التحفظ على استخدام نوع دون آخر، أما عن أسباب اختيار النوعين من الاختبارات دون غيرهما فكانت:

١ - يعد هذان النوعان أكثر الاختبارات شيوعا، على مختلف المستويات التعليمية (Gay, 1980).

٢ - انتهاء كل منهما إلى نوع مغاير من الاختبارات. فالاختيار من متعدد ينتمي إلى الاختبارات ذات الإجابة المنتقاة. أما الإجابة القصيرة فينتمي إلى الاختبارات ذات الإجابة المصوغة (Hopkins and Antes, 1978).

٣ - تضارب نتائج الدراسات حول أفضلية أي منها كاستراتيجية تعلم، وكذلك حول مستوى صعوبة كل منها. فقد أشارت بعض الدراسات مثل : (Duchstel, 1981; Sax and Collet, 1968; Halpin & Halpin, 1982). إلى أن أداء الطلاب الذين تعرضوا لاختبارات الاختيار من متعدد كان أعلى إحصائياً من أداء الذين تعرضوا لاختبارات الإجابة القصيرة على اختبار من نوع الاختيار من متعدد، ولقياس الاحتفاظ. كما أظهرت نتائج دراسات أخرى تفوق أثر الاجابات القصيرة أو عدم وجود فروق لصالح أي منها، وذلك حسب نوع فقرات الاختبار التي تقيس الاحتفاظ. فقد وجد جيه (Gay, 1980) أنه لا يوجد فرق بين أداء الطلاب الذين تعرضوا للنوعين على اختبار الاحتفاظ من نوع الاختيار من متعدد أو من النوعين معا. بينما وجد أن الفرق لصالح الإجابة القصيرة عندما كان اختبار الاحتفاظ من نوع الإجابة القصيرة. أما فوس (Voss, 1974) فكان قد وجد فرقاً لصالح مجموعة الإجابة القصيرة عندما كان اختبار الاحتفاظ من النوعين معا.

انصبت التفسيرات في حالة وجود الفرق لصالح أي منها أو في حالة غياب الفرق على مستوى صعوبة هذين النوعين من الاختبارات. فقد ذكر ساكس وريدر (Sax, and Reader, 1964) أن الاختبار الأصعب يحقق تعلماً أفضل. ولعل التضارب في نتائج هذه الدراسات يعود إلى عدة أمور تتعلق بعينات الدراسات كالعمر والمستوى العلمي، أو بطبيعة المادة، وبالمعالجات الإحصائية والتصاميم التجريبية.

مشكلة الدراسة وفرضياتها

هدفت هذه الدراسة إلى مقارنة أثر اختبارات الاختيار من متعدد واختبارات الإجابة القصيرة في الاحتفاظ بنواتج التعلم في المجال المعرفي، وبنواتج التعلم في كل من المستويين الرئيسيين للمجال المعرفي حسب تصنيف بلوم، وهما: المعرفة، والقدرات الذكائية (كل ما فوق مستوى المعرفة).

واختير لهذا الغرض ثلاث مجموعات من الطلبة، تعرضت إحداها لسلسلة من اختبارات الاختيار من متعدد، وتعرضت الأخرى لسلسلة من اختبارات الإجابة القصيرة، بينما لم تتعرض الثالثة لأي نوع من الاختبارات، وإنما طلب منها عمل التمرينات الواردة في الوحدة الدراسية. وذلك لفحص الفرضيات التالية (على مستوى الدلالة ٠,٠٥):

١ - لا توجد فروق ذات دلالة إحصائية بين متوسطات علامات طلاب المجموعات الثلاثة

على اختبار الاحتفاظ من نوع «الإجابة القصيرة» لنواتج التعلم في مستوى المعرفة من المجال المعرفي .

٢ - لا توجد فروق ذات دلالة إحصائية بين متوسطات علامات طلاب المجموعات الثلاثة على اختبار الاحتفاظ من نوع «الإجابة القصيرة» لنواتج التعلم في مستوى القدرات الذكائية من المجال المعرفي .

٣ - لا توجد فروق ذات دلالة إحصائية بين متوسطات علامات طلاب المجموعات الثلاثة على اختبار الاحتفاظ من نوع «الإجابة القصيرة» لنواتج التعلم في المجال المعرفي . .

٤ - لا توجد فروق ذات دلالة إحصائية بين متوسطات علامات طلاب المجموعات الثلاثة على اختبار الاحتفاظ من نوع «الاختبار من متعدد» لنواتج التعلم في مستوى المعرفة من المجال المعرفي .

٥ - لا توجد فروق ذات دلالة إحصائية بين متوسطات علامات طلاب المجموعات الثلاثة على اختبار الاحتفاظ من نوع «الاختبار من متعدد» لنواتج التعلم في مستوى القدرات الذكائية من المجال المعرفي .

٦ - لا توجد فروق ذات دلالة إحصائية بين متوسطات علامات طلاب المجموعات الثلاثة على اختبار الاحتفاظ من نوع «الاختبار من متعدد» لنواتج التعلم في المجال المعرفي .

٧ - لا توجد فروق ذات دلالة إحصائية بين متوسطات علامات طلاب المجموعات الثلاثة على اختبار الاحتفاظ من النوعين (الاختبار من متعدد والإجابة القصيرة) لنواتج التعلم في مستوى المعرفة من المجال المعرفي .

٨ - لا توجد فروق ذات دلالة إحصائية بين متوسطات علامات طلاب المجموعات الثلاثة على اختبار الاحتفاظ من النوعين «الاختبار من متعدد والإجابة القصيرة» لنواتج التعلم في مستوى القدرات الذكائية من المجال المعرفي .

٩ - لا توجد فروق ذات دلالة إحصائية بين متوسطات علامات طلاب المجموعات الثلاثة على اختبار الاحتفاظ (المكون من اختبائي الاختبار من متعدد والإجابة القصيرة) . لنواتج التعلم في «المجال المعرفي» .

الطريقة والإجراءات

مجتمع وعينة الدراسة

بلغ عدد أفراد عينة الدراسة ٣٠١ طالب موزعين على تسع شعب للصف الأول الثانوي من ثلاث مدارس اختيرت عشوائيا من ١٧ مدرسة تابعة لمكتب تربية إيدون في

محافظة إربد، وفيها ٩٣٩ طالب من الذكور موزعين في ٣٧ شعبة. وقد توزعت الشعب التسعة في عينة الدراسة بالتساوي على المدارس الثلاثة. ويدرس الشعب الثلاثة في كل مدرسة المعلم نفسه. واعتبرت واحدة منها مجموعة الاختيار من متعدد، وأخرى مجموعة الإجابة القصيرة، والثالثة مجموعة التمرينات، وهي المجموعة الضابطة التي لم تتعرض لأي نوع من الاختبارات.

هذا وقد بلغ عدد الطلاب في مجموعة الاختيار من متعدد ١٠٤ طلاب ومجموعة الإجابة القصيرة ١٠١ طالب، والمجموعة الضابطة ٩٦ طالبا.

أدوات الدراسة

أولا : سلسلة الاختبارات : تم إعداد ثلاثة اختبارات تحصيلية على وحدة الموضوعات من كتاب مذكرة في قواعد اللغة العربية، للصف الأول الثانوي، وكان لكل من هذه الاختبارات صورتان، إحداهما اختيار من متعدد، والثانية من نوع الإجابة القصيرة. وقد اعتمدت تقديرات المختصين في مبحث اللغة العربية على لائحة المواصفات، وصياغة الفقرات، ودرجة توافقها مع الأهداف التدريسية للوحدة، كمؤشرات على صدق المحتوى لهذه الاختبارات.

ثانيا : اختبار الاحتفاظ : وهو اختبار تحصيلي على الوحدة نفسها من نوع الاختيار من متعدد، تم تحويله إلى اختبار من نوع الإجابة القصيرة بإبقاء الأرومات على حالها مع حذف البدائل. وكان عدد الفقرات ٧٥ فقرة. منها (٢٢) فقرة بمستوى المعرفة و (٥٣) فقرة بمستوى القدرات الذكائية.

جرب الاختبار بصورتيه على عينة من مجتمع الدراسة، بلغت ٦٦ طالبا. (٣٤) طالبا للاختيار من متعدد، ٣٢ طالبا للإجابة القصيرة) بغرض تقدير الزمن اللازم للإجابة، وتحليل الفقرات. وقد تم إعادة تجريب الاختبار (بعد تنقيح فقراته وتعديلها على ضوء التجربة الأولى) على عينة من ١٠٠ طالب أعطي نصفها اختيار الاختبار من متعدد، وأعطى النصف الآخر اختبار الإجابة القصيرة. وذلك لتقدير ثبات الاختبار.

وحتى لا تكون طريقة التصحيح عاملاً مؤثراً في ثبات الاختبارين ونتائج الدراسة، فقد روعيت المعرفة الجزئية في الإجابات القصيرة وصححت الإجابة عن كل فقرة بعلامة تتراوح بين صفر وثلاث درجات. وبالمقابل، فإن أمام الطالب الفرصة أن يحصل على علامة بالتخمين العشوائي، أو بالتخمين الذكي، وهذا مالا يتوفر في حالة

الإجابة القصيرة التي لا مجال فيها للشعور به أيضا. ولذلك طلب من الطلبة في مجموعة الاختيار من متعدد، الإجابة على جميع الفقرات، ويسمح له باختيار أكثر من إجابة للفقرة نفسها (وهي أربعة بدائل لكل فقرة)، على أن يبرز البديل المفضل بالنسبة له. وعلى هذا الأساس صححت الإجابات بثلاث طرق هي: الكل، أو العدم (١، صفر)، التصحيح الأثر التخمين، والاختيار المتدرج (Subset selection) وقد دلت نتائج تحليل الفرق بين متوسطين، وتجانس التباين على أن الاختبارين متكافئان عندما صحح الاختيار من متعدد، وبالطريقة الأولى. فقد كان الوسط الحسابي والانحراف المعياري لكل من مجموعتي الاختيار من متعدد، والإجابة القصيرة هي (٤، ٥٥، ١٥، ٥٢)، (١٦، ٨٥، ٥٠، ٧٢)، على التوالي.

بعد اختيار الطريقة المناسبة للتصحيح حسبت معاملات الثبات للاختبارين بالطريقة النصفية، والاتساق الداخلي، وبين الجدول رقم (١) ملخصا لها.

الجدول رقم (١)

معاملات الثبات بالطريقة النصفية، والاتساق الداخلي للاختبارين على مستوى المعرفة، والقدرات الذكائية، وعلى مستوى المجال المعرفي

نوع الاختبار		الاختيار من متعدد		الإجابة القصيرة	
نوع الثبات	معرفة	قدرات ذكاء	مجال معرفي	معرفة	قدرات ذكاء
النصفي (فردى - زوجى) الاتساق الداخلي (ألفا)	٧٣،	٨٦،	٩١،	٨٢،	٩٠،
	٧٤،	٨٦،	٩٠،	٨٧،	٨٩،
					٩٢، ٩٣،

ويلاحظ أن هذه المعاملات مرتفعة نسبيا مقارنة بالاختبارات التحصيلية غير المقتنة. كما حسب معامل الارتباط بين المعرفة والقدرات الذكائية بطريقة سبيرمان، وكان ٧٣، للاختيار من متعدد، و (٦٨،) للإجابة القصيرة. وهذا مؤشر على صدق التصنيف للمجال المعرفي إلى المستويين المعتمدين في هذه الدراسة.

أما عن صدق المحتوى، (وهو الأهم في هذه الدراسة). فقد اعتمدت طريقة التحكيم المشار إليها في سلسلة الاختبارات من خلال لائحة المواصفات. والدلالة

الإحصائية للفروق بين الأعداد المتوقعة والملاحظة في خلايا هذه اللائحة باستخدام الإحصائي كاي تربيع.

خطوات التجربة

للحصول على البيانات الإحصائية اللازمة لفحص فرضيات الدراسة، سارت التجربة وفق خطوات أخصها بما يلي:

١ - بعد التقسيم العشوائي للشعب الثلاثة التي يدرّسها المعلم نفسه وفي كل مدرسة وتحديد نوع الاختبارات كمعالجة التي تُعطى لكل شعبة. بدأ المعلمون بتدريس الوحدة. وقد تم إعلام الطلبة في المجموعات التجريبية أنهم سيخضعون لثلاثة اختبارات في هذه الوحدة (أي بعد كل درس من دروسها). وكان الفاصل الزمني بين كل اختبار والذي يليه أسبوعين. كما تم إعلام الطلبة في كل مجموعة بنوع الاختبارات في السلسلة الخاصة بها. وكانت تصحح إجاباتهم، ويتم مناقشتها مع الطلبة بعد كل اختبار. أما المجموعة الضابطة، فكانت تحل تمرينات الكتاب في وقت الاختبار ومناقشة الإجابات.

٢ - وبعد مرور أسبوعين من انتهاء سلسلة الاختبارات أعطيت المجموعات الثلاثة اختبار الاحتفاظ بصورتيه، فقد قسمت كل مجموعة إلى نصفين بطريقة عشوائية. أعطي أحدهما اختبار الاختيار من متعدد، وأعطى الآخر اختبار الإجابة القصيرة. صححت الإجابات بالطريقة المذكورة في حساب معاملات الثبات. وقد اعتمدت متوسطات علامات ثلاثة مصححين في حالة الإجابة القصيرة. كما حولت علامات الطلاب في كل من الاختبارين، وعلى كل مستوى، إلى علامات مئوية.

تحليل ومناقشة النتائج

تتلخص نتائج التحليل الإحصائي لعلامات الطلبة على اختبار الاحتفاظ التحصيلي في حساب المتوسطات والانحرافات المعيارية. وكذلك تحليل التباين الأحادي لفحص فرضيات الدراسة، وإجراء مقارنات بين المتوسطات، على ضوء نتائج تحليل التباين. وقد عرضت هذه النتائج في ثلاثة جداول. الجدول رقم (٢) ويتضمن المتوسطات الحسابية لعلامات الطلاب في كل مجموعة (مجموعة الإجابة القصيرة، ومجموعة الاختيار من متعدد، مجموعة التمرينات، وهي المجموعة الضابطة) على اختبار الاحتفاظ من نوع الإجابة القصيرة، والاختيار من متعدد، والنوعين معا. بمستوى

المعرفة، ومستوى القدرات الذكائية، وبالمستويين معا ليكونا المجال المعرفي من مجالات تصنيف الأهداف.

وقبل إجراء تحليل التباين الذي يجيب على كل فرضية من الفرضيات التسعة، ثم الكشف عن مدى تحقق افتراض تجانس التباين باستخدام طريقة بارتلت (Bartlett test).

وكانت جميع قيم χ^2 المحسوبة غير دالة إحصائيا على مستوى دلالة ٠,٠١، حيث تراوحت القيم التسعة بين ٣,٢٠ و ٩,٠٤ بينما كانت القيمة الحرجة ٩,٢١.

ثم أجري تحليل التباين الأحادي لعلامات الطلاب في المجموعات الثلاثة على اختباري الاحتفاظ (الإجابة القصيرة، والاختيار من متعدد) والنوعين معا، لنواتج التعلم في مستوى المعرفة، والقدرات الذكائية، والمجال المعرفي (المستويين معا). وذلك لفحص الفرضيات التسعة، حيث وضعت نتائج التحليل في الجدول رقم (٣).

ويلاحظ من الجدول أن جميع قيم F المحسوبة أكبر من القيم الحرجة المقابلة لها، وهذا يعني رفض الفرضيات الصفرية التسعة، التي نصت على أنه لا يوجد فروق ذات دلالة إحصائية بين تحصيل الطلبة على اختبار الاحتفاظ. وقد ترتب على هذه النتيجة إجراء مقارنات بعدية. بطريقة شافية، وقد وضعت نتائج المقارنات في الجزء الأسفل من الجدول رقم (٣).

ولتوضيح نتائج المقارنات وعلاقتها بالمتوسطات الحسابية للمجموعات الواردة في الجدول رقم (٢)، فإن الجدول رقم (٣) يفي بهذا الفرض. حيث تشير الأسهم إلى وجود فروق دالة إحصائية بمستوى دلالة ٠,٠٥، ولصالح المجموعة التي يبدأ السهم عند المتوسط الحسابي.

وتشير نتائج هذه المقارنات إلى تفوق الطلاب الذين تعرضوا لسلسلة من الاختبارات الصفية في أثناء دراستهم للمادة على الطلاب الذين لم يتعرضوا للاختبارات الصفية من أي نوع في الاحتفاظ بنواتج التعلم بالمستوى المعرفي، أو في الاحتفاظ بنواتج التعلم بالمستوى المعرفي، أو مستوى القدرات الذكائية، أو المجال المعرفي (المستويين معا). بمعنى أن الاختبارات تؤثر بشكل ما في الاحتفاظ بالمادة المتعلمة، ويمكن عزو ذلك إلى الاستراتيجية التعليمية التي اتبعها الطلاب في المجموعتين التجريبيتين في أثناء دراستهم للمادة، بالإضافة إلى تأثيرها في الدافعية والتركيز الدراسي. كما يمكن أن يعزى تفوق مجموعتي الاختبارات إلى كون الاختبارات مواقف تعليمية اكتسب الطلاب من

الجدول رقم (٣)
نتائج تحليل التباين ومقارنة المتوسطات بطريقة شافية لكل ناتج من نواتج التعلم (معرفة، قدرات ذكائية، المجال المعرفي) مقاسة باختبار الإجابة القصيرة، والاختبار من متعدد، والنوعين معاً.

مقارنة المتوسطات بطريقة شافية (ش)				ف الدرجة	ف	متوسط المربعات	درجات الحرية	مصدر التباين	ناتج التعلم (الاختبار المستخدم)
ش الدرجة	٢س	١س	٢س						
٦,١٤	٣١,٣٧ ٥,٧٣	١٠,٣٦ ٢س	٣,٠٧	٣,٠٧	١٥,٣٥	٨٠٤٥,٤٦ ٥٢٤,٢٨	٢ ١٤١	بين المجموعات داخل المجموعات	معرفة (إجابة قصيرة)
٦,١٤	٢٣,٠٤ ٦,٥٩	٥,٠١ ٢س	٣,٠٧	٣,٠٧	١١,٢٠	٤٥٨٨,١٩ ٤٠٩,٤١	٢ ١٤١	بين المجموعات داخل المجموعات	القدرات الذكائية (إجابة قصيرة)
٦,١٤	٢٧,٢٩ ٦,٥١	٧,١٨ ٢س	٣,٠٧	٣,٠٧	١٣,٢٧	٥٥٣١,٠٣ ٤١٦,٨٢	٢ ١٤١	بين المجموعات داخل المجموعات	المجال المعرفي (إجابة قصيرة)
٦,١٤	١٦,٨٧ ١٠,٩٥	٨٦ ٢س	٣,٠٧	٣,٠٧	٩,٣٨	٤١٦٨,٠٨ ٤٤٤,٣٨	٢ ١٤٠	بين المجموعات داخل المجموعات	معرفة (اختبار من متعدد)
٦,١٤	١٥,٦٨ ٣,٨٨	٥,٦٩ ٢س	٣,٠٧	٣,٠٧	٧,٨٧	٢٥٣٣,٧٩ ٣٣٢,٠٧	٢ ١٤٠	بين المجموعات داخل المجموعات	قدرات ذكائية (اختبار من متعدد)

تابع الجدول رقم (٣)

مقارنة المتوسطات بطريقة شاذية				ف	ف	متوسط المرينات	درجات الحرية	مصدر البيانات	نتائج التعلم الاختبار المستخدم
ش	س ^١	س ^٢	(ش)						
٦,١٤	١٦,٦٠ ١,٣٠	٢,٨٩	س ^١ س ^٢	٣,٠٧	٨,٤٢	٢٧٨٨,٤٣ ٣٣١,٣٤	٢ ١٤٠	بين المجموعات داخل المجموعات	المجال المعرفي (اختبار من متعدد)
٦,٠٨	٤١,٦٣ ١٤,٢١	٧,٣٦	س ^١ س ^٢	٣,٠٤	٢٠,٠	١١٢٧٩,٩٩ ٥٦١,٦٥	٢ ٢٨٤	بين المجموعات داخل المجموعات	معرفة (النوعين معا)
٦,٠٨	٣٦,٥٤ ١٠,٣١	٩,٠٠	س ^١ س ^٢	٣,٠٤	١٨,٢٩	٦٩٠٧,٥٣ ٣٧٧٥٨	٢ ٢٨٤	بين المجموعات داخل المجموعات	قدرات ذكائية (النوعين معا)
٦,٠٨	٤٠,٨ ١٢,٨	٨,٨٨	س ^١ س ^٢	٣,٠٤	١٩,٩٩	٧٩٨٥,٨٧ ٣٩٩,٥٣	٢ ٢٨٤	بين المجموعات داخل المجموعات	المجال المعرفي (النوعين معا)

الجدول رقم (٤)
ملخص نتائج المقارنات الثنائية بين المتوسطات الحسابية للمجموعات الثلاثة بطريقة شافيه

نوع الاختبار									
التعيين معاً			الاختيار من متعدد			الإيجابية القصيرة			معرفة
معرفة	قدرات ذكائية	معرفة	معرفة	قدرات ذكائية	معرفة	معرفة	قدرات ذكائية	معرفة	
٥٢,٣٩ ↓ ٤٣,٨٩ ↓ ٣٣,٧٧	٥٢,٨٣ ↓ ٤٤,٥١ ↓ ٣٥,٥٠	٥١,٧٩ ↓ ٤٢,٦٢ ↓ ٢٩,٧٣	٥٥,٨٥ ↓ ٤٩,٥٩ ↓ ٤٠,٢٩	٥٥,٦٥ ↓ ٤٧,٩٤ ↓ ٤٠,٧٣	٥٧,٤١ ↓ ٥٣,٤٦ ↓ ٣٩,٢٦	٤٩,٠٨ ↓ ٣٨,٠٨ ↓ ٢٧,٣٧	٥٠,١٣ ↓ ٤١,٠٢ ↓ ٣٠,٢٧	٤٦,٣٩ ↓ ٣١,٥٧ ↓ ٢٠,٢١	مجموعة الإيجابية القصيرة مجموعة الاختيار من متعدد مجموعة التمرينات (الضابطة)

ملحظة : يشير السهم إلى وجود فرق ذي دلالة إحصائية بين المتوسطين ولصالح المجموعة التي يبدأ السهم عند متوسطها الحسابي.

خلالها مهارة انتقاء المهم من جوانب المادة التعليمية وتخزينه في الذاكرة. وهكذا فإن توالي الاختبارات تعمل على إكساب الطلاب مهارة اختيار المهم من محتوى الاختبار واختزانه، وتعمل على إكسابهم مهارة البحث في مخزن الذاكرة عن المعلومات المتعلقة بكل سؤال من أسئلة الاختبار، وتجميعها، وتنظيمها، وإعادة صياغتها بالشكل المناسب للإجابة عن السؤال، وهذا بدوره عزز عملية التثبيت (Enhanced consolidation) لنواتج التعلم.

وتتفق هذه النتيجة مع نتائج عدة دراسات سابقة، ومن هذه الدراسات: دراسة آلن وآخرين (Allen, et., 1969)، ودراسة لابورت وفوس (laporte and Voss, 1975)، ودراسة دكستل (Duchastel, 1979) حيث وجدت أن تعرض الطلبة لسلسلة من الاختبارات يزيد احتفاظهم بالمادة المتعلمة وعدت الاختبار وسيلة قوية لتحسين التعلم.

بينما لم تتفق هذه النتيجة مع نتائج دراسات أخرى مثل دراسة نانجستر ودكستل (Nungester and Duchastel, 1982)، ودراسة هالين وهالين (Halpin and Halpin, 1982) حيث وجدت هذه الدراسات أن الطلاب الذين تعرضوا للاختبارات بعد دراستهم للمادة لم يتفوقوا في الاحتفاظ بتلك المادة على اختبار الاحتفاظ إلا من خلال الفقرات المعادة. وأن فوائد الاختبار كانت محدودة على الفقرات القديمة، ولم يمتد تأثيرها إلى الفقرات الجديدة وربما يعود ذلك لاختلاف في طبيعة المادة، وأنواع التصميم المستخدمة.

وتشير النتائج أيضا إلى أن درجة الاحتفاظ تتأثر بنوع الاختبار الذي تعرض له المجموعة، إذ يلاحظ من الجدول رقم (٤) أيضا أن هناك اختلافا بين النتائج فيما بين العمود الثالث، والعمودين الثاني والأول. حيث يلاحظ من خلال العمود الثالث أن مجموعة الإجابة القصيرة متفوقة وعلى كل المستويات على مجموعة الاختيار من متعدد. ويمكن أن يُعزى ذلك إلى اختبارات الإجابة القصيرة كمعالجة بصرف النظر عن نوع الاختبار النهائي. إلا أن التمايز بدأ يتضح عندما فصلت المجموعات، حيث أعطي نصف كل منها اختبار الإجابة القصيرة، وأعطى النصف الآخر اختبار الاختيار من متعدد، مما يشير إلى إضاعة المعلومات، لو اقتصر التحليل على العمود الأول. ولو اقتصر التحليل على العمود الثالث، لأمكن الباحث أن يوصي بضرورة استخدام الاختبارات الصفية من نوع الإجابة القصيرة كمعالجة بصورة مطلقة لقياس الأهداف المعرفية لو ثبتت باقي الظروف التعليمية.

وبالنظر إلى العمود الأول يلاحظ تفوق مجموعة الإجابة القصيرة على مجموعة الاختيار من متعدد، على اختبار الاحتفاظ نوع الإجابة القصيرة، على مستوى الأداء الكلي، ومستوى المعرفة. ويمكن أن يُعزى ذلك إلى قوة الأثر التعليمي لاختبارات الإجابة القصيرة، بالمقارنة مع الأثر التعليمي لاختبار الاختيار من متعدد، وأن اختبارات الإجابة القصيرة أحدثت تعلمًا أفضل، وأكثر ديمومة من اختبارات الاختيار من متعدد، وأن اختبارات الإجابة القصيرة تحت الطالب على إعمال فكره في البحث في الذاكرة عن إجابات للأسئلة، كما تحته على المراجعة النشيطة. وأنها أكسبته قدرة على استدعاء المعلومات من الذاكرة، أعلى من تلك التي أكسبته إياها اختبارات الاختيار من متعدد، هذه الاختبارات التي تتطلب من المفحوص أن يتعرف (Recognize) الإجابة الصحيحة من بين مجموعة البدائل الخاطئة المرافقة. ولا شك أن عملية الاستجابة من الذاكرة، عن طريق الاسترجاع تتطلب درجة كبيرة من المراجعة العقلية، بالمقارنة مع اختبارات الاختيار من متعدد، حيث إن اختبارات الاختيار من متعدد تنقص درجة البحث المطلوب من الذاكرة، بسبب تزويد المفحوص بالإجابة الصحيحة من بين البدائل. بمعنى أن البحث في الذاكرة، والمراجعة العقلية النشيطة، واستدعاء المعلومات من الذاكرة، كلها عوامل تزيد من قدرة الطالب على التمييز واسترجاع التعلم المسبق، وتزيد من انتقال أثر التعلم عن طريق تحويل المادة المتعلمة من الذاكرة القصيرة المدى إلى الذاكرة الطويلة المدى. وبالتالي زيادة تثبيت نواتج التعلم. ولكن يظهر من النتائج أن زيادة التثبيت درجات فإن كان قد حصل لمدى مجموعة الاختيار من متعدد، إلا أنه قد حصل بشكل أفضل لدى مجموعة الإجابة القصيرة.

وبالنظر إلى العمود الثاني من الجدول نفسه يلاحظ عدم وجود فروق ذات دلالة إحصائية ($\infty = 0,05$) بين متوسطات مجموعة الإجابة القصيرة ومتوسطات مجموعة الاختيار من متعدد على الاحتفاظ نوع الاختيار من متعدد، وعلى مستويات الأداء الثلاثة.

وجاءت هذه النتائج متفقة مع نتائج دراسات كل من جيه (Gay, 1980) وفوس (Voss, 1974)، ودكستل (Duchastel, 1981) حيث أشارت إلى تفوق طلاب مجموعة الإجابة القصيرة على طلاب مجموعة الاختيار من متعدد، على فقرات اختبار الاحتفاظ من نوع الإجابة القصيرة. بينما لم تتفق مع نتائج دراسات أخرى مثل دراسة ساكس وكولت (Sax and Collet, 1968) ودراسة دكستل ونانجستر (Duchastel and

(Nungester, 1982) حيث وجدت فروق غير دالة إحصائية بين مجموعتي الاختيار من متعدد، والإجابة القصيرة على اختبار الاحتفاظ من نوع الإجابة القصيرة.

ويمكن تفسير ذلك بأن قوة الأثر التعليمي لاختبارات الإجابة القصيرة قد ظهر على النوع الآخر من الاختبارات. كما ظهر على النوع نفسه. حيث تمكن طلاب مجموعة الإجابة القصيرة أن يحصلوا على متوسطات أعلى من طلاب مجموعة الاختيار من متعدد، وعلى اختبار الاحتفاظ نوع الاختيار من متعدد، وإن لم يكن هناك فروق ذات دلالة إحصائية بين متوسطات علامات المجموعتين. وهكذا يمكن القول إن تعرض الطلاب إلى سلسلة من الاختبارات لم يكن عملية تدريب أو تمرين فقط، فلو كانت كذلك لفوق طلاب كل مجموعة على النوع الذي تعرضوا له في أثناء عملية التدريس. لكن النتائج أظهرت خلاف ذلك. وبدل تفوق مجموعة الإجابة القصيرة على أن تعريض الطلاب إلى سلسلة من الاختبارات هو استراتيجية تعليمية كما يدل على أن اختبارات الإجابة القصيرة قد رفعت كفاءة الطلاب في الاحتفاظ بالمادة وزادت قدرتهم على استخدام نواتج التعلم في مواقف جديدة.

وقد اتفقت هذه النتائج أيضا مع نتائج دراسة جيه (Gay, 1980) ودراسة نانجستر ودكستل (Nungester and Duchastel, 1982)، التي أشارت إلى تساوي أداء طلاب المجموعتين على اختبار الاحتفاظ من نوع الاختيار من متعدد، بينما لم تتفق مع نتائج دراسة ساكس وكولت (Sax and Collet, 1968)، ودراسة دكستل ونانجستر (Duchastel and Nungester, 1982)، التي أشارت إلى تفوق مجموعة الإختيار من متعدد على اختبار الاحتفاظ من نوع الاختيار من متعدد.

كما تظهر النتائج أنه لا توجد فروق ذات دلالة إحصائية ($\alpha = 0.05$) بين متوسط علامات طلاب مجموعة الاختيار من متعدد، ومتوسط علامات طلاب المجموعة الضابطة في مستوى الفهم من اختبار الاحتفاظ نوع الاختيار من متعدد.

ويمكن تفسير هذه النتيجة على أساس أن طلاب المجموعة الضابطة قد لجأوا إلى التخمين، أو ربما أن اختبارات الاختيار من متعدد التي تعرض لها الطلاب سابقا قد كونت لديهم نزعة لاختزان الإجابات الخاطئة، مثل خزنها للإجابات الصحيحة وعند محاولة التذكر فيما بعد وجد الطلاب صعوبة في التمييز بين الإجابات الخاطئة والإجابات الصحيحة على المستويات العقلية العليا. وأن مجموعة الاختيار من متعدد استخدمت التشابه في المواقف التعليمية بصورة سلبية، حيث اختزنت المادة الخام المتعلقة بالإجابات

الخاطئة كما هي بالنسبة للمادة المتعلقة بالإجابات الصحيحة. لذلك ينصح المعلمون بأن يعملوا على محور الإجابات الخاطئة من خلال التغذية الراجعة. وتعزيز الإجابات الصحيحة لدى طلابهم.

كما تظهر النتائج أيضا عدم وجود فروق ذات دلالة إحصائية على مستوى الدلالة (٠,٠٥) بين متوسط علامات طلاب مجموعة الاختيار من متعدد، ومتوسط علامات طلاب المجموعة الضابطة، على مستوى المعرفة من اختبار الاحتفاظ، نوع الإجابة القصيرة.

ويمكن تفسير هذه النتيجة على أساس القول إن اختبارات الاختيار من متعدد، لم تستطع تزويد الطلاب بمهارة استرجاع المعلومات في مستوى المعرفة، عند تعرضهم لاختبار مغاير للنوع الذي سبق ان تعرضوا له، ويعود ذلك إلى اتساع الفجوة بين اختيار البدائل بالتحليل الإحصائي، والبدائل التي يمكن اختيارها من الإجابات القصيرة، عند محاولة ترجمة الإجابات القصيرة إلى اختيار من متعدد. لذلك ينصح المعلمون بأنهم إذا رغبوا في تكوين معرفة فقرات من نوع الاختيار من متعدد، فالأفضل أن يعطوا طلابهم اختبار إجابة قصيرة أولاً، لتكون أكثر الإجابات الخاطئة تكراراً بدائل في فقرات ذلك الاختبار.

كما يظهر العمود الأول أيضا عدم وجود فروق ذات دلالة إحصائية على مستوى الدلالة (٠,٠٥) بين متوسط علامات طلاب مجموعة الإجابة القصيرة، ومتوسط علامات طلاب مجموعة الاختيار من متعدد، في مستوى الفهم من اختبار الاحتفاظ، نوع الإجابة القصيرة. وربما كان ذلك يعني أن قوة الأثر التعليمي لاختبارات الاختيار من متعدد، قد ظهرت على النوع الآخر من الاختبارات، هو اختبار الإجابة القصيرة بالإضافة إلى أثرها في النوع نفسه كما في العمود الثاني، كما يمكن أن يعزى إلى طبيعة المادة التعليمية، (وهي هنا اللغة العربية) التي تميل إلى المستوى المعرفي، حيث إن الأسئلة في مستوى الفهم تعتمد على القاعدة المعرفية. وفي كثير من المواد التي تعتمد على القاعدة المعرفية، لا يكون هناك حاجة إلى أن يقرأ المفحوص البدائل في اختبار الاختيار من متعدد، ليتعرف الإجابة الصحيحة، بل يتذكر الإجابة مباشرة من خلال قراءته لأرومة السؤال، فيجيب بطريقة إجابة المفحوص نفسه عن السؤال من نوع الإجابة القصيرة.

كما يبدو واضحاً أن اختبارات الإجابة القصيرة أصعب من اختبارات الاختيار من متعدد، وهذا بالطبع غير ناتج عن مستوى الهدف الذي تقيسه الفقرة - حيث يظهر ذلك من خلال مقارنة متوسطات أداء كل مجموعة، بمتوسطاتها على اختبار الإجابة القصيرة، وعلى اختبار الاختيار من متعدد، وعلى مستويات الأداء الثلاثة حيث يلاحظ أن متوسطات أداء المجموعتين على اختبار الاختيار من متعدد أعلى من متوسطات أدائها على اختبار الإجابة القصيرة، وعلى المستويات الثلاثة وهذا يتفق مع ما ذكره براون (Brown, 1983) الذي ذكر أن فقرات الإجابة القصيرة أصعب من فقرات الاختيار من متعدد. ويتعارض مع ما ذكره جرونلند (Gronlund, 1976)، الذي رتبها بالعكس تماماً.

الخلاصة

وما تجدر الإشارة إليه هنا هو أن نواتج التعلم المقاسة في هذه الدراسة مبنية على صدق التصنيف من الدرجة الأولى للأهداف في المجال المعرفي (المعرفة، والقدرات الذكائية)، وقد لوحظ من نتائج التحليلات الإحصائية الجانبية إن اختبار الإجابة القصيرة أكثر قدرة على التمييز بين المستويين في اختبار الاختيار من متعدد. فقد كان معامل ثبات الفرق بين المستويين في اختبار الاختيار من متعدد ٠,٢٦، بينما كان هذا المعامل في اختبار الإجابة القصيرة ٠,٦٣، إلا أن معاملي ثبات الفرق بين مستوى المعرفة ومستوى القدرات الذكائية منخفض بشكل عام. وهذا يجعل الثقة في التحليل لنواتج التعلم في المجال المعرفي أعلى منها في التحليل لكل مستوى بمفرده. خاصة أن معامل الارتباط بين مستوى المعرفة والقدرات الذكائية في الاختيارين كان عاليا نسبياً، حيث بلغ ٠,٧٣، للاختبار من متعدد، ٠,٦٨، للإجابة القصيرة. وقد لا تشجع هذه النتيجة باحثاً آخر في إعادة إجراء هذه الدراسة باعتماد التصنيف من الدرجة الثانية للأهداف في المجال المعرفي. ولكن يمكن التوصية بإجراء دراسات أخرى في هذا المجال على ضوء الإجراءات المتبعة في هذه الدراسة، ومن خلال النتائج وتفسيرها، أوجزها بما يلي:

- ١ - دراسة الخصائص السيكومترية لاختبارات الاختيارات من متعدد المطورة من خلال إجابات الطلبة على اختبارات الإجابة القصيرة المصممة لقياس الأهداف نفسها.
- ٢ - إعادة الدراسة باستخدام تصاميم تجريبية، وطرق تصحيح غير التي اتبعت في الدراسة.
- ٣ - إجراء دراسات تبحث أثر هذين النوعين من الاختبارات في الاحتفاظ وعلاقته باتجاهات الطلاب، وطرق التدريس، ونوع التغذية الراجعة.

المراجع

- 1 - Al-Mahmoud, Naser, Y. «self Reported Teaching Proficiencies and their sources», Ananalysis of Teachers perception of their competencies. (Unpublished Docctoral Dissertation, University of illinois, Urbana-Champaign, 1984).
- 2 - Borgdon, R.E. (1978). In search of competencies for preparing middle school teachers. **Peabody Journal of Education**, 55-56, 145-151.
- 3 - **Campbell, Larry, F. «Astudy of pupil gain scores as acriterion of Teacher effectiveness». Dissertation Abstracts international, 29, October-December, 1968.**
- 4 - Guba, E.G. (1978). Toward amethodology of naturalistic inquiry in Educational evaluation. CSE Monograph services in Evaluation 8, center for the study of Evaluation, University of California, Los angeles.
- 5 - Ingersol, G.M. (1976). Assessing inservice training needs through teacher responses. **Journal of Teacher Education** 27, 169-173.
- 6 - Medley, D.M. (1979). The effectiveness of teachers. Inp. L. person, 8H. J. Walberg (Eds) Research on teaching. 11-17.
- 7 - Medly, D.M. (1973). Closing the gap between research teacher effectiveness and the teacher Education curriculum, **Journal of Research and Development Education**, 7, 39-46.
- 8 - Moure, K.D. Hanly, P.E. (1982). Anidentification of Elementary Teacher needs, **American Educational Research Journal**, 19, (1), 137-144.
- 9 - Percy, E.B. (1967). The Teacher and the public School system (2nd ed). PP. 176-180.
- 10 - Gargiulo, R.M & Pigge, F. (1978). Teacher Competencies: Need, Proficiency and wher proficiency was developed. **Journal of Teacher Education**, 29 70-76.
- 11 - Ryans, D.G. (1960) Characteristics of Teachers. Washington, D.C.: American Council on Education.
- 12 - Shannon, J.R. (1941). Elements of edcellence in teaching. **Educational Adminstration and Supervision**, 27, 168-176.
- 13 - Young, F. (1939). Some factors affecting teaching efficiency, **Journal of Educational Research**, 32, 649-652.

- Sax, G. and Reader, M. Achievement as a function of Test Difficulty level. **American Educational Research Journal**, 1964, Vol. 1, pp. 22-25.
- Voss, J.F. Acquisition and Nonspecific Transfer Effects in prose Learning as a Function of Question form. **Journal of Educational psychology**, Vol. 6, No. 5, 1974, pp.736-740.

مراجع اضافية

- Adkins D.C. Test Construction. Ohio: A Bell and howell Company, 1974.
- Allen, G.A., Mahler, W.A., and Estes, W.K. "Wffects of Recall Tests on long-Term Retention of paired Associates." **Journal of varbal Behavior**, 1969, 8, pp. 463-471.
- Brown, F.G. Principles of Educational and Psychological Testing. And edition, Holt Rinehart and Winston, NewYork, 1983.
- Duchastel, P.C. Retention of Prose Materials: The effect of testing. **Journal of Educational Research**. Vol. 72, No. 5, 1979, pp. 299-300.
- Duchastel, P.C. Retention of prose following Testing with diferent types of tests. **Contemporary Educational Psychology**. 6, 1981, pp. 217-226.
- Duchastel, P.C. and Nungester, R.J. Testing Effects Measured with Alternate Test Form. **Journal of Eduational Research**, Vol. 75, No. 5, 1982, pp. 309-313
- Gay, L.R. "The Comparative Effects of Multiple choice versus Short Answer Tests on Retention. **Journal of Educational Measurement**, Vol. 17, No. 1, Spring, 1980, pp. 45-50.
- Gay, L.R. and Gallagher, P.D. The comparative Effectiveness. **The Journal of Educational REsearch**, Vol, 70, 1976. pp. 59-61.
- Gronlund , N.E. Measurement and Evaluation in teaching. 3rd ed. Macmilan, NewYork, 1976.
- Gronlaund, N.E. Constructing Achievement Test. 2nd. ed. London prentice-Hall, 1977.
- Halpin, G. and Halpin, G. Experimental investigation of the effects of study. and testing on student learning, Retention, and Ratings, of instruction. **Journal of Educational Psychology**, Vol. 74, No. 1, 1982, pp. 32-38.
- Hogan, R.M. and Kintesch, W. Differential Effects of Study and test. Trials on long Term Recognition and Recall, **Journal of verbal learning and berbal Behavior**, Vol. 10, 1971, pp. 562-567.
- Hopkins, C.D. and Antes R.L. Classroom Measurement and Evaluation. F.E. Peacock publishers, INC. 1978, p. 124.

- Laporte, R.E. and Voss, J.F. "Retention of prose materials as a function of postacquisition Testing. **Journal of Educational psychology**, Vol. 67, No. 2, 1975. pp. 259-266.
- Nungester, R.J. and Duchastel, P.C. Testing versus Review: Effects on Retention. **Journal of Educational Psychology**, Vol. 74. No. 1, 1982, pp. 18-22.
- Sax, G. and Collet, L.S. An Empirical Comparison of the Effects on Recall and multiple-Choice Tests on student Achievement. **Journal of Educational Measurement**. Vol. 5, No. 2, Summer, 1968, pp. 169-173.