

تحديد العدد الأمثل لخيارات الفقرة في اختبار الاختيار من متعدد في ضوء بعض خصائصه السيكمترية

د. نعمان الموسوي

– أستاذ القياس والتقويم التربوي المساعد – كلية التربية / جامعة البحرين

المجلة التربوية (٢٠١)

القدرات التنظيمية، وقدرة الطالب على التعبير وإبداء الرأي بينما يشجع استخدام الأسئلة الموضوعية لقياس مدى إمتلاك الطالب لمهارات التذكر والفهم والتطبيق ضمن عينة أوسع من محتوى المادة المراد معرفة درجة إتقان الطالب لها.

ومع أن الأسئلة الموضوعية، وأسئلة الاختيار من متعدد على وجه التحديد، تتعرض باستمرار، للكثير من الانتقادات لأنها «تقيس عمليات معرفية بسيطة تنحصر في إعادة التعرف على معلومات مفتتة سبق للطالب التعرف عليها من قبل» (وهبة، ١٩٩٦)، بالإضافة إلى كونها لا تعبّر عن القدرات العقلية العليا لدى الطالب، إلا أنها لا زالت تمثل النوع الأكثر شيوعاً في إختبارات القدرة والتحصيل نظراً لقدرتها على قياس درجة إمتلاك الطالب للمهارات المطلوبة ضمن مساحة أكبر من محتوى المادة الدراسية في فترة زمنية قصيرة، ونظراً لسرعة وعدم تدخل الذاتية في تصحيحها، وإرتفاع درجة صدقها وثباتها، كما أنها لا تتيح للطالب فرصة اللف والدوران أو التهرب من الإجابة المباشرة عن السؤال.

ومن بين أنماط الأسئلة الموضوعية المستخدمة في قياس التحصيل في البلدان العربية لا زالت أسئلة الاختيار من متعدد بأربعة خيارات تستأثر بنصيب الأسد فيها. ويعود ذلك إلى عوامل كثيرة أهمها شيوع الفكرة السائدة في أوساط المشتغلين بالقياس والتقويم التربوي بأنه كلما زاد عدد خيارات الفقرة عن ثلاثة، قلت فرصة تخمين الإجابة الصحيحة من جانب الطالب (خضر، ١٩٨٧، أبو زينة، ١٩٩٢؛ أبو عودة، ١٩٩٣). ولهذا السبب أصبح من المتعارف عليه أن تتألف أسئلة الاختيار من متعدد من أربعة خيارات فما فوق دون أن يجهد التربويون في البلاد العربية أنفسهم بالبحث في إمكانية استخدام خيارات أقل من أربعة، هذا على الرغم من وجود عدد كبير من الدراسة الأجنبية التي برهنت على إرتفاع درجة الصدق والثبات في أسئلة الاختيار من متعدد بثلاثة خيارات قياساً لمثيلتها ذات الخيارات الأربعة فأكثر، كما أنها أوصت مراراً باستخدامها لقياس مهارات معينة لدى المتعلمين.

وإنطلاقاً مما سبق جاءت هذه الدراسات لتتحرى إمكانية تقليص عدد الخيارات في أسئلة الاختيار من متعدد إلى ثلاثة عبر إجراء مقارنة بين الخصائص السيكمترية (الصدق والثبات والقدرة التمييزية) لإختبار من نوع الاختبار من متعدد بأربعة خيارات أولاً، وبثلاثة خيارات ثانياً، أي بعد حذف أحد الخيارات من كل سؤال في الاختبار، أخذاً بالحسبان تباين المستويات التحصيلية لعينة المفحوصين التي يتم تطبيق الاختبار عليها بصورتيه المذكورتين.

الدراسات السابقة :

لا تتوافر في حدود علم الباحث وإطلاعاته دراسة عربية واحدة تتناول مباشرة مسألة تحديد العدد الأمثل optimum number للخيارات في أسئلة الاختيار من متعدد. كما أن استخدام هذه الأسئلة بعدد من الخيارات يقل عن أربعة لم يطرح إطلاقاً في دراسات الباحثين التربويين العرب. وقد يعود ذلك لترسخ قناعة لدى معظم المشتغلين بالقياس والتقويم التربوي في البلاد العربية بأن هذا النوع من الأسئلة يجب ألا يتضمن عدداً من الخيارات يقل عن أربعة، لأن تقليل عدد الخيارات «يرفع فرصة العثور على الإجابة الصحيحة عن سؤال الاختيار من متعدد بطريق الصدفة أو التخمين العشوائي من الربع في حالة أربع خيارات إلى الثلث في حالة ثلاثة خيارات» (أبو لبدة، ١٩٩٤، ص. ٣٢٧). وبالمقابل فإن زيادة عدد الأجوبة الممكنة عن سؤال الاختيار من متعدد «تقيّد عامل الحظ في اختيار الطالب للإجابة الصحيحة، وتحدّ من أثره، حيث يخف احتمال إختيار جواب ما وإستبعاد آخر عن طريق القرعة، ويتعزز في ذلك دور العلم الحقيقي، كما يتعزز الشعور بضرورة الإمام بالموضوع المطروح إماماً صحيحاً» (عطية، ١٩٩٢، ص. ٢٠٩). وبالتالي فإن تحديد عدد الخيارات التي يجب أن تتضمنها أسئلة الاختيار من متعدد كي يتحقق فيها الحد الأقصى الممكن من الصدق والثبات والقدرة التمييزية لم يطرح كمسألة قيد البحث لدى المتهمين بقضايا القياس والتقويم في الدول العربية.

وبالمقابل حظت مسألة تحديد العدد الأمثل لخيارات فقرة الاختيار من متعدد بإهتمام الباحثين الأجانب، وعلى الأخص منهم السيكمترين الأمريكيين، قبل فترة زمنية طويلة. ففي عام ١٩٦٤م برهن تفيرسكي رياضياً على أنه يمكن، من الناحية النظرية، تحقيق الحد الأمثل لبعض خصائص الاختيار المتعددة كقوة الاختبار، وقدرته التمييزية بين المفحوصين، وكمية المعلومات المستقاة منه، في حالة إحتوائه على ثلاثة خيارات قياساً للإختبارات الأخرى بعدد مغاير من الخيارات شريطة أن يظل العدد الإجمالي للخيارات في الاختبار الواحد ثابتاً (Tversky, 1964).

وقد حقّزت هذه الإستنتاجات النظرية كوستين (Costin, 1970) على بناء اختبار تحصيلي بأربع خيارات في مساق علم النفس لعينة من أفراد القوات الجوية، قام بعدها بحذف خيار واحد من كل فقرة فيه بطريقة عشوائية ليصبح عدد خيارات الفقرة ثلاثة فقط. وقد وجد كوستين أن الفقرة ذات الخيارات الثلاثة قد أبرزت قدرة تمييزية أكبر، ومعامل ثبات أعلى، من مثيلتها ذات الخيارات الأربعة. وفي تطوير لهذا البحث (Costin, 1972) إنتقى كوستين عشوائياً ٥٠ فقرة من اختبار يتضمن ١٠٠ فقرة، وتم إختزال عدد خيارات كل فقرة من هذه الفقرات الخمسين إلى ثلاثة خيارات، ثم أعطيت هذه الفقرات، بالإضافة إلى الفقرات الخمسين التي لم يطالها الإختزال، لعينة قوامها ١٢٠٠ طالباً. وبمقارنة هاتين المجموعتين من الفقرات وجد كوستين أنه لا توجد فروق ذات دلالة إحصائية بينهما عند مستوى ٠,٠٥ سواءً في متوسطات الدرجات الخام، وفي معامل ثبات الإتساق الداخلي (كودر - ريتشاردسون ٢٠)، وكذلك في متوسط معامل الارتباط الثنائي الأصيل (Point- Biserial Correlation). وتبين أن معامل ثبات التجزئة النصفية لصيغة الخيارات الثلاثة أعلى من مثيلتها ذات الخيارات الأربعة في حالة إبقاء العدد الإجمالي للخيارات ثابتاً في الاختبار الواحد، أي بموجب الشرط السابق الذي وضعه تفيرسكي.

وقد أبرزت عدة دراسات تجريبية لاحقة صُممت على أساس الإفتراض

بعشوائية تخمين المفحوص للإجابة الصحيحة عن سؤال الاختيار المتعدد تقوُّق أسئلة الخيارات الثلاثة من حيث الكفاءة على مثلتها بأربعة أو بخمسة خيارات (Grier, 1975; Lord, 1977).

فباستخدام نموذج المنحنى المميّز للفقرة Item Characteristic Curve Model لتحليل بيانات مستقاة من إختبارات الإستعداد الدراسي Scholastic Aptitude Test (SAT) وجد لورد أن العدد الأقل من الخيارات يحقق معامل ثبات أكبر لدرجات المفحوصين ذوي المستويات التحصيلية العليا، وأقل لدرجات أقرانهم ذوي المستويات الدنيا في حالة ثبات العدد الإجمالي للخيارات. وفي معرض تفسيره لهذه النتيجة ذكر لورد بأن المفحوصين ذوي مستوى التحصيل المتدني يراهنون بشكل أكثر على التخمين العشوائي الذي يقود بدوره إلى التقليل من صدق الإختبار خصوصاً عندما تكون أسئلة الإختبار بخيارات أقل (Lord, 1997, p. 36).

في الآونة الأخيرة زاد إهتمام الباحثين الأمريكيين بمقارنة الخصائص السيكمترية لأسئلة الإختبار من نوع الإختيار من متعدد عندما تكون متفاوتة في أعداد خياراتها، ومطبقة على مجموعات من المفحوصين متباينة في مستوياتها التحصيلية. وقد جرت العادة أن يتم الحصول على صورة الإختبار بثلاثة خيارات عن طريق إجراء تحليل لفقرات الإختبار ذي الخيارات الخمسة، وحذف الخيار الذي نال أدنى نسبة إختيار من قبل المفحوصين من كل سؤال في هذا الإختبار بصورة دورية.

فقد أجرت وير (Weber, 1978) مقارنةً لمعاملات الثبات لصورتَي إختبار الإختيار من متعدد بثلاثة وخمسة خيارات من ناحية، وبثلاثة وأربعة خيارات من ناحية أخرى، وكانت نتيجة المقارنة أنه كلما زادت خيارات السؤال إرتفع معامل الثبات لدرجات مجموعة الطلبة منخفضي ومتوسطي التحصيل عندما يتم تثبيت طول وزمن الإختبار.

وبدورها أظهرت المقارنة التي أجرتها جرين وآخرون بين صيغ إختبارات الإختيار من متعدد بثلاثة وأربعة وخمس خيارات فروقاً ذات دلالة إحصائية عند مستوى ٠,٠٥ بين معاملات الثبات لدرجات مجموعة الطلبة منخفضي التحصيل، وعند مستوى ٠,١٠ لمجموعة الطلبة متوسطي التحصيل وذلك لصالح صيغة الخيارات الأربعة. وقد بينت الدراسة أن أربعة خيارات هو العدد الأمثل في إختبار الإختيار المتعدد للطلبة من مختلف المستويات التحصيلية (Green, Sax and Michael, 1991).

وبالمثل أبرزت دراسة المقارنة المماثلة التي قام بها تريفيسان وآخرون فروقاً ذات دلالة إحصائية عند مستوى ٠,٠٥ بين معاملات الثبات لمجموعة الطلبة منخفضي التحصيل، لكنها لم تبرز فروقاً مشابهة لمجموعتي الطلبة متوسطي ومرتفعي التحصيل. كما وجد الباحثون أن ثلاثة خيارات هو العدد الأمثل في إختبار الإختيار من متعدد للطلبة على اختلاف مستوياتهم التحصيلية (Trevisan, Sax and Michael, 1991). وقد توصل هؤلاء الباحثون إلى نفس النتيجة في دراسة لاحقة جرت فيها المقارنة بين معاملات الثبات لصور الإختيار من متعدد بخيارين وثلاثة وأربعة وخمسة خيارات، غير أن الحصول على صورة الإختيار بثلاثة خيارات فأكثر تم في هذه المرة عن طريق إضافة خيار إلى كل سؤال في صيغة الخيارين بصورة منتظمة (Trevisan, Sax and Michael, 1994).

يتضح من العرض السابق أن هناك اتفاقاً بين بعض المشتغلين الأجانب بالقياس والتقويم التربوي على أن صيغة الخيارات الثلاثة هي التي تحقق لإختبار الإختيار من متعدد القدر الأكبر من الصدق والثبات. غير أن هذا الإتفاق يزول حينما يدور الحديث عن تطبيق صيغ متباينة في أعداد خياراتها على عينات من المفحوصين متفاوتة في مستوياتها التحصيلية. فبينما تتفق نتائج دراسات الباحثين في هذا الموضوع على وجود فروق ذات دلالة إحصائية عند مستوى ٠,٠٥ بين معاملات ثبات الأسئلة المختلفة في عدد الخيارات، والمطبقة على مجموعات من المفحوصين تتميز بمستويات تحصيلية منخفضة، نلاحظ بروز تضارب في هذه

النتائج حول وجود مثل هذه الفروق عند مستوى ٠,٠٥ بين معاملات الثبات المحسوبة على مجموعات من المفحوصين ذوي مستويات تحصيلية متوسطة أو مرتفعة، وكذلك حول ما إذا كان العدد الأمثل للخيارات في أسئلة الاختيار المتعدد ثلاثة أم أكثر. كما أن قدرة هذا الاختبار على التنبؤ بدرجة المفحوص في مقياس المحك في فترة زمنية لاحقة، وكذلك قدرته على التمييز بين المفحوصين لم تحظيا بالقدر الكافي من الدراسة في الأبحاث الآتية الذكر.

ومن هنا تنشأ الحاجة إلى تحديد العدد الأمثل للخيارات في أسئلة الاختيار المتعدد عن طريق إجراء مقارنة لمعاملات الصدق والثبات والقدرة التمييزية لصيغ متفاوتة في أعداد خياراتها أثناء تطبيقها على مجموعات من الطلبة متباينة في مستوياتها التحصيلية.

مشكلة البحث :

تتلخص مشكلة البحث الحالي في تحديد العدد الأمثل للخيارات في أسئلة الاختيار من متعدد الذي يحقق أعلى درجة ممكنة من الصدق والثبات والقدرة التمييزية للاختبار في حالة تطبيقه على مجموعات من المفحوصين متفاوتة في مستوياتها التحصيلية.

أهداف البحث :

- ١ - تحديد العدد الأمثل للخيارات في سؤال الاختيار من متعدد.
- ٢ - الكشف عن العلاقة القائمة بين عدد الخيارات في أسئلة الاختيار من متعدد من جهة، ومستوى تحصيل الطالب، وخصائص هذه الأسئلة، من جهة أخرى.
- ٣ - تحديد الفروق القائمة بين معاملات الصدق والثبات والقدرة التمييزية لأسئلة الاختيار من متعدد المختلفة في عدد خياراتها، والمطبقة على عينات متفاوتة في المستوى التحصيلي.

أهمية البحث:

يستمد هذا البحث أهميته من الجوانب التالية:

- ١ - تفتقر الأدبيات التربوية العربية إلى الدراسات المتصلة باستخدام أسئلة الاختيار من متعدد بخيارات أقل من أربعة لغرض تقويم تحصيل الطلبة في مساق دراسي محدد.
- ٢ - يوفر هذا البحث للمعلم فرصة تقليل عدد الخيارات في فقرة الاختيار من متعدد من أربعة فأكثر، كما هو شائع في صيغ الاختبارات التحصيلية المستخدمة على نطاق واسع، إلى ثلاثة خيارات فقط. ومن شأن ذلك أن يؤدي إلى حث المعلم على إعداد كمية أكبر من الأسئلة ضمن الموضوع الاختباري الواحد، وبالتالي تغطية مساحة أكثر شمولاً من المادة الدراسية قيد الفحص. كما يسهم ذلك في تقليص الفترة الزمنية المخصصة للإمتحان، والحد من الآثار النفسية المصاحبة له كالقلق والإنفعالية والرغبة لدى المفحوص.
- ٣ - يقدم البحث الحالي أداة قياس جديدة صالحة للاستخدام مستقبلاً لغرض تقويم تحصيل الطلبة الجامعيين في مساق القياس والتقويم التربوي، لا سيما أنها تتمتع بدرجة عالية من الصدق، والثبات، والقدرة التمييزية.

منهج البحث:

لما كان البحث الحالي يهدف إلى تحديد العدد الأمثل لسؤال من نوع الاختيار من متعدد عن طريق مقارنة الخصائص السيكمترية لأسئلة من هذا النمط بأربعة وثلاثة خيارات وُضعت في مادة القياس والتقويم التربوي، فقد اعتمد الباحث المنهج الوصفي التحليلي كإسلوب مناسب لبناء الاختبار ذي الخيارات الأربعة، وتحديد خصائصه وذلك عن طريق:

- الإطلاع على بعض أدوات القياس المستخدمة في التربية.
- وصف محتوى المادة الدراسية قيد الفحص، وإشتقاق أهداف سلوكية لها.

- بناء جدول المواصفات .
 - وضع أسئلة تمثل هذه الأهداف وتكشف عن مدى تحققها .
 - تطبيق الإختبار على عينة ممثلة للجمهور المستهدف .
 - تحديد معاملات الصدق والثبات لدرجات الإختبار .
 - إستخراج مؤشر القدرة التمييزية لكل فقرة من فقرات الإختبار .
 - إحتساب نسبة إختيار المفحوصين للخيار في كل فقرة من فقرات الإختبار .
- وقد تم الحصول على صيغة الخيارات الثلاثة عن طريق حذف الخيار الذي حاز على أقل نسبة إختيار بين المفحوصين في كل فقرة من فقرات الإختبار بأربعة خيارات . وفي حالة حيازة أكثر من خيار في الفقرة على أقل نسبة إختيار يتم حذف أحدها بصورة عشوائية .

وتجدر الإشارة هنا إلى أنه كان من المفترض في هذه المرحلة من إعداد الإختبار أن تُعاد صياغة بدائل الفقرات التي تبيّن أن قدرتها على التمييز بين المفحوصين مرتفعي ومنخفضي التحصيل ضعيفة أو سالبة، إلا أننا اضطررنا إلى عدم القيام بهذه الخطوة بالنظر إلى ضرورة إجراء مقارنة في وقت لاحق بين مؤشرات التمييز لفقرات الإختبار التحصيلي بأربعة خيارات وبثلاثة خيارات، كما سيتضح فيما بعد .

التعريفات الإجرائية لمصطلحات البحث :

إختبار الإختيار من متعدد: تتكون الفقرة في هذا النوع من الإختبارات من المتن (الأرومة) التي تأخذ شكل السؤال أو الجملة الناقصة يتبعها إثنان من الخيارات أو أكثر . واحد هذه الخيارات يمثل الإجابة الصحيحة (أو الأصح)، وباقي الخيارات عبارة عن بدائل مُوهمة للإجابة الصحيحة . ويشار إلى هذا المصطلح في البحث الحالي بالإختبار .

الصورة الأولى: ونقصد بذلك أسئلة الإختبار التحصيلي بأربعة خيارات لكل سؤال .

الصورة الثانية: ونقصد بذلك أسئلة الإختبار التحصيلي بثلاثة خيارات لكل سؤال .

العدد الأمثل للخيارات: هو أدنى عدد من الخيارات يحقق للإختبار من نوع الإختيار من متعدد أقصى قيمة ممكنة من الصدق والثبات والقدرة التمييزية .

معامل الصدق: يقصد بهذا المصطلح في البحث الحالي معامل الصدق التنبؤي للإختبار الذي يعبر عنه معامل الارتباط بين درجة المفحوص في الإختبار ودرجته النهائية في المساق الدراسي قيد الفحص .

معامل الثبات: يقصد بهذا المصطلح في البحث الحالي معامل ثبات الإتساق الداخلي للإختبار المحسوب باستخدام معالة كرونباخ - ألفا .

القدرة التمييزية للإختبار: قدرة الإختبار على التمييز بين الطلبة مرتفعي التحصيل من جهة ، والطلبة منخفضي التحصيل من جهة أخرى وإبراز الفروق الفردية بين المفحوصين .

فئة منخفضي التحصيل: هي مجموعة المفحوصين الذين تتراوح درجتهم النهائية في المقرر قيد الفحص بين ٦٠ - ٦٧ من أصل مائة درجة هي الدرجة النهائية العظمى .

فئة متوسطي التحصيل: هي مجموعة المفحوصين الذين تتراوح درجتهم النهائية في المقرر قيد الفحص بين ٦٨ - ٧٦ من أصل مائة درجة هي الدرجة النهائية العظمى .

فئة مرتفعي التحصيل: هي مجموعة المفحوصين الذين تتراوح درجتهم النهائية في المقرر قيد الفحص بين ٧٧ - ٩٠ من أصل مائة درجة هي الدرجة النهائية العظمى .

فرضيات البحث:

- ١ - لا توجد فروق ذات دلالة إحصائية بين معامليّ الصدق التنبؤي لصورتَي الإختبار بأربعة وثلاثة خيارات لكل فئة ولجميع فئات الطلبة على إختلاف مستوياتهم التحصيلية.
- ٢ - لا توجد فروق ذات دلالة إحصائية بين معامليّ الثبات لصورتَي الإختبار بأربعة وثلاثة خيارات لكل فئة ولجميع فئات الطلبة على إختلاف مستوياتهم التحصيلية.
- ٣ - لا توجد فروق ذات دلالة إحصائية بين متوسطات القدرة التمييزية لصورتَي الإختبار بأربعة وثلاثة خيارات لفئات الطلبة لكل فئة ولجميع فئات الطلبة على إختلاف مستوياتهم التحصيلية.
- ٤ - العدد الأمثل لخيارات فقرة إختبار الإختيار من متعدد لكل فئة ولجميع فئات الطلبة على إختلاف مستوياتهم التحصيلية هو ثلاثة خيارات.

حدود البحث:

يتحدد البحث الحالي بالجوانب التالية:

- ١ - تقتصر الدراسة على دراسة التفاعل بين عدد الخيارات في أسئلة الإختيار من متعدد ومتغير واحد هو مستوى تحصيل الطالب على الخصائص السيكمومترية لهذه الأسئلة. أما المتغيرات الأخرى كجنس الطالب، وتخصصه، ومستواه الدراسي، والكلية التي يدرس فيها، فلم تشملها الدراسة الحالية.
- ٢ - تقتصر عينة الدراسة على ١٤٤ طالباً وطالبة من طلبة البكالوريوس بجامعة البحرين يمثلون ثلاث كليات هي كلية الآداب والعلوم والهندسة للعام الدراسي ١٩٩٦/٩٥ م.
- ٣ - يعتمد الباحث في بحثه بعض التعريفات الإجرائية الآنفه الذكر، ويلتزم قدر الإمكان وهو يضع الإختبار التحصيلي بمادة القياس والتقويم

التربوي بالقواعد الأساسية الخاصة بصياغة أسئلة الاختيار من متعدد، وهي: إحتواء متن السؤال على فكرة أساسية واحدة محددة المعالم، تساوي الخيارات في الطول وتجانسها منطقياً ولغوياً، توزيع الإجابات الصحيحة عشوائياً في خيارات الأسئلة.

عينة البحث:

بلغ حجم عينة البحث ١٤٤ طالباً وطالبة يدرسون مادة القياس والتقييم كأحد المقررات التربوية المطلوبة على الطلبة الذين إختاروا التربية كتخصص فرعي إلى جانب تخصصهم الأساسي في الجامعة. وقد أختير أفراد عينة البحث بصورة عشوائية من بين طلاب وطالبات التخصص الفرعي في التربية المنتظمين بكليات الآداب والعلوم والهندسة بجامعة البحرين للفصل الأول من العام الدراسي ١٩٩٦/٩٥ م. وقد أتبعت طريقة جداول الأرقام العشوائية في إختيار أفراد العينة.

وتتألف العينة من ٦٤ طالباً و ٨٠ طالبةً بينهم ٢٢ طالباً و ٣٢ طالبةً بكلية الآداب بمتوسط معدل تراكمي قدره ٢,٢٠؛ ٢٨ طالباً و ٤٢ طالبة بكلية العلوم بمتوسط معدل تراكمي قدره ٢,٧٨ بالإضافة إلى ١٤ طالباً و ٦ طالبات بكلية الهندسة بمتوسط معدل تراكمي قدره ٢,٣١. ومن بين أفراد العينة ٤٨ طالباً وطالبة منتظمين في المستوى الدراسي الرابع والخامس و ٩٦ طالباً وطالبة في المستوى الدراسي السادس والسابع، وبيّن الجدول رقم ١ توزيع أفراد العينة حسب الجنس والكلية والمستوى الدراسي ومتوسط المعدل التراكمي.

وتمثّل هذه العينة نحو ١٠/ من حجم المجتمع الأصلي الذي يصل إلى ١٤٤٦ من طلبة التخصص الفرعي في التربية المنتظمين بالكليات الآنفة الذكر في الفصل الأول من العام الدراسي ١٩٩٦/٩٥ م.

جدول رقم ١
توزيع أفراد عينة البحث وفق الكلية والجنس
والمستوى الدراسي ومتوسط المعدل التراكمي

الكلية	الجنس		المستوى الدراسي		متوسط المعدل التراكمي GPA
	ذكور	إناث	(٥، ٤)	(٧، ٦)	
الآداب	٢٢	٣٢	٢٠	٣٤	٢,٢٠
العلوم	٢٨	٤٢	٢٤	٤٦	٢,٧٨
الهندسة	١٤	٦	٤	١٦	٢,٣١

إجراءات تصميم وتطبيق أداة البحث :

أولاً: بناء الإختبار التحصيلي (الصورة الأولى).

تم بناء إختبار من نوع إختيار من متعدد يحتوي على ٤٠ سؤالاً بأربع خيارات لكل منها في إحدى وحدات مقرر القياس والتقويم التربوي الذي يقوم الباحث بتدريسه لعينة أفراد البحث الحالي، وهي وحدة «إعداد الإختبارات التحصيلية». وقد إقتضى بناء أسئلة الإختبار ووضعه في صيغته النهائية الإسترشاد بالأسس العامة المتبعة في تصميم إختبارات التحصيل الصفية (Hopkins et. al, 1978; Gronlund, 1990). وعلى أساس مَرَّت عملية بناء الإختبار التحصيلي بالمراحل التالية.

١ - تحديد الغرض من الإختبار: يهدف الإختبار التحصيلي التكويني إلى تقويم مدى تقدم الطلبة في تحقيق الأهداف السلوكية المتوخاة من تدريس وحدة «إعداد الإختبارات التحصيلية» في مساق القياس والتقويم التربوي، وتوفير التغذية الراجعة لهم.

٢ - صياغة الأهداف السلوكية: في هذه المرحلة تم تحديد النتائج التعليمية المراد قياسها في الوحدة المذكورة وإشتقاق أهداف سلوكية مرتبطة بها في مستويات المعرفة والفهم والتطبيق وفقاً لتصنيف بلوم (Bloom, 1956).

٣ - إعداد جدول المواصفات: هنا جري ربط مستويات الأهداف السلوكية بمحتوى المادة الدراسية موضوع الاختبار (ملحق رقم ١ - ٣).

٤ - صياغة أسئلة الاختبار: تم إختيار عينة تمثيلية من الأسئلة الموضوعية من نمط الاختيار من متعدد بأربعة خيارات بحيث تكون مطابقة للأهداف السلوكية المنشودة وجرى تحديد عدد أسئلة الاختبار في ضوء الفترة الزمنية المخصصة له، كما رُوِعت القواعد الآنفه الذكر أثناء صياغة الأسئلة.

٥ - التأكد من صدق محتوى الاختبار: تم عرض الصورة الأولية للاختبار على أربعة من المختصين في مادة القياس والتقويم والتربوي وذلك لأخذ وجهات نظرهم في مدى صدق الفقرة الاختبارية في قياس الهدف السلوكي المحدد، وإقتراح ما يرونه من تعديل. وقد بلغ متوسط نسبة الإتفاق بين المحكمين حول أسئلة الاختبار ٨٥٪، وتم إجراء التعديلات اللازمة عليها في ضوء مقترحاتهم. وقد أُرِفقت الصورة الأولى للاختبار في شكلها النهائي بالملحق رقم ١ - ١.

ثانياً: إجراءات تطبيق وتصحيح الاختبار:

تم تطبيق الصورة الأولى من الاختبار بأربعة خيارات لكل فقرة باعتبارها إختبار منتصف الفصل المقرر على الطلبة ضمن خطة تقويم أداء الطالب في المساق الدراسي وذلك قبل إجازة منتصف الفصل الأول. وبعد أن قام المفحوصون بملء البيانات المطلوبة منهم وقراءة تعليمات الاختبار طُلب منهم أن ينتقوا الخيار الذي يمثل برأيهم أصح إجابة بين الإجابات المعطاة لكل فقرة، وأن يدوّنوا رمز الخيار المرغوب في ورقة الأجابة المرفقة (ملحق رقم ١ - ١). وقد مُنح الطلبة ٤٠ دقيقة للإجابة عن الأسئلة. وحيث أنهم إستطاعوا إنهاء الإجابة عنها خلال هذه الفترة الزمنية فإن السرعة لم تكن عاملاً مؤثراً على أدائهم.

وبعد تصحيح أوراق الاختبار باستخدام المفتاح المثقب المعد وفقاً للإجابات الصحيحة (ملحق رقم ١ - ٤) تبين أن متوسط أداء الطلبة في الاختبار يساوي ٢٤,٣٦، ومعامل الصدق التنبؤي لجميع الفئات التحصيلية يساوي ٠,٧٣، ومعامل ثبات كرونباخ - ألفا يساوي ٠,٨، ومتوسط القدرة التمييزية يساوي ٠,٣٣. وقد دوت هذه البيانات في الجدول رقم ٢.

جدول رقم ٢

معاملات الصدق والثبات المحسوبة لصورتَي الاختبار التحصيلي والقيم المصححة لكل منها حسب مستوى تحصيل فئات العينة

معامل الإرتباط بين الصورتين	ثلاثة خيارات		أربعة خيارات		الصورة	
	معامل القيمة الثبات	معامل القيمة المصححة	معامل القيمة الثبات	معامل القيمة المصححة	الفئة	
*٠,٣٢	*٠,٧٤	٠,٦٨	*٠,٧٩	٠,٠٨	٠,٠٧	الدنيا
*٠,٥٩	*٠,٨٠	٠,٧٥	*٠,٧٣	٠,٠٨	٠,٠٧	الوسطى
*٠,٥٧	*٠,٧٠	٠,٦٣	*٠,٧٨	٠,١٦	٠,١٤	العليا
*٠,٦٧	*٠,٧٩	٠,٧٤	*٠,٨٠	*٠,٧٣	٠,٦٥	الجميع

* القيمة المحسوبة دالة إحصائياً عند مستوى ٠,٠٥

ثالثاً - بناء أسئلة الاختبار التحصيلي (الصورة الثانية):

تم إتباع الخطوات التالية في الحصول على صورة الاختبار بثلاثة خيارات:

- ١ - تقسيم المفحوصين حسب المستوى التحصيلي: لغرض تحليل مفردات الصورة الأولى للاختبار بأربعة خيارات تم تقسيم الطلبة إلى ثلاث فئات متساوية وهي: فئة مرتفعي التحصيل، فئة متوسطي التحصيل، وفئة منخفضي التحصيل، بحيث تضم كل منها ٤٨ طالباً وذلك وفقاً للدرجة النهائية المستحقة للطالب في المقرر. وقد حددت الدرجة القطعية لكل فئة وفقاً للتعريف الإجرائي المعتمد للفئة المذكور أعلاه.

٢ - تحليل فقرات الاختبار: هدف تحليل مفردات الصورة الأولى للاختبار بأربعة خيارات إلى تحديد درجة صعوبة كل فقرة، وقدرتها التمييزية بين الطلبة، وكذلك احتساب فاعلية كل خيار من خياراتها الخاطئة (البدائل) وفقاً للمعادلات المستخدمة لهذا الغرض (أبو زينة، ١٩٩٢، ص: ٢٨٤ - ٢٨).

٣ - تقليص عدد خيارات الفقرة: على أثر احتساب نسبة إختيار المفحوصين لكل بديل من البدائل الخاطئة لكل سؤال تم حذف البديل الذي حصل على أقل نسبة إختيار في كل فقرة، أو البديل الذي اختاره المفحوصون ذوي التحصيل العالي أكثر من أقرانهم من ذوي التحصيل المتدني، أيهما أقل. وفي حالة تساوي نسب الإختيار لأكثر من بديل في الفقرة يتم إستبعاد أحد هذه البدائل عشوائياً. ويوضح الملحق رقم ١ - ٥ رموز خيارات الأسئلة التي تم إستبعادها من الصورة الأولى للاختبار.

٤ - إعادة ترميز الخيارات: بعد أن أصبحت كل فقرة تضم ثلاثة خيارات فقط أعيد ترميز الخيارات الجديدة، وبذلك تم إستكمال بناء الاختبار التحصيلي بثلاثة خيارات. وقد أرفقت الصورة الثانية للاختبار في شكلها النهائي بالملحق رقم ١ - ٢.

رابعاً: تطبيق وتصحيح الاختبار (الصورة الثانية):

أعيد تطبيق الاختبار التحصيلي الذي صاري يحتوي على ٤٠ سؤالاً بثلاثة خيارات لكل منها بعد أسبوعين من تاريخ تطبيق الصورة الأولى بأربعة خيارات. وقد حددت الفترة الزمنية بين الإختبارين بأسبوعين أخذاً بالحسبان العوامل التي يمكن أن تؤثر سلباً على معامل ثبات الاختبار كالتعلم، والتذكر، والنسيان، وطلب من المفحوصين إختيار الإجابة الصحيحة لكل فقرة من الفقرات التي سبق لهم التعرف عليها، ومنحوا ٣٠ دقيقة فقط إستطاعوا خلالها إنهاء الإجابة عنها. وتم تصحيح الإجابات بال مفتاح المثقب، ورصدت لكل طالب الدرجة الجديدة المستحقة له.

المعالجة الإحصائية:

أُستخدِمت الأساليب الإحصائية التالية: المتوسط الحسابي، الانحراف المعياري، معامل ارتباط سبيرمان للرتب، معامل الارتباط الثنائي الأصيل، معامل الثبات بمعادلة كرونباخ - ألفا، كما استخدم الباحث U Statistic التي طوّرها ماراسكويو لفحص الفروق بين معاملات الارتباط للعينات المستقلة (Marascuilo, 1966)، وكذلك M Statistic التي وضعها عالما الإحصاء هاكشتين والين (Hakstain & Whalen, 1976) لفحص الفروق بين معاملات الثبات للعينات المستقلة. وحُلِّلت البيانات باستخدام الحزمة الإحصائية للعلوم الاجتماعية SPSS.

عرض نتائج البحث في ضوء الفرضيات:

تم احتساب المتوسطات، والانحرافات المعيارية، وأطوال مدى الدرجات، ومتوسطات للمعدل التراكمي، ومعاملات الصدق التنبؤي، ومعاملات الثبات، ومعاملات الارتباط الثنائي الأصيل لكل صورة من صور الاختبار عند مختلف فئات العينة حسب مستوى تحصيلها، وكذلك عند أفراد العينة ككل، ودونت النتائج في الجدول رقم ٢ - ٥. تلي ذلك اختبار الفرضيات البحثية سالفه الذكر للتأكد من صحتها. وجاءت النتائج كما يلي:

أولاً: نتائج الفرضية الأولى:

تنص هذه الفرضية على ما يلي: «لا توجد فروق ذات دلالة إحصائية بين معاملي الصدق التنبؤي لصورتَي الاختبار بأربعة وثلاثة خيارات لكل فئة ولجميع فئات الطلبة على اختلاف مستوياتها التحصيلية».

جدول رقم ٣

المتوسطات الحسابية والانحرافات المعيارية وأطوال المدى المحسوبة
لصوريّ الاختبار ومتوسطات المعدلات التراكمية حسب مستوى تحصيل الفئة

عدد أفراد المجموعة	متوسط المعدل التراكمي	ثلاثة خيارات			أربعة خيارات			الصورة
		المتوسط	الانحراف	طول	المتوسط	الانحراف	طول	الفئة
		الحسابي	المعياري	المدى	الحسابي	المعياري	المدى	
٤٨	١,٩٠	١١	٥,٥٨	٢٢,٠٠	١٥	٦,٧٦	٢١,٤١	الدنيا
٤٨	٢,٥٦	١٢	٥,٧٨	٢٤,٢٥	٩	٥,٤٦	٢٤,١٣	الوسطى
٤٨	٣,١٨	١٠	٥,٥٢	٢٧,٥٩	١٤	٧,٠٦	٢٦,٩٤	العليا
١٤٤	٢,٤٣	١٤	٦,٨٨	٢٤,٤٢	٢٠	٧,٨٦	٢٤,٣٦	الجميع

وللتحقق من هذه الفرضية تم حساب معامل الارتباط بين درجات أفراد العينة على اختلاف مستوياتهم التحصيلية في كل صيغة من صيغتي الاختبار من جهة، ودرجاتهم النهائية في المحك، من جهة أخرى. وقد تم اختيار الدرجة النهائية للطالب في المقرر الدراسي الذي وُضع الاختبار لقياس درجة تحصيل الطالب فيه محكاً لاستخراج الصدق التنبؤي للاختبار. ويُقصد هنا بالدرجة النهائية مجموع الدرجات المستحقة للطالب في المقرر.

وأخذاً بالحسبان وجود بعض العوامل المؤثرة في صلاحية المحك، ووجود قدر من الخطأ أثناء تحديد درجة الطالب في المحك، فقد تم تصحيح القيم المستخرجة لمعاملات الصدق التنبؤي باستخدام معادلة التصحيح Formula for Correction for Attenuation عن طريق قسمة معامل الصدق التنبؤي للاختبار على الجذر التربيعي لمعامل ثباته كما يلي:

$$م ص ح = \frac{م ص}{\sqrt{م ث}}$$

حيث أن: م ص = معامل الصدق التنبؤي للاختبار.

م ث = معامل ثبات الاختبار.

م ص ح - معامل الصدق التنبؤي بعد التصحيح .

وقد تم تدوين معاملات الصدق التنبؤي الأصلية والمصححة لكل فئة على حدة، وكذلك لجميع الفئات التحصيلية في الجدول رقم ٢ .

ويُتضح من الجدول أن قيمة معامل الصدق التنبؤي للإختبار أكثر دقة عندما يتم حساب معامل الارتباط لجميع الفئات التحصيلية. وتتفق هذه النتيجة مع تلك التي توصل إليها تريفيسان وزملائه (Trevisan, Sax and Michael, 1991, P. 834)، لكنها لا تتفق مع نتائج دراسة جرين وزملائها بخصوص معامل الارتباط المحسوب لجميع فئات التحصيل (Green, Sax and Michael, 1982, P. 241).

وقد تم إستخراج الفروق بين معاملات الصدق باستخدام U Statistic التي طورها ماراسكويلو (Marascuilo, 1966). وهذه الإحصاءة تتوافق مع توزيع كاي تربيع Chi-square، وبالتالي تتيح الفرصة لإختبار الفرضية الصفرية القائلة بأن معاملات الصدق المحسوبة لعينات مستقلة تكون متساوية، وبوبت النتائج في الجدول رقم ٤ .

جدول رقم ٤

قيم الفروق بين معاملات الصدق والثبات ومتوسطات القدرة التمييزية المحسوبة لصورتي الإختبار ودلالاتها الإحصائية حسب مستوى تحصيل أفراد العينة

	معامل الصدق التنبؤي			معامل الثبات			متوسط القدرة التمييزية		
الصورة	١	٢	كا ^٢	١	٢	كا ^٢	١	٢	كا ^٢
الدنيا	٠,٠٨	-٠,٤٦	٤,٢٢*	٠,٧٩	٠,٧٤	٠,١٠	٠,٣٢	٠,٢٧	٠,٢٤
الوسطى	٠,٠٨	٠,٤٧	٤,٢٣ ^٥	٠,٧٣	٠,٨٠	٠,٣٨	٠,٣٣	٠,٢٩	٠,٠٤
العليا	٠,١٦	٠,٤٦	٢,٢٥	٠,٧٨	٠,٧٠	٠,٧٢	٠,٢١	٠,٢٥	٠,٠٦
الجميع	٠,٧٣	٠,٧٩	١,٨٨	٠,٨٠	٠,٧٩	١,٠٩	٠,٣٣	٠,٢٩	٠,١١

* قيمة الفرق المحتسبة دالة إحصائياً عند مستوى ٠,٠٥ لصالح صيغة الخيارات الأربعة.

^٥ قيمة الفرق المحتسبة دالة إحصائياً عند مستوى ٠,٠٥ لصالح صيغة الخيارات الثلاثة.

وقد أبرزت النتائج وجود فروق دالة إحصائية عند مستوى ٠,٠٥ بين معامليّ الصّدق لفئة الطلبة منخفضي التحصيل ($\text{كا}^2 = 4,22$) لصالح صيغة الخيارات الأربعة، ومتوسطي التحصيل ($\text{كا}^2 = 4,32$) لصالح صيغة الخيارات الأربعة، ومتوسطي التحصيل ($\text{كا}^2 = 4,32$) لصالح صيغة الخيارات الثلاثة، وعدم وجود تلك الفروق عند مستوى ٠,٠٥ لفئة مرتفعي التحصيل ($\text{كا}^2 = 2,25$) أو لإجمالي أفراد العينة ($\text{كا}^2 = 1,88$) حيث أن هاتين القيمتين أقل من ٣,٨٤ وهي القيمة المقابلة لكاي تربيع عند مستوى ٠,٠٥ بدرجة حرية: ح = ١.

وتتفق نتائج البحث مع ما خلصت إليه دراسة تريفيسان حول الفروق بين معاملات الصّدق للفئتين الدنيا والعليا، ولجميع الفئات التحصيلية. غير أنها لا تتفق من نتائج هذه الدراسة فيما يتصل بالفئة الوسطى (Trevisan, Sax and Michael, 1991, P. 833). ومما سبق تتحقق صحة الفرضية الأولى جزئياً فيما يتصل بعدم وجود فروق ذات دلالة إحصائية بين معامليّ الصّدق التنبؤي لصورتَي الاختيار بأربعة وثلاثة خيارات للفئة العليا ولجميع فئات الطلبة على اختلاف مستويات التحصيلية.

ثانياً: نتائج الفرضية الثانية:

تنص هذه الفرضية على ما يلي؛ «لا توجد فروق ذات دلالة إحصائية بين معامليّ الثبات لصورتَي الاختبار بأربعة وثلاثة خيارات لكل فئة ولجميع فئات الطلبة على اختلاف مستوياتها التحصيلية».

وللتحقّق من هذه الفرضية تم حساب معامل ثبات درجات أفراد العينة على اختلاف مستوياتهم التحصيلية في كل صيغة من صيغتي الاختبار باستخدام معادلة كرونباخ - ألفا.

وأخذاً في الحسبان اختلاف عدد الخيارات في الصيغتين تم تعديل قيم

معاملات ثبات صيغة الثلاث خيارات وذلك باستخدام معادلة سبيرمان - براون للإختبار المطّول على إعتبار أن هذه الصيغة تكون مكافئة الأربعة خيارات في عدد الفقرات حينما يتم إطالة عدد فقراتها بنسبة ٤ : ٣ ، أي أن :

$$M_p = \frac{p \times M_{th}}{p + (1 - p) \times M_{th}}$$

حيث : M_p = معامل ثبات صيغة الخيارات الثلاثة .

M_{th} = معامل ثبات صيغة الخيارات الأربعة .

$$p = \text{مقدار الإطالة} = 3/4 = 0.75$$

وتم تدوين قيم معاملات الثبات في كلا الصيغتين وكذلك القيم المعدلة في الجدول رقم ٢ .

وقد جرى إستخراج الفروق بين معامليّ الثبات باستخدام M Statistic التي طوّرها كل من هاكشتين و والين (١٩٧٦)، والمشار إليها آنفاً، والتي تتوافق أيضاً مع توزيع كاي تربيع، تحقّق شروط نموذج تحليل التباين بموجب عاملين بآثار عشوائية Two-Factor Random Effects of the ANOVA، وتتيح الفرصة لإختبار الفرضيّة الصفرية بأن معاملات الثبات المحسوبة لعينات مستقلة تكون متساوية. وقد تم تدوين النتائج في الجدول رقم ٤ .

وتبين النتائج عدم وجود فروق دالة إحصائية عند مستوى ٠,٠٥ بين معامليّ ثبات صيغتيّ الإختبار لفئة الطلبة منخفضي التحصيل (كا^٢ ٠,١٠) أو متوسطي التحصيل (كا^٢ ٠,٣٨)، أو مرتفعي التحصيل (كا^٢ ٠,٧٢)، أو لإجمالي أفراد العينة ككل (كا^٢ ١,٠٩)، حيث أن جميع هذه القيم أقل من ٣,٨٤ وهي القيمة المقابلة لكاي تربيع عند مستوى ٠,٠٥ بدرجة حرية : ح = ١ .

وتتفق هذه النتائج مع النتائج التي توصلت إليها دراسة جرين وزملائها حول الفروق بين معاملات الثبات للفئة العليا (Green, Sax and Michael, 1982).

(p. 242)، وكذلك مع النتائج التي خلصت إليها دراسة تريفيسان وزملائه حول الفروق بين معاملات الثبات للفتتين الوسطى والعليا (Trevisan, Sax and Michael, 1991, p. 833) غير أن النتائج لا تتفق مع الدراستين المذكورتين فيما يتصل بالفئة الدنيا، ومع نتائج دراسة جرين الآنف الذكر فيما يتصل بالفئة الوسطى، وبإجمالي أفراد العينة ككل.

ومما سبق تتحقق صحة الفرضية الثانية فيما يتصل بعدم وجود فروق ذات دلالة إحصائية بين معاملي الثبات لصورتَي الاختبار بأربعة وثلاثة خيارات لكل فئة ولجميع فئات الطلبة على اختلاف مستوياتها التحصيلية.

ثالثاً: نتائج الفرضية الثالثة:

تنص هذه الفرضية على ما يلي: «لا توجد فروق ذات دلالة إحصائية بين متوسطات القدرة التمييزية لصورتَي الاختبار بأربعة وثلاثة خيارات لكل فئة ولجميع فئات الطلبة على اختلاف مستوياتها التحصيلية».

وللتحقق من هذه الفرضية تم احتساب معامل الارتباط الثنائي الأصيل بين الدرجة الكلية لأفراد عينة البحث على اختلاف مستوياتهم التحصيلية لكل صيغة من صيغتي الاختبار من جهة، ودرجاتهم في كل فقرة من فقرات الاختبار، من جهة أخرى. بعدها أُحْتُسِب متوسط القدرة التمييزية بإعتباره متوسط معامل الارتباط لفقرات الاختبار، وبوبت النتائج في الجدول رقم ٥. ولإستخراج الفروق بين متوسطي القدرة التمييزية لصيغتي الاختبار تم تطبيق إحصائه U Statistic السالفة البيان، ودونت النتائج في الجدول رقم ٤.

وقد أبرزت النتائج عدم وجود فروق دالة إحصائية عند مستوى ٠,٠٥ بين معاملي الارتباط لفئة الطلبة منخفضي التحصيل ($\alpha^2 = ٠,٢٤$)، أو متوسطي التحصيل ($\alpha^2 = ٠,٠٤$)، أو مرتفعي التحصيل ($\alpha^2 = ٠,٠٦$)، أو لأفراد العينة ككل ($\alpha^2 = ٠,١١$) حيث أن جميع هذه القيم أقل من ٣,٨٤ وهي القيمة المقابلة لكاي تربيع عند مستوى ٠,٠٥ بدرجة حرية: ح ١.

جدول رقم ٥

متوسطات القدرة التمييزية المحسوبة لصورتي الاختبار التحصيلي
والنسب المئوية المقابلة لها حسب مستوى تحصيل الفئة

الصورة		أربعة خيارات			ثلاثة خيارات		
الفئة	متوسط القدرة التمييزية	النسب المئوية المقابلة			متوسط القدرة التمييزية	النسب المئوية المقابلة	
		جيد	مقبول	ضعيف		جيد	مقبول
الدنيا	٠,٣٢	٣٨	٢٠	٤٢	٠,٢٧	٣٦	٢٧
الوسطى	٠,٣٣	٤٢	٢٢	٣٦	٠,٢٩	٤٠	٣٠
العليا	٠,٢١	٤٠	١٨	٤٢	٠,٢٥	٢٠	٣٣
الجميع	٠,٣٣	٢٨	٥٨	١٤	٠,٢٩	٣٠	٤٠

ومما سبق نتحقق صحة الفرضية الثالثة فيما يتصل بعدم وجود فروق ذات دلالة إحصائية بين متوسطات القدرة التمييزية لصورتي الاختبار بأربعة وثلاثة خيارات لكل فئة ولجميع فئات الطلبة على اختلاف مستوياتها التحصيلية.

ثالثاً - نتائج الفرضية الرابعة:

تنص هذه الفرضية على ما يلي: «العدد الأمثل لخيارات فقرة اختبار الاختيار من متعدد لكل فئة ولجميع فئات الطلبة على اختلاف مستوياتهم التحصيلية هو ثلاثة».

وللتحقق من هذه الفرضية تمت مقارنة معاملات الصدق والثبات ومتوسطات القدرة التمييزية المحسوبة لكل صورة من صورة الاختبار التحصيلي ولمختلف المستويات التحصيلية. وحيث أن النتائج بينت وجود فروق دالة إحصائية عند مستوى ٠,٠٥ بين معاملات الصدق التنبؤي للفئة الوسطى لصالح الصورة الثانية، ولم تكشف عن فروق دالة إحصائية بين معاملات الصدق

التنبؤي للفئة العليا ولجميع الفئات التحصيلية، وبين معاملات ثبات الاختبار التحصيلي بأربعة وبثلاثة خيارات، أو بين متوسطات القدرة التمييزية لكل فئة ولجميع الفئات التحصيلية، فإن العدد الأمثل لخيارات فقرة الاختيار من متعدد هو إذن ثلاثة خيارات فقط الأمر الذي يعني تحقق صحة الفرضية الرابعة.

مناقشة وتفسير نتائج البحث:

أولاً: صدق وثبات الاختبار:

تبين نتائج المقارنة بين صورتَي الاختبار بثلاثة وأربعة خيارات عدم وجود فروق دالة إحصائية بين معاملي الصدق التنبؤي لكلا الاختبارين للفئة العليا ولجميع فئات الطلبة على اختلاف مستوياتها التحصيلية. ويعني ذلك أن تقليص عدد الخيارات في اختبار الاختيار من متعدد من أربعة، كما هو في الحالة التقليدية الأكثر شيوعاً، إلى ثلاثة لا يقلل من قدرة الاختبار التنبؤي بمستوى التحصيل الأكاديمي للطلاب في المستقبل بصرف النظر عن مستواه الراهن لحظة إجراء الاختبار، علماً بأن قيمة معامل الصدق التنبؤي تصل إلى حدها الأقصى في حالة احتسابها لجميع الفئات.

غير أننا نلاحظ أن نمط ترتيب القيم المطلقة لمعامل الصدق (أي بغض النظر عن الإشارة) لا يبقى ثابتاً في حالة الانتقال إلى صيغة الخيارات الثلاثة. كما أن ترتيب قيم معاملات الثبات لا تتسق مع النمط المذكور، ناهيك أن اتجاه معامل الارتباط بين الاختبار والمحك يتغير ضمن الفئة الواحدة في حالة تقليص عدد الخيارات من أربعة إلى ثلاثة.

ويمكن تفسير ذلك بعدة عوامل أولها اختلاف طبيعة المتغيرين، حيث أن المتغير الأول، أي درجة الطالب في الاختبار، ترتبط أساساً بمدى إستيعابه للوحدة الدراسية قيد الفحص، وبعوامل ذاتية قد تتدخل في تشكيلها، ويصعب

ضبطها بصورة حاسمة، والحيلولة دون تأثيرها كالتخمين، والتذكر، والنسيان، والتعلّم خلال الفترة الزمنية الفاصلة بين التطبيقين. أما المتغير الثاني، أي درجة الطالب النهائية في المقرر، فهي وحدة مركبة من مجموعة علامات تمثل بوجه عام متوسط أداء الطالب في الإختبارات، والأوراق البحثية، والأنشطة الصفية واللاصفية المرتبطة بالمقرر مثلما تتأثر بمستوى التحصيل الأكاديمي للطلاب وبطريقة تقويم أستاذ المقرر لأدائه. وبعبارة أخرى فإن الحديث يدور عن بعض التحيزات التي تؤثر في صدق الدرجة النهائية للمفحوص بوصفها محكاً بسبب «تأثر الدرجة الخاصة بالمحك بدرجة الإختبار، وعدم تساوي وحدات القياس في المكونات الفرعية، وعدم تناسب أوزانها» (فرج، ١٩٨٩، ص. ٢٧٦). وبالتالي فمن الصعب أن نتوقع توافقاً تاماً بين درجة المفحوص في أي من صور الإختبار ودرجته النهائية في المقرر.

غير أنه من المهم هنا التحقق من دقة التنبؤ بالدرجة النهائية للطلاب في المقرر بمعلومية درجته في الإختبار التحصيلي عند قيمة محددة لمعامل الارتباط بين الدرجتين لجميع الفئات التحصيلية، ومما إذا كانت هذه الدقة تتأثر بعدد خيارات الفقرة. ولغرض تحقيق ذلك تم اشتقاق معاملات الارتباط (معاملات الصدق التنبؤي) المبينة في الجدول رقم ٢ لكل صورة من صور الإختبار، ثم جرى تقسيم المفحوصين إلى ثلاثة مستويات متساوية في طول مدى كل منها وذلك وفقاً لدرجاتهم في الإختبار. بعدها تم احتساب نسبة الطلبة الذين حافظوا على مستواهم الأصلي في الإختبار بحصولهم على درجة نهائية تقع في نفس الفئة التحصيلية المناظرة له من بين الفئات الثلاث التي قُسم الطلبة فيما بعد وفقاً لمستوى الطالب التحصيلي، وتميز هذه النسب بخطوط غامقة اللون في الخانات القطرية للجدول رقم ٦. كما احتُسبت النسب المئوية للطلبة الذين لم يحافظوا على مستواهم الأصلي في الإختبار بحصولهم على درجات نهائية تقع في الفئة التحصيلية المغايرة له، ودونت النتائج في نفس الجدول.

جدول رقم ٦

النسبة المئوية لدقة التنبؤ بوقوع الدرجة النهائية للطالب في المقرر
في الفئة المناظرة لمستواه الأصلي في الاختبار وفقاً لعدد خيارات الفقرة

المستوى	الصورة			أربعة خياراً (ر = ٠,٧٣)			ثلاثة خيارات (ر = ٠,٧٩)		
	النسبة المئوية لدقة التنبؤ بوقوع الدرجة النهائية للطالب في الفئة								
موقع الطالب في الإختبار في المستوى	الأعلى	الوسطى	الدنيا	الأعلى	الوسطى	الدنيا	الأعلى	الوسطى	الدنيا
	٦٣,١٦	٣١,٥٨	٥,٢٦	٦٦,٦٧	٢٦,٦٦	٦,٦٧	٦٦,٦٧	٢٦,٦٦	٦,٦٧
	٢٠,٨٣	٢٩,١٧	٥٠,٠٠	٢٨,٥٨	٣٢,١٤	٣٩,٢٨	٢٨,٥٨	٣٢,١٤	٣٩,٢٨
	٠,٠٠	٤٢,٨٦	٥٧,١٤	٠,٠٠	٢٨,٥٧	٧١,٤٣	٠,٠٠	٢٨,٥٧	٧١,٤٣

ويلاحظ من القيم المدوّنة في الخانات القطرية للجدول أن النسبة المئوية لدقة التنبؤ بالدرجة النهائية للطالب تزداد في حالة تقليل عدد خيارات الفقرة. أي أن احتمال وقوع الطالب الحاصل على درجة ما في إختبار منتصف الفصل (من نوع إختيار من متعدد بأربعة خيارات مثلاً) في مقرر دراسي معين في فئة الطلبة منخفضي التحصيل في المقرر نفسه مستقبلاً يزيد في حالة الإكتفاء بثلاثة خيارات فقط. وفي هذه الحالة يلاحظ أيضاً أن احتمال حصول الطالب على درجة متدنية مستقبلاً في الإختبار النهائي تعكس مستواه المنخفض أصلاً في الإختبار يزيد عن الإحتمال المماثل له بالنسبة للطالب ذي المستوى المرتفع في الإختبار.

وحيث أن واقع الممارسات التربوية القائمة في مجال وضع العلامات تؤكد مثل هذه النتيجة، حيث يميل الطلبة ذو الأداء المتدني في الإختبارات الشهرية إلى الحصول على درجات منخفضة في المساق الدراسي المعني أكثر مما يميل أقرانهم ذو الأداء المرتفع للحصول على درجات عالية في المساق نفسه، فإن القدرة

التنبؤية لإختبار الإختيار من متعدد بثلاثة خيارات بدرجة الطالب مستقبلاً تزيد عن مثيلتها لنفس الإختبار بأربعة خيارات.

وقد يبدو ثمة تناقض شكلي يتمثل في عدم التوافق بين قيم النسب المئوية لدقة التنبؤ بالدرجة النهائية للطلاب المدونة في الجدول رقم ٦ وقيم معاملات الارتباط المحتسبة لكل فئة من فئات التحصيل المبوبة في الجدول رقم ٢. غير أنه تجدر الإشارة هنا إلى أن قيمة معامل الارتباط المحسوب لكل فئة لا تتوقف على قيم الدرجات بقدر ما تتوقف على ترتيب هذه الدرجات، وبالتالي فإن احتمال التنبؤ بوقوع الدرجة النهائية للطلاب في المقرر في الفئة التحصيلية المناظرة لموقعة الأصلي في الإختبار قد لا يتوافق مع معامل الارتباط بين درجته في الإختبار ودرجته في المحك.

إلا أن جدوى أي إختبار في التنبؤ «لا يعتمد على الدرجة التي يرتبط بها الإختبار مع مقياس المحك فقط، ولكن أيضاً على كمية المعلومات التي يزودنا بها» (ثورندايك وآخرون، ١٩٨٩، ص. ٦٢). ومن ضمن هذه المعلومات معاملات الارتباط بين صورتى الإختبار لجميع فئات التحصيل، ولكل فئة على حدة (جدول رقم ٢).

ويتضح من الجدول أن قيمة معامل الارتباط بين درجات الطلبة في إختبار الإختيار من متعدد بأربعة خيارات ومثيلاتها في الإختبار ذي الخيارات الثلاثة دالة إحصائياً عند مستوى ٠,٠٥ لجميع الفئات، ولكل فئة على حدة. ومع أنه لا يجوز تفسير معامل الارتباط بمنطق السببية الحتمية، إلا أن القيمة المحتسبة لجميع فئات الطلبة (+ ٠,٦٧) تعني أن نحو ٥٠٪ من التباين في درجاتهم في الصورة الثانية للإختبار يمكن أن يُعزى إلى تباين الدرجات في الصورة الأولى للإختبار.

وينبغي أثناء تفسير قيم معاملات الارتباط بين صورتى الإختبار الأخذ بعين

الإعتبار حقيقة أن «السقف النظري للإرتباط بين مقياسين هو الجذر التربيعي لحاصل ضرب معامليّ ثباتهما» (ثورندايك وآخرون، ١٩٨٩، ص. ٨٥). ويمكن التحقق من صحة ذلك من خلال البيانات المدوّنة في الجدول رقم ٢.

وهكذا فإن وجود معامل إرتباط موجب ودال إحصائياً بين درجات الطلبة في كل صورة من صور الإختبار يسمح لنا بالقول بأن تقليل عدد الخيارات في أسئلة الإختيار من متعدد من أربعة إلى ثلاثة لا يؤثر سلباً بوجه عام على مستوى أداء الطالب في الإختبار، الأمر الذي تؤكده أيضاً حقيقة عدم وجود فروق ذات دلالة إحصائية عند مستوى ٠,٠٥ بين معامليّ صدق صورتيّ الإختبار (جدول رقم ٤).

وأخذاً بالحسبان طبيعة العلاقة بين صدق وثبات الإختبار نستطيع القول بأن القدرة التنبؤية للإختبار تتحدد بتأثير ثبات مقياس المحك أكثر مما هو بتأثير مقياس التنبؤ، ومن الصعب تحديد قيمة معامل ثبات الدرجات النهائية للطلبة وذلك للعوامل المشار إليها أعلاه.

وعلى الرغم من عدم وجود فروق دالة إحصائياً بين معاملات الثبات المحسوبة بطريقة كرونباخ - ألفا لكل صورة من صور الإختبار، إلا أن الجدول رقم ٤ يبيّن أن معامل الثبات يقلّ بالنسبة للفئتين الدنيا والعليا، بينما ترتفع قيمته بالنسبة للفئة الوسطى في حالة الإقتصار على ثلاثة خيارات فقط بدلاً من أربعة. ويمكن تفسير ذلك بعدة عوامل نذكر منها تجانس أفراد المجموعة من حيث القدرات الذهنية، وعدم إمكانية ضبط العوامل المؤثرة على هذه القدرات كالتخمين والحرز وغيرها. لذا يمكننا أن نعزو إنخفاض قيمة معامل الثبات بالنسبة للفئتين الدنيا والعليا إلى تقلص طول مدى درجات كل منهما بسبب تجانس أداء أفراد هاتين الفئتين في إختبار الإختيار من متعدد بثلاثة خيارات، وإزدیاد هذا الطول بالنسبة للفئة الوسطى نتيجة لتباين مستويات القدرة لدى أفرادها في هذا الإختبار (جدول رقم ٣).

ثانياً: القدرة التمييزية للاختبار:

يتضح من مقارنة متوسطي القدرة التمييزية لصورتَي الاختبار أن متوسط القدرة التمييزية للاختبار الاختيار من متعدد بثلاثة خيارات لا يقل عن مثيله للاختبار بأربعة خيارات مع أن متوسطات القدرة التمييزية أكبر من حيث القيمة في حالة الخيارات الأربعة (فيما عدا قيم المتوسطات بالنسبة لأفراد المجموعة العليا). وهذه الحقيقة تفسّر إلى حد ما انخفاض قيم معاملات الثبات عند تقليص أعداد خيارات الفقرة إلى ثلاثة (جدول رقم ٢).

ولغرض المفاضلة بين متوسطات القدرة التمييزية في حالة أربعة خيارات وثلاثة خيارات تم الحكم على جودة متوسطات القدرة التمييزية لصورتَي الاختبار ولفئات التحصيل المختلفة في ضوء المعايير التالية (Trevusan, Sax and Michael, 1991, p. 834): أكثر من ٠,٢٩ - مؤشر تمييز جيّد.

أكثر من ٠,١٠ إلى ٠,٢٩ - مؤشر تمييز مقبول.

أقل من ٠,١٠ - مؤشر تمييز ضعيف.

وتأسيساً على ذلك تم حساب النسب المئوية المقابلة لكل نوع من أنواع المؤشرات السالفة الذكر لكل صورة من صور الاختبار ولمختلف الفئات التحصيلية، وتفرغ النتائج في الجدول رقم ٥. ويتبيّن من هذا الجدول أن نسب مؤشرات التمييز الجيدة لفقرة الأربعة خيارات تكاد تبقى ثابتة في حالة تقليص عدد الخيارات إلى ثلاثة باستثناء النسب المحسوبة لأفراد المجموعة العليا، حيث تنقلص إلى النصف بسبب تضاعف نسبة مؤشرات التمييز المقبولة، وفقاً للمعيار المذكور.

وهذه النتيجة تؤكد صحة ما توصل إليه البحث بأن اختبار الاختيار من متعدد بثلاثة خيارات لا يقل من حيث قدرته على التمييز بين المفحوصين على اختلاف مستوياتهم التحصيلية عن مثيله ذي الخيارات الأربعة.

خلاصة البحث:

تبين نتائج البحث الحالي بوجه عام عدم وجود فروق جوهرية بين صدق وثبات الاختبار وقدرته التمييزية لصيغتي الاختيار من متعدد بأربعة وبثلاثة خيارات. كما أن الاختبار ذي الخيارات الثلاثة أبرز قدرة تنبؤية أكبر بدرجة الطالب في مقياس المحك في فترة زمنية لاحقة قياساً لمثيله ذي الخيارات الأربعة بصرف النظر عن مستوى تحصيل الطالب. ويعني ذلك أن العدد الأمثل لخيارات فقرة الاختيار من متعدد هو ثلاثة وليس أربعة حيث أن الخصائص السيكمترية للاختبار لا تتأثر عند تقليص عدد الخيارات إلى ثلاثة.

وهذا الاستنتاج يدحض ما يكاد أن يكون شبه إجماع في أوساط المشتغلين بالقياس والتقويم التربوي في البلاد العربية بأن عدد خيارات السؤال من نوع الاختيار من متعدد يجب ألا يقل عن أربعة. ومن ناحية أخرى فإن من شأن بناء أسئلة الاختيار من متعدد بثلاثة خيارات فقط أن يوفر الوقت والجهد الذي يبذله المعلمون في بناء هذا النمط من الأسئلة بخيارات أربعة فأكثر، ناهيك عن إمكانية تغطية قدر أكبر من محتوى المادة الدراسية المراد معرفة مدى إتقان التلاميذ لها في زمن إختباري أقصر.

توصيات البحث:

في ضوء نتائج البحث الراهن يوصي الباحث بما يلي:

- ١ - الإكتفاء بثلاثة خيارات فقط أثناء بناء أسئلة الاختيار من متعدد لغرض قياس مدى تحصيل الطلبة في المواد الدراسية قيد الفحص سواء في الاختبارات الفصلية أو النهائية.
- ٢ - إجراء دراسات مماثلة لخصائص أخرى لأسئلة الاختيار من متعدد بحيث تُطبق على عينات من المفحوصين أكثر عدداً، وتضم كما أكبر من الأسئلة في إطار مسابقات دراسية أخرى وذلك بهدف تدعيم نتائج البحث الحالي بما يكفل إمكانية تعميمها على نطاق أوسع.

ملاحق البحث

ملحق رقم ١ - ١ : الإختبار التحصيلي (الصورة الأولى)

الإسم : _____
 الرقم الجامعي : _____
 الكلية : _____
 السنة الدراسية : _____
 المعدل التراكمي : _____
 الجنس : _____

عزيزتي الطالبة، عزيزي الطالب

تحية طيبة وبعد،

فهذا الإختبار يهدف إلى قياس درجة تحصيلك في وحدة «إعداد الإختبارات التحصيلية» التي درستها ضمن مقرّر القياس والتقويم التربوي في هذا الفصل. فالرجاء التكرم بقراءة تعليمات الإختبار المدونة أدناه بدقة قبل البدء في الإجابة عن الأسئلة.

مع خالص الشكر والتقدير لتعاونك معنا، ،

تعليمات الإختبار :

- ١ - إقرأ كل سؤال بحرص وتأني قبل الإجابة عنه.
- ٢ - تأكد من المطلب الدقيق المحدد في متن السؤال.
- ٣ - إختار الخيار الذي تعتقد أنه أفضل الخيارات المطروحة.
- ٤ - دوّن رمز الخيار الذي تختارته في ورقة الإجابة المرفقة.
- ٥ - حاول الإجابة عن جميع الأسئلة في زمن لا يتجاوز ٤٠ دقيقة.

أسئلة الإختبار التحصيلي (الصورة الأولى : أربعة خيارات)

- ١ - ما هي الخطوة الأولى لبناء الإختبار التحصيلي؟
 - أ - تحديد الغرض من الإختبار.
 - ب - تحديد عدد أسئلة الإختبار.
 - ج - إختيار شكل أسئلة الإختيار.
 - د - صياغة الأهداف السلوكية.
- ٢ - أي من العبارات التالية تعبر عن مفهوم «تقويم التحصيل»؟
 - أ - الكشف عن نواحي الضعف لدى المتعلم في وحدة دراسية معينة
 - ب - تحديد القدرات الكامنة عند المتعلم لتعلم مساق دراسي محدد.
 - ج - تقرير درجة إمتلاك الطالب للمعارف والمهارات ضمن المادة الدراسية.
 - د - تمييز الخصائص والقدرات اللازمة لتقرير النجاح المستقبلي للمتعلم.
- ٣ - أحد الإعتبارات الهامة في بناء الإختبار التحصيلي الصفي يتلخص في :
 - أ - إعداد أسئلة يتطلب حلها مهارات عقلية عليا لدى الطالب.
 - ب - تطوير صيغتين متطابقتين من نفس الإختبار.
 - ج - مطابقة الأسئلة بالأهداف السلوكية المراد قياسها
 - د - إختيار فقرات تتناسب مع مستوى الطلبة الممتحنين.
- ٤ - تكون الأهداف التعليمية أكثر فائدة عندما :
 - أ - تكون مناسبة لجميع أفراد الصف.
 - ب - تربط بين المهارات المطلوب إتقانها وأهداف الأداء.
 - ج - تكون محددة بالنسبة للمعرفة التي تتضمنها.
 - د - يمكن استخدامها أساساً لبناء إختبارات مقننة.

٥ - تحقق الأهداف التعليمية أغراض الإختبار التحصيلي عندما تُصاغ بطريقة يظهر فيها:

- أ - المحتوى الدراسي .
- ب - سلوك المتعلّم .
- ج - الأنشطة التعليمية
- د - أنواع الأسئلة .

٦ - «يستوعب الطالب أعمال المشرف الإجتماعي». يُعتبر ذلك هدفاً تعليمياً ضعيفاً لأنه:

- ١ - لا تتوفر وسائل لقياسه في ظروف الصف .
- ب - يكتفي بوصف أحد جوانب العملية التعليمية .
- ج - لا يبيّن الظروف التي يتوقع فيها الأداء .
- د - لا يحدّد أداء قابلاً للملاحظة والقياس .

٧ - أي الأفعال التالية يفتقر إلى الوضوح والتحديد؟

- أ - يعدّد
- ب - يفهم .
- ج - يلخّص .
- د - يحسب .

٨ - «أذكر أربعة أسباب أدت إلى بدء حرب الخليج». لكي نصنّف هذا السؤال ضمن تصنيف بلوم وزملائه نحتاج أن نعرف:

- أ - ما تلقاه الطالب من محاضرات وقراءات عن حرب الخليج .
- ب - درجة إدراك الطلبة للشئون الإقليمية والدولية .
- ج - ما حدّده الأستاذ كإجابة نموذجية عن السؤال المذكور .
- د - ما إذا كان الحديث يدور حول حرب الخليج الأولى أو الثانية .

٩ - «يستخلص التلميذ عيوب مقالة صحفية». تصنف هذه العبارة وفقاً لتصنيف بلوم وزملائه في مستوى:

- أ - الفهم.
- ب - التطبيق.
- ج - التحليل.
- د - التقويم.

١٠ - «أن يعرض الطالب الأفكار وفقاً للقواعد النحويّة والبلاغية»، هذا الهدف التعليمي يصنف وفقاً لتصنيف بلوم وزملائه في مستوى:

- أ - الفهم.
- ب - التطبيق.
- ج - التحليل.
- د - التقويم.

١١ - في أي مستوى من مستويات الأهداف التعليمية تقع العبارة التالية بموجب تصنيف بلوم وزملائه: «أن يتنبأ الطالب بآثار تلوث البيئة في منطقة الخليج»؟

- أ - المعرفة.
- ب - الفهم.
- ج - التطبيق.
- د - التقويم.

١٢ - في أي مستوى من مستويات تصنيف بلوم يمكن تصنيف العبارة التالية: «أن يختبر الطالب صحة الفرضيات التي طرحها في بداية البحث»؟

- أ - المعرفة.
- ب - الفهم.
- ج - التطبيق.
- د - التحليل.

١٣ - في أي مستوى من مستويات الأهداف التعليمية تقع العبارة التالية بموجب تصنيف بلوم وزملائه: «أن يصمم التلميذ تجربة عملية لقياس

درجة تلوث ماء البحر»؟

أ - المعرفة .

ب - الفهم .

ج - التطبيق .

د - التركيب .

١٤ - في أي مستوى من مستويات الأهداف التعليمية تقع العبارة التالية بموجب تصنيف بلوم وزملائه: «أن يتعرف الطالب على الأخطاء

النحوية في الجملة المعطاة»؟

أ - المعرفة .

ب - الفهم .

ج - التطبيق .

ج - التحليل .

د - التقويم .

١٥ - ما الميزة الرئيسية لإستخدام جدول المواصفات عند إعداد الإختبار التحصيلي؟

أ - يحدد الأهمية النسبية لكل موضوع من مواضيع الإختبار .

ب - يجعل قياس الهدف السلوكي أمراً سهلاً

ج - يزيد في تحسين التوازن والشمول في أسئلة الإختبار .

د - يحدد العدد النهائي لأسئلة الإختبار وطريقة توزيعها .

١٦ - في أي مما يلي يعتبر جدول المواصفات قليل الفائدة؟

أ - تحديد شكل الأسئلة التي سيتضمنها الإختبار .

ب - الحكم على الأهمية النسبية التي تعطى لمستوى الهدف التعليمي .

ج - تحديد عدد الأسئلة التي تقيس مستوى معيناً من الأهداف التعليمية .

د - ربط مضمون السؤال بالهدف السلوكي الذي يقيسه .

- ١٧ - ما هو الأصح أثناء استخدام الأسئلة المقالية والموضوعية في تقييم التحصيل؟
- أ - يتوقف إختيار شكل سؤال على أهداف التقييم.
- ب - يجب استخدام الأسئلة المقالية فقط لأنها تسمح بحرية التعبير.
- ج - يجب استخدام الأسئلة الموضوعية فقط لأنها أكثر صدقاً وثباتاً.
- د - يجب استخدام الأسئلة المقالية والموضوعية معاً في نفس الإختبار.
- ١٨ - الإختبار المقالي أكثر ملاءمة من الإختبار الموضوعي في قياس:
- أ - معرفة الحقائق والمفاهيم.
- ب - تنظيم الأفكار والمعلومات.
- ج - إستيعاب المبادئ الأساسية.
- د - تطبيق المعارف والمفاهيم.
- ١٩ - ما أبرز ما يميز الإختبارات الموضوعية؟
- أ - تغطي أسئلتها قسماً كبيراً من المقرر الدراسي.
- ب - يخضع إعداد أسئلتها لقواعد محددة.
- ج - لا تتدخل ذاتية المعلم أثناء رصد الدرجات.
- د - ترتبط الأسئلة مباشرة بالأهداف التعليمية المراد قياسها.
- ٢٠ - أي مما يلي يُعتبر أبرز عيوب إختبارات المقال؟
- أ - تشجيعها على الحفظ الغيبي للحقائق والمفاهيم.
- ب - تحفيزها للطلاب للتحايل على الإجابة التي يجملها.
- ج - مساهمتها في إطلاق حرية الطالب في الإجابة.
- د - تأثر درجاتها بعوامل خارجة عن الإجابة.
- ٢١ - لأي سبب لا يفضل بناء الإختبار الصفي بالكامل من أسئلة مقالية؟
- أ - عينة المحتوى المراد تغطيتها تكون محدودة.
- ب - تصحيح هذه الأسئلة يتطلب وقتاً طويلاً.
- ج - زمن الحصة الدراسية لا يكفي لحل جميع الأسئلة.
- د - بناء نماذج للإجابة يكلف الوقت والجهد الكثير.

٢٢ - لأي سبب مما يلي لا يفضل السماح للطلاب بالإختيار في أسئلة الإختبار
المقالي؟

- أ - لا يمنح الإختيار للطلاب الفرصة لإظهار قدراتهم الفعلية .
 - ب - يفسح المجال لتدخل الذاتية أثناء التصحيح .
 - ج - يقلل من قابلية مقارنة درجات الطلبة في الإختبار بعضهم ببعض .
 - د - يضيع وقت الطالي في تحديد السؤال الذي يرغب في الإجابة عنه .
- ٢٣ - أي أشكال الأسئلة التالية تغطي عينة أكبر من محتوى المادة في زمن
قصير؟

- أ - الإجابة القصيرة .
 - ب - الصواب والخطأ .
 - ج - الاختيار من متعدد .
 - د - المزاوجة .
- ٢٤ - أي من أنواع الأسئلة التالية هو الأقل موضوعية؟

- أ - الصواب والخطأ .
 - ب - الإختيار من متعدد .
 - ج - المزاوجة .
 - د - الإكمال .
- ٢٥ - أي أشكال الأسئلة التالية تُعتبر الأفضل في رصد سوء فهم الطالب
للحقائق؟

- أ - الإجابة القصيرة .
- ب - الصواب والخطأ .
- ج - الإختيار من متعدد .
- د - المقال .

٢٦ - في أيّ من الأسئلة التالية يحصل التلميذ على أعلى علامة عبر التخمين العشوائي؟

أ - الصواب والخطأ.

ب - الاختيار من متعدد.

ج - المزاوجة.

د - الإكمال.

٢٧ - أي أشكال الأسئلة التالية تُعتبر الأفضل في قياس تعرف الحقائق وتمييزها؟

أ - الاختيار من متعدد.

ب - الإجابة القصيرة.

ج - الإكمال.

د - المقال.

٢٨ - أي أشكال الأسئلة التالية تُعتبر الأفضل إستخداماً في قياس التحصيل؟

أ - التكملة والإجابة القصيرة.

ب - الصواب والخطأ.

ج - الاختيار من متعدد.

د - المقال.

٢٩ - في أي مما يلي تتفوق أسئلة الإجابة القصيرة على أسئلة الاختيار من متعدد؟

أ - تميّزها بمعامل ثبات أكبر.

ب - القدرة على التمييز أكثر بين المفحوصين.

ج - تقليل فرصة التخمين العشوائي إلى الحد الأدنى.

د - إمكانية تغطيتها لعينة أوسع من محتوى المادة.

٣٠ - تقل إمكانية التخمين العشوائي في حالة الإجابة عن سؤال الاختيار من متعدد حينما:

أ - يزداد طول الخيار.

ب - يقل عدد الخيارات.

ج - يرتفع عدد الخيارات الخاطئة.

د - يزداد تجانس الخيارات.

٣١ - أي مما يلي لا يفضل إتباعه عند بناء أسئلة الصواب والخطأ؟

أ - جعل عبارات الصواب أكثر من عبارات الخطأ.

ب - حصر كل سؤال في فكرة واحدة.

ج - تجنب استخدام عبارات النفي قدر الإمكان.

د - الإمتناع عن استخدام عبارات طويلة ومعقدة.

٣٢ - لأي من نواتج التعليم التالية يكون السؤال من نوع الاختيار من متعدد أقل ملاءمة؟

أ - تعريف مفهوم أساسي.

ب - التعرف على سبب حدوث ظاهرة معينة.

ج - تقديم أمثلة ملموسة عن مجالات توظيف طريقة ما.

د - إختيار العنوان الأنسب للنص المعطى.

٣٣ - أي مما يلي لا يفضل إتباعه عند بناء أسئلة الاختيار من متعدد؟

أ - جعل كل الخيارات بنفس الطول.

ب - إنتقاء خيارات مرتبطة بمحتوى الأرومة.

ج - استخدام نفي النفي في متن السؤال.

د - توحيد عدد الخيارات في كل فقرات الاختبار.

٢٤ - أي مما يلي يُعتبر من أوجه القصور في اختبار المزاوجة؟

أ - إفتقاره إلى الموضوعية والدقة في القياس.

ب - صعوبة استخدام أسئلته في معظم المساقات الدراسية.

ج - وجود عدد أكبر من البدائل في قائمة الإستجابات.

د - محدودية مجال استخدامه لقياس المهارات العقلية العليا.

- ٣٥ - بأيّ مما يلي يمكن تقليل النجاح عن طريق الصدفة في أسئلة المزاجية؟
- أ - استخدام عبارات تتطلب تمييزاً دقيقاً.
- ب - جعل عدد الإستجابات أكثر من المقدمات.
- ج - عدم السماح باستخدام الإجابات أكثر من مرة.
- د - إطالة عبارات المقدمات والإستجابات.
- ٣٦ - الام تفتقر صياغة السؤال على النحو التالي: قارن بين «عصبة الأمم» و«هيئة الأمم المتحدة»؟
- أ - التناسق اللغوي.
- ب - بلاغة التعبير.
- ج - تحديد المطلب.
- د - الوضوح اللغوي.
- ٣٧ - ما هو العيب في صياغة السؤال التالي من نوع الصواب والخطأ: «خالد بن الوليد، بطل معركة اليرموك، مدفون في مدينة حمص»؟
- أ - تسهّل الإجابة عليه.
- ب - يشتمل على أكثر من فكرة.
- ج - أخذ حرفياً من الكتاب المدرسي.
- د - يحتوي على جملة اعتراضية.
- ٣٨ - ما هو العيب في صياغة السؤال التالي من نوع الصواب والخطأ: «أحلامنا تعبير عن رغباتنا المكبوتة وإحباطاتنا»؟
- أ - يعبر السؤال عن رأي معين.
- ب - تتسم فكرته بالطابع العام.
- ج - تتضمن صياغته أخطاء نحوية.
- د - يحتوي على مفاهيم معقدة.

٣٩ - أي مما يلي يعتبر ضرورياً لتطبيق الإختبار بفاعلية؟

- أ - وضوح ودقة التعليمات .
 - ب - عدم التقيّد بزمن محدد للتطبيق .
 - ج - منع طرح الأسئلة أثناء الإختبار .
 - د - زيادة عدد المراقبين لكل طالب .
- ٤٠ - يمكن تقليص قلق الإختبار لدى الطالب عن طريق .
- أ - تقليل الوقت المخصّص للإختبار ما أمكن .
 - ب - إعطاء سلسلة إختبارات قصيرة بصورة دورية .
 - ج - رصد درجات الممتحنين قبل إنتهاء زمن التطبيق .
 - د - إكساب الطالب مهارات جيدة لتقديم الإختبار .

ورقة الإجابة

الرقم	أ	ب	ج	د	الرقم	أ	ب	ج	د
١					١٢				
٢					٢٢				
٣					٢٣				
٤					٢٤				
٥					٢٥				
٦					٢٦				
٧					٢٧				
٨					٢٨				
٩					٢٩				
١٠					٣٠				
١١					٣١				
١٢					٣٢				
١٣					٣٣				
١٤					٣٤				
١٥					٣٥				
١٦					٣٦				
١٧					٣٧				
١٨					٣٨				
١٩					٣٩				
٢٠					٤٠				

ملحق رقم ١ - ٢ : الإختبار التحصيلي (الصورة الثانية)

الإسم : الرقم الجامعي :
 الكلية : السنة الدراسية :
 المعدل التراكمي : الجنس :

عزيزتي الطالبة، عزيزي الطالب

تحية طيبة وبعد

فهذا الإختبار يهدف إلى قياس درجة تحصيلك في وحدة «إعداد الإختبارات التحصيلية» التي درستها ضمن مقرر القياس والتقويم التربوي في الفصل . فالرجاء التكرم بقراءة تعليمات الإختبار المدونة أدناه بدقة قبل البدء في الإجابة عن الأسئلة .

مع خالص الشكر والتقدير لتعاونك معنا،

تعليمات الإختبار:

- ١ - إقرأ كل سؤال بحرص وتأني قبل الإجابة عنه .
- ٢ - تأكد من المطلب الدقيق المحدد في متن السؤال .
- ٣ - إختار الخيار الذي تعتقد أنه أفضل الخيارات المطروحة .
- ٤ - دوّن رمز الخيار الذي تختاره في ورقة الإجابة المرفقة .
- ٥ - حاول الإجابة عن جميع الأسئلة في زمن لا يتجاوز ٣٠ دقيقة .

أسئلة الاختبار التحصيلي الصورة الثانية : ثلاثة خيارات

- ١ - ما هي الخطوة الأولى لبناء الإختبار التحصيلي؟
 - أ - تحديد الغرض من الإختبار .
 - ب - تحديد عدد أسئلة الإختبار .
 - ج - صياغة الأهداف السلوكية .
- ٢ - أيّ من العبارات التالية تعبر عن مفهوم «تقويم التحصيل»؟
 - أ - الكشف عن نواحي الضعف لدى المتعلّم في وحدة دراسية معيّنة .
 - ب - تحديد القدرات الكامنة عند المتعلّم لتعلّم مساق دراسي محدّد .
 - ج - تقرير درجة إمتلاك الطالب للمعارف والمهارات ضمن المادة الدراسية .
- ٣ - أحد الإعتبارات الهامة في بناء الإختبار التحصيلي الصفي يتلخص في :
 - أ - إعداد أسئلة يتطلب حلّها مهارات عقلية عليا لدى الطالب .
 - ب - مطابقة الأسئلة بالأهداف السلوكية المراد قياسها .
 - ج - إختيار فقرات تتناسب مع مستوى الطلبة الممتّحين .
- ٤ - تكون الأهداف التعليمية أكثر فائدةً عندما :
 - أ - تكون مناسبةً لجميع أفراد الصف .
 - ب - تربط بين المهارات المطلوب إتقانها وأهداف الأداء .
 - ج - تكون محدّدةً بالنسبة للمعرفة التي تتضمّنها .
- ٥ - تحقّق الأهداف التعليمية أغراض الإختبار التحصيلي عندما تُصاغ بطريقة يظهر فيها :
 - أ - المحتوى الدراسي .
 - ب - سلوك المتعلّم .
 - ج - أنواع الأسئلة .

٦ - «يستوعب الطالب أعمال المشرف الإجتماعي». يعتبر ذلك هدفاً تعليمياً ضعيفاً لأنه:

- أ - لا تتوفر وسائل لقياسه في ظروف الصف.
- ب - لا يبيّن الظروف التي يتوقع فيها الأداء.
- ج - لا يحدد أداءاً قابلاً للملاحظة والقياس.

٧ - أي الأفعال التالية يفتقر إلى الوضوح والتحديد؟

- أ - يعدّد.
- ب - يفهم.
- ج - يحسب.

٨ - «أذكر أربعة أسباب أدت إلى بدء حرب الخليج». لكي نصنّف هذا السؤال ضمن تصنيف بلوم وزملائه نحتاج أن نعرف:

- أ - ما تلقاه الطالب من محاضرات وقراءات عن حرب الخليج.
- ب - ما حدّده الأستاذ كإجابة نموذجية عن السؤال المذكور.
- ج - ما إذا كان الحديث يدور حول حرب الخليج الأولى أو الثانية.

٩ - «يستخلص التلميذ عيوب مقالة صحفية». تُصنّف هذه العبارة وفقاً لتصنيف بلوم وزملائه في مستوى:

- أ - الفهم.
- ب - التحليل.
- ج - التقويم.

١٠ - «أن يعرض الطالب الأفكار وفقاً للقواعد النحويّة والبلاغية». هذا الهدف التعليمي يُصنّف وفقاً لتصنيف بلوم وزملائه في مستوى:

- أ - الفهم.
- ب - التطبيق.
- ج - التحليل.

١١ - في أي مستوى من مستويات الأهداف التعليمية تقع العبارة التالية بموجب تصنيف بلوم وزملائه: «أن يتنبأ الطالب بأثار تلوث البيئة في

منطقة الخليج»؟

أ - المعرفة .

ب - الفهم .

ج - التقويم .

١٢ - في أي مستوى من مستويات تصنيف بلوم يمكن تصنيف العبارة التالية: «أن يختبر الطالب صحة الفرضيات التي طرحها في بداية البحث»؟

أ - الفهم .

ب - التطبيق .

ج - التحليل .

١٣ - في أي مستوى من مستويات الأهداف التعليمية تقع العبارة التالية بموجب تصنيف بلوم وزملائه: «أن يصمم التلميذ تجربةً عمليةً لقياس درجة تلوث ماء البحر»؟

أ - الفهم .

ب - التطبيق .

ج - التركيب .

١٤ - في أي مستوى من مستويات الأهداف التعليمية تقع العبارة التالية بموجب تصنيف بلوم وزملائه: «أن يتعرّف الطالب على الأخطاء النحويّة في الجملة المعطاة»؟

أ - التطبيق .

ب - التحليل .

ج - التقويم .

١٥ - ما الميزة الرئيسية لإستخدام جدول المواصفات عند إعداد الإختبار التحصيلي؟

- أ - يحدد الأهمية النسبية لكل موضوع من مواضيع الإختبار.
- ب - يزيد في تحسين التوازن والشمول في أسئلة الإختبار.
- ج - يحدد العدد النهائي لأسئلة الإختبار وطريقة توزيعها.

١٦ - في أي مما يلي يعتبر جدول المواصفات قليل الفائدة؟

- أ - تحديد شكل الأسئلة التي سيتضمنها الإختبار.
- ب - تحديد عدد الأسئلة التي تقيس مستوى معيناً من الأهداف التعليمية.
- ج - ربط مضمون السؤال بالهدف السلوكي الذي يقيسه.

١٧ - ما هو الأصح أثناء إستخدام الأسئلة المقالية والموضوعية في تقييم التحصيل؟

- أ - يتوقف إختيار شكل السؤال على أهداف التقييم.
- ب - يجب إستخدام الأسئلة الموضوعية فقط لأنها أكثر صدقاً وثباتاً.
- ج - يجب إستخدام الأسئلة المقالية والموضوعية معاً في نفس الإختبار.

١٨ - الإختبار المقالي أكثر ملاءمة من الإختبار الموضوعي في قياس:

- أ - معرفة الحقائق والمفاهيم.
- ب - تنظيم الأفكار والمعلومات.
- ج - تطبيق المعارف والمفاهيم.

١٩ - ما أبرز ما يميز الإختبارات الموضوعية؟

- أ - تغطي أسئلتها قسماً كبيراً من المقرر الدراسي.
- ب - لا تتدخل ذاتية المعلم أثناء رصد الدرجات.
- ج - ترتبط الأسئلة مباشرة بالأهداف التعليمية المراد قياسها.

٢٠ - أي مما يلي يُعتبر أبرز عيوب إختبارات المقال؟

- أ - تشجيعها على الحفظ الغيبي للحقائق والمفاهيم.

- ب - تحفيزها للطلاب للتحايل على الإجابة التي يجهلها.
- ج - تأثر درجاتها بعوامل خارجة عن الإجابة.
- ٢١ - لأي سبب لا يفضل بناء الإختبار الصفي بالكامل من أسئلة مقالية؟
- أ - عينة المحتوى المراد تغطيتها تكون محدودة.
- ب - تصحيح هذه الأسئلة يتطلب وقتاً طويلاً.
- ج - زمن الحصة الدراسية لا يكفي لحل جميع الأسئلة.
- ٢٢ - لأي سبب مما يلي لا يفضل السماح للطلاب بالإختيار في أسئلة الإختبار المقالي؟
- أ - لا يمنح الإختيار للطلاب الفرصة لإظهار قدراتهم الفعلية.
- ب - يقلل من قابلية مقارنة درجات الطلبة في الإختبار بعضهم ببعض.
- ج - يضيع وقت الطالب في تحديد السؤال الذي يرغب في الإجابة عنه.
- ٢٣ - أي أشكال الأسئلة التالية تغطي عينة أكبر من محتوى المادة في زمن قصير؟
- أ - الصواب والخطأ.
- ب - الإختيار من متعدد.
- ج - المزاوجة.
- ٢٤ - أي من أنواع الأسئلة التالية هو الأقل موضوعية؟
- أ - الصواب والخطأ.
- ب - المزاوجة.
- ج - الإكمال.
- ٢٥ - أي أشكال الأسئلة التالية تُعتبر الأفضل في رصد سوء فهم الطالب للحقائق؟
- أ - الصواب والخطأ.
- ب - الإختيار من متعدد.
- ج - المقال.

٢٦ - في أيّ من الأسئلة التالية يحصل التلميذ على أعلى علامة عبر التخمين العشوائي؟

أ - الصواب والخطأ.

ب - الاختيار من متعدد.

ج - المزاوجة.

٢٧ - أي أشكال الأسئلة التالية تُعتبر الأفضل في قياس تعرّف الحقائق وتمييزها؟

أ - الاختيار من متعدد.

ب - الإجابة القصيرة.

ج - الإكمال.

٢٨ - أي أشكال الأسئلة التالية تُعتبر الأفضل إستخداماً في قياس التحصيل؟

أ - التكملة والإجابة القصيرة.

ب - الاختيار من متعدد.

ج - المقال.

٢٩ - في أي مما يلي تتفوق أسئلة الإجابة القصيرة على أسئلة الاختيار من متعدد؟

أ - القدرة على التمييز أكثر بين المفحوصين.

ب - تقليل فرصة التخمين العشوائي إلى الحد الأدنى.

ج - إمكانية تغطيتها لعينة أوسع من محتوى المادة.

٣٠ - تقلل إمكانية التخمين العشوائي في حالة الإجابة عن سؤال الاختيار من متعدد حينما:

أ - يقل عدد الخيارات.

ب - يرتفع عدد الخيارات الخاطئة.

ج - يزداد تجانس الخيارات.

- ٣١ - أي مما يلي لا يفضل إتباعه عند بناء أسئلة الصواب والخطأ؟
 أ - جعل عبارات الصواب أكثر من عبارات الخطأ.
 ب - تجنب استخدام عبارات النفي قدر الإمكان.
 ج - الإمتناع عن استخدام عبارات طويلة ومعقدة.
- ٣٢ - لأي من نواتج التعلم التالية يكون السؤال من نوع الاختيار من متعدد أقل ملاءمة؟
 أ - تعريف مفهوم أساسي.
 ب - تقديم أمثلة ملموسة عن مجالات توظيف طريقة ما.
 ج - إختيار العنوان الأنسب للنص المعطى.
- ٣٣ - أي مما يلي لا يفضل إتباعه عند بناء أسئلة الاختيار من متعدد؟
 أ - إلتفاء خيارات مرتبطة بمحتوى الأرومة.
 ب - استخدام نفي النفي في متن السؤال.
 ج - توحيد عدد الخيارات في كل فقرات الاختبار.
- ٣٤ - أي مما يلي يُعتبره من أوجه القصور في اختبار المزاجية؟
 أ - إفتقاره إلى الموضوعية والدقة في القياس.
 ب - صعوبة استخدام أسئلته في معظم المساقات الدراسية.
 ج - محدودية مجال استخدامه لقياس المهارات العقلية العليا.
- ٣٥ - بأي مما يلي يمكن تقليل النجاح عن طريق الصدفة في أسئلة المزاجية؟
 أ - استخدام عبارات تتطلب تمييزاً دقيقاً.
 ب - جعل عدد الإستجابات أكثر من المقدمات.
 ج - عدم السماح باستخدام الإجابات أكثر من مرة.
- ٣٦ - إلام تفتقر صياغة السؤال على النحو التالي: قارن بين «عصبة الأمم» و «هيئة الأمم المتحدة»؟
 أ - التناقض اللغوي.
 ب - تحديد المطلب.
 ج - الوضوح اللغوي.

- ٣٧ - ما هو العيب في صياغة السؤال التالي من نوع الصواب والخطأ: «خالد بن الوليد، بطل معركة اليرموك، مدفون في مدينة حمص»؟
- أ - يشتمل على أكثر من فكرة.
- ب - أخذ حرفياً من الكتاب المدرسي.
- ج - يحتوي على جملة إعتراضية.
- ٣٨ - ما هو العيب في صياغة السؤال التالي من نوع الصواب والخطأ: «أحلامنا تعبير عن رغباتنا المكبوتة وإحباطاتنا»؟
- أ - يعبر السؤال عن رأي معين.
- ب - تتسم فكرته بالطابع العام.
- ج - يحتوي على مفاهيم معقدة.
- ٣٩ - أي مما يلي يعتبر ضرورياً لتطبيق الإختبار بفاعلية؟
- أ - وضوح ودقة التعليمات.
- ب - عدم التقيّد بزمن محدّد للتطبيق.
- ج - منع طرح الأسئلة أثناء الإختبار.
- ٤٠ - يمكن تقليص قلق الإختبار لدى الطالب عن طريق:
- أ - إعطاء سلسلة إختبارات قصيرة بصورة دورية.
- ب - رصد درجات الممتحنين قبل إنتهاء زمن التطبيق.
- ج - إكساب الطالب مهارات جيّدة لتقديم الإختبار.

ورقة الإجابة

الرقم	أ	ب	ج	د	الرقم	أ	ب	ج	د
١					١٢				
٢					٢٢				
٣					٢٣				
٤					٢٤				
٥					٢٥				
٦					٢٦				
٧					٢٧				
٨					٢٨				
٩					٢٩				
١٠					٣٠				
١١					٣١				
١٢					٣٢				
١٣					٣٣				
١٤					٣٤				
١٥					٣٥				
١٦					٣٦				
١٧					٣٧				
١٨					٣٨				
١٩					٣٩				
٢٠					٤٠				

ملحق رقم ١ - ٣
جدول مواصفات الإختبار التحصيلي

عدد الأسئلة	توزيع أعداد الأسئلة حسب مستويات الأهداف			المواضيع
	التطبيق (٣٠٪)	الفهم (٥٠٪)	المعرفة (٢٠٪)	
١٠	٣	٥	٢	تخطيط الإختيار (٢٥٪)
٦	٢	٣	١	تصنيف الأهداف (١٥٪)
٢٤	٧	١٢	٥	إعداد الفقرات (٦٠٪)
٤٠	١٢	٢٠	٨	المجموع

ملحق رقم ١ - ٤
رموز الإجابات الصحيحة لأسئلة الصورة الأولى للإختبار

١	أ	٦	د	١١	ب	١٦	أ	٢١	أ	٢٦	أ	٣١	أ	٣٦	ج
٢	ج	٧	ب	١٢	ج	١٧	أ	٢٢	ج	٢٧	أ	٣٢	د	٣٧	ب
٣	ج	٨	أ	١٣	د	١٨	ب	٢٣	ب	٢٨	ج	٣٣	ج	٣٨	أ
٤	ب	٩	د	١٤	د	١٩	ج	٢٤	د	٢٩	ج	٣٤	د	٣٩	أ
٥	ب	١٠	ب	١٥	ج	٢٠	د	٢٥	ب	٣٠	د	٣٥	ب	٤٠	د

ملحق رقم ١ - ٥
رموز الخيارات التي تم حذفها من أسئلة الإختبار للصورة الثانية

١	ج	٦	ب	١١	ج	١٦	ب	٢١	د	٢٦	د	٣١	ب	٣٦	ب
٢	د	٧	ج	١٢	أ	١٧	ب	٢٢	ب	٢٧	د	٣٢	ب	٣٧	أ
٣	ب	٨	ب	١٣	أ	١٨	ج	٢٣	أ	٢٨	ب	٣٣	أ	٣٨	ج
٤	د	٩	ب	١٤	أ	١٩	ب	٢٤	ب	٢٩	أ	٣٤	ج	٣٩	د
٥	ج	١٠	د	١٥	ب	٢٠	ج	٢٥	أ	٣٠	أ	٣٥	د	٤٠	أ

المراجع

- ١ - أبو زينة، فريد كامل (١٩٩٢). أساسيات القياس والتقويم في التربية (الطبعة الأولى). الكويت: مكتبة الفلاح.
- ٢ - أبو عودة، أحمد سليمان (١٩٩٣). القياس والتقويم في العملية التدريسية (الطبعة الثانية). إربد: دار الأمل للنشر والتوزيع.
- ٣ - أبو لبدة، سبع محمد (١٩٩٤). مبادئ القياس النفسي والتقويم التربوي (الطبعة الرابعة) عمان: جمعية أعمال المطابع التعاونية.
- ٤ - ثورندايك، روبرت وآخرون (١٩٨٩). القياس والتقويم في علم النفس والتربية. ترجمة: د. عبدالله الكيلاني وآخرون. عمان: مركز الكتب الأردني.
- ٥ - خضر، فخري رشيد (١٩٨٧). التقويم التربوي. دبي: دار القلم للنشر والتوزيع.
- ٦ - عطية، نعيم (١٩٩٢). التقييم التربوي الهادف (الطبعة الثالثة). بيروت: الشركة العالمية للكتاب.
- ٧ - فرج، صفوت (١٩٨٩). القياس النفسي (الطبعة الثانية). القاهرة: مكتبة الأنجلو المصرية.
- ٨ - وهبه، نخلة إلياس (١٩٩٦). قضايا تربوية بحرينية. البحرين: المطبعة الحكومية.

المصادر الأجنبية:

- 9 - Bloom, B.S. (Ed.) (1956). **Taxonomy of Educational Objectives: Cognitive Domain**. NY: David Mckay.
- 10 - Costin, F. (1970). The optimal number of alternatives in multiple choice achievement tests: Some empirical evidence

- for a mathematical proof. **Educational and Psychological Measurement**, 30, 353 - 358.
- 11 - Costin, F. (1972). Three choice versus four choice items: Implications for validity and reliability of objective achievement tests. **Educational and Psychological Measurement**, 32, 1035 - 1038.
 - 12 - Crehan, K., Haladyna, T., and Brewer, B. (1993) Use of an inclusive option and the optimal number of options for multiple - choice items.. **Educational and Psychological Measurement**, 53, 241 - 247.
 - 13 - Green, K., Sax, G., and Michael, W.B. (1982). Validity and reliability of tests having different number of options for students of differing levels of ability. **Educational and Psychological Measurement**, 42, 239 - 245.
 - 14 - Grier, J.B. (1975). The number of alternatives for optimum test reliability. **Journal of Educational Measurement**, 12, 109 - 113
 - 15 - Gronlun, N.E. and Linn, R.L. (1990). **Measurement and Evaluation in Teaching**. New York: Macmillan Publishing Co., Inc.
 - 16 - Hakstian, A.R. and Whalen T.E. (1976). A k-sample significance test for independent alpha coefficients. **Psychometrika**, 41, 219 - 231.
 - 17 - Hopkins, C.D. and Antes, R.L. (1978). **Classroom Measurement and Evaluation**. Itasca: F.E. Peacock Publishers, Inc.
 - 18 - Lord, F.M. (1977). Optimal number of choices per item a comparison of four approaches. **Journal of Educational Measurement**, 14, 33-38.
 - 19 - Marascuilo, L. (1966). Large-sample multiple comparisons. **Psychological Bulletin**, 65, 280 - 290.
 - 20 - Tversky, A. (1964) On the optimal number of alternatives at a choice point. **Journal of Mathematical Psychology**, 1, 386 - 391.

- 21 - Trevisan, M.S., Sax, G., and Michael, W.B. (1991). The effects of the number of options per item and student ability on test validity and reliability. **Educational and Psychological Measurement**, 51, 829 - 837.
- 22 - Trevisan, M.S., Sax, G., and Michael, W.B. (1994). Estimating the optimum number of options per item using an incremental option paradigm. **Educational and Psychological Measurement**, 54, 86 - 91.
- 23 - Weber, M.B. (1978). **The effect of choice format on internal consistency**. Paper presented at the National Council on Measurement in Education Annual Meeting, Toronto, Canada.

DETERMINING THE OPTIMUM NUMBER OF OPTIONS TO A MULTIPLE - CHOICE TEST ITEM

Dr. Nu'man Almusawi

*College of Education
The University of Bahrain*

This study aims to determine the optimum number of options to a multiple - choice test item considering validity, reliability and item discrimination of the test for examinees of different achievement levels.

A 4-option achievement test, containing 40 items covering a unit in “Educational measurement and evaluation” course, was constructed by the author and was administered to a random sample of 144 students, enrolled at the University of Bahrain. One option was deleted from each test item to create a 3-option test which was administered to the same sample after two weeks. Students were later divided into three groups based on student’s final course score which was subsequently used as the criteriion in calculating predictive validity coefficient for each test.

Significant differences were found between predictive validity coefficients of the two test formats for the average group in favor of the 3-option test. No significant differences were found between reliability coefficients and between item discrimination means for each group and for the combined sample. This result suggests that the optimum number of options to a multiple - choice test item is three.