

دليلك إلى صياغة اختبار الأسئلة متعدد الخيارات

Your Guide to Write Multiple Choice Test Questions

دمشق 2016

المساهمون

الدكتورة عزيزة رحمة	الإعداد
الدكتورة ميسون دشاش	المراجعة اللغوية والعلمية
الدكتور عدنان علي نظام	المراجعة اللغوية
عبد المنعم حسين	الإخراج الطباعي
المهندسة زينب مسلماني	متابعة التصميم والإخراج الطباعي



المحتوى

5 تقديم وزير التعليم العالي
7 مقدمة
8 خصائص الاختبار الجيد
30 مواصفات اختبار الأسئلة متعددة الخيارات
57 قواعد اختبار الأسئلة متعددة الخيارات
65 التحليل الإحصائي للاختبارات
79 تطبيقات
85 الملاحق
90 المراجع

Test

1. A B C D
2. A B C D
3. A B C D
- A B C D
- 
- A close-up photograph of a multiple-choice test paper. The paper is white with a black border. It features three numbered questions, each with four options labeled A, B, C, and D. Each option is followed by a small square checkbox. In the first question, the checkbox for 'B' is checked with a red mark. In the second question, the checkboxes for 'A' and 'C' are checked with red marks. A person's finger is pointing at the 'C' checkbox in the second question. The third question and the start of a fourth question are visible but their checkboxes are not checked.



تقديم

نظراً إلى أهمية معرفة الأساسيات المتعلقة ببناء الاختبارات، ولما تمثله الأسئلة متعددة الخيارات من أهمية في تقويم العمليات التعليمية في الجامعات السورية اعتماداً على قياس عدة مستويات معرفية، فإن هذا الدليل يوضح القواعد العامة لكتابة البنود الاختبارية ولتطوير بنود تقيس مستويات عقلية عليا إلى جانب ميزاتها وعيوبها وطرائق تحليلها، وتصميم جدول المواصفات لضمان الشمولية والقياس المطلوب. وقد أصبحت هذه المعرفة مطلباً أساسياً في التعليم العالي اليوم من أجل أداء أفضل وتحسين جودة المنتج التعليمي. أشكر كل من شارك في إعداد هذا الدليل وأثنى جهود كل من يساهم في رفع سوية العملية التعليمية في الجمهورية العربية السورية.

وزير التعليم العالي
الدكتور محمد عامر المارديني





مقدمة

تعد الأسئلة متعددة الخيارات من أكثر أنواع الأسئلة شيوعاً في عمليات التقويم ومع ذلك يجد الكثير من مصمميها صعوبة في بناء بنود اختبارية تتجاوز في قياسها عمليات الحفظ والاسترجاع. ونعرض في هذا الدليل كيفية بناء بنود اختبارية للأسئلة متعددة الخيارات تقيس عدة مستويات معرفية. وهذا يتطلب الإلمام ببعض الأساسيات المتعلقة ببناء الاختبار يأتي في مقدمتها معرفة خصائص الاختبار الجيد وتبني أحد التصنيفات المعرفية في بناء وصوغ الأسئلة. وتصميم جدول مواصفات وآلية حدد عدد الأسئلة الكلي للاختبار ومدى توزيعها على المستويات المعرفية لضمان شمولها وقياسها لعدة مستويات معرفية. وأيضاً يشرح هذا الدليل القواعد العامة التي ينصح بها لكتابة البنود الاختبارية. ولتطوير بنود تقيس مستويات عقلية عليا. وكذلك ميزاتها وعيوبها وطرائق خليلها.

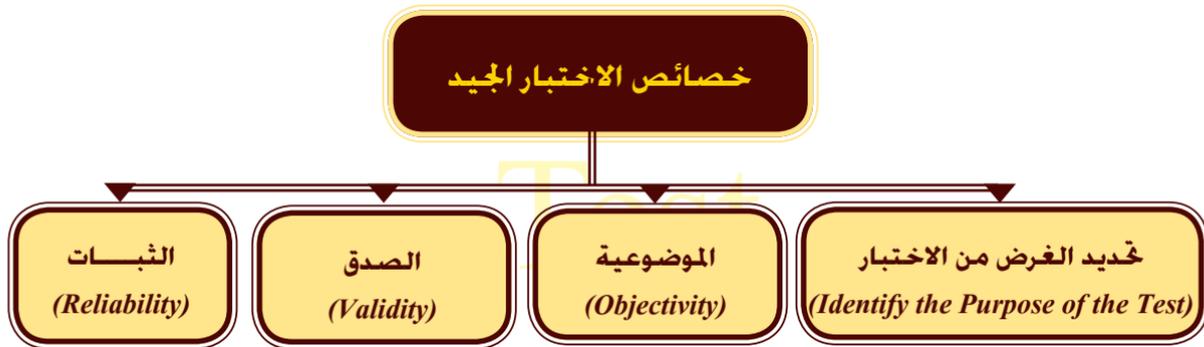
مدير عام مركز القياس والتقويم

الدكتورة ميسون دشاش



خصائص الاختبار الجيد (The Characteristics of Good Test)

تتمثل خصائص الاختبار الجيد في أربع خصائص هي:



◉ تحديد الغرض من الاختبار (Identify the Purpose of the Test)

يجب أن يكون لكل اختبار غرض واضح ومحدد تماماً. ومن المتوقع من مطور الاختبار أن يوضح الصفة أو الخاصية التي يقيسها الاختبار. ويحدد الفئة المستهدفة (أي من سيأخذ الاختبار). وكيف يستفاد من نتائج



القياس؛ فالاختبار التشخيصي يحتاج أشكال وطرائق أسئلة تختلف في المحتوى والصعوبة عما يرد في اختبار قياس الإنجاز.

◉ الموضوعية (Objectivity)

إخراج رأي المصحح أو حكمه الشخصي من عملية التصحيح أو عدم توقف درجة المفحوص على من يصحح إجابته أو اختلاف درجته باختلاف المصححين. وأن يكون جواب السؤال محددًا بحيث لا يختلف عليه جميع المصححين. فإذا أردنا أن تتصف الاختبارات بدرجة عالية من الموضوعية يجب أن:

Test

أ - فيما يتعلق بالأسئلة

- ◉ شاملة لجميع المقرر.
- ◉ يجب أن تمتاز الأسئلة بالوضوح من الغموض.
- ◉ اللغة واضحة.

ب - فيما يتعلق بالتصحيح

- ◉ إذا كانت الأسئلة موضوعية فيجب وضع مفتاح للإجابة.
- ◉ يجب توضيح الإجابة أمام المتعلمين.





◉ الصدق (Validity)

يرتبط مفهوم صدق الاختبار بصحة صلاحيته للاستعمال. فالاختبار الصادق يقيس الأهداف التي وضع من أجلها. ويمكن أن نعني به ذلك الاختبار المصمم لقياس سلوك معين أو سلوك ما. وترتبط بنوده كلها بالسلوك المراد قياسه. فاختبار القياس والتقويم يكون صادقاً إذا قاس مقدرة المتعلم على استعمال المهارات القياسية. أما إذا احتوى على بنود أو أسئلة تمثل قواعد اللغة العربية، فإن هذا الاختبار لا يكون صادقاً لأنه خالف ذلك. ويمكن تعريفه أيضاً بأنه «مدى شمولية الاختبار للغرض الذي يجب أن يحققه أو مدى قيامه بالوظيفة المفترض قيامه بها عندما يطبق على فئة وضع من أجلها».

خصائص الصدق

◉ **تتوقف خصائص الصدق على عاملين:** هما الغرض من الاختبار، أي وظيفة الاختبار التي يقيسها، والفئة المستهدفة التي يطبق عليها.

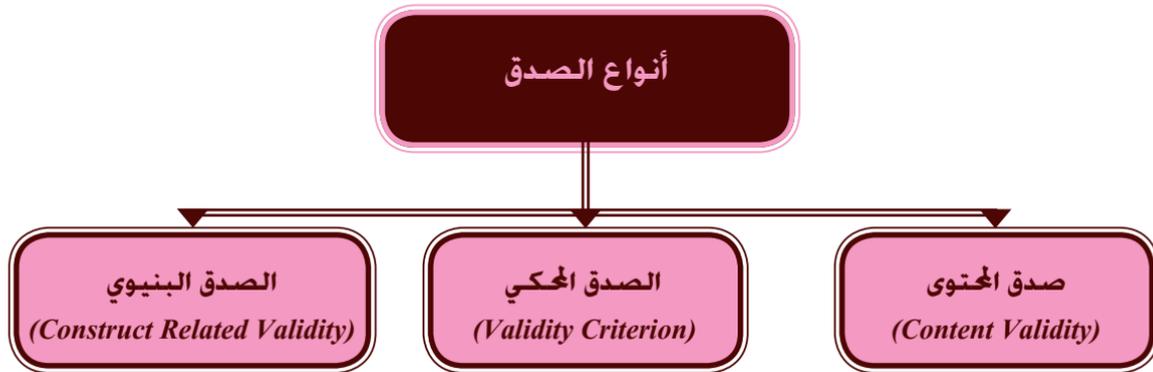
◉ **الصدق نسبي:** يمكن أن يكون الاختبار صادقاً لقياس سلوك طلاب / جماعة في صف ما في بيئة معينة. وليس صادقاً في قياس سلوك المتعلمين في الصف نفسه في بيئة أخرى؛ فالصدق يعتمد على طبيعة الجماعة التي تطبق عليها الاختبار. فالاختبار الذي يكون صادقاً في قياس سلوك ذكاء الراشدين قد لا يكون صادقاً بالدرجة نفسها في قياس ذكاء الأطفال.



- ◉ **قد يكون الصدق نوعياً:** إن الاختبار الصادق الذي يقيس معارف المتعلمين في الفيزياء لا يكون صادقاً إذا استعمل في قياس مهارات المتعلمين في التربية الفنية.
- ◉ **الصدق يتضمن الثبات:** الاختبار الصادق يقيس سلوكاً ما بدقة، فإن معلم الجغرافية إذا وضع اختباراً هادفاً لقياس مهارات استعمال الخريطة، فإن هذا الاختبار لا يكون هادفاً إذا أعطى نتائج مختلفة في حالة استعماله ثانية في قياس السلوك نفسه، ويكون صادقاً ويمتاز بالثبات إذا أعطى النتائج نفسها.
- ◉ **الصدق ليس درجة مطلقة:** فالاختبار الصادق ليس صادقاً بدرجة كاملة فقد ترتفع درجة الصدق في اختبار قياس الذكاء إلى (0.9)، ويكون اختباراً صادقاً بدرجة كاملة، وتقل درجة الصدق لاختبار الذاكرة إلى (0.7)، ويطلق عليه اختباراً صادقاً، فالمهم هو درجة صدق الاختبار، ويجب ألا تقل درجة الصدق عن (0.64) – (0.66).



أنواع الصدق



◉ صدق المحتوى (Content Validity)

- يدل على مدى تمثيل أداة القياس للنطاق السلوكي للسمة المراد الاستدلال عليها.
- يعتمد على اتفاق المحكمين. وإذا اتفق الجميع على أن بنود الاختبار تمثل مجال المحتوى تمثيلاً مناسباً يمكن القول إنه يتصف بصدق المحتوى.



- دليل صدق المحتوى هو دليل منطقي وليس إحصائياً.
- يناسب صدق المحتوى الاختبارات التحصيلية (جدول المواصفات).
- صدق المحتوى ليس صدق ظاهرياً.

○ الصدق المحكي (Validity Criterion)

يبين الدليل المتعلق بالمحك المدى الذي ترتبط به درجات أداة القياس مع متغير خارجي مستقل (بالمحك). يعتقد أنه يقيس على نحو مباشر السلوك أو الخاصية المعينة. **مثل:** اختبار الاستعداد والمعدل التراكمي.

خواص المحك

- 1 - أن يكون وثيق الصلة بالموضوع (المحك يمثل الأداء الناجح للسلوك المعنى).
- 2 - أن يتصف المحك بالمصادقية والثبات (ينبغي للمحك أن يكون مقياساً ثابتاً للخاصة على مدار الزمن أو من موقف لآخر). فإذا لم يكن المحك مستقراً فلا يمكن أن نتوقع ارتباطه على نحو ثابت بأي أداة تنبؤ.
- 3 - أن يكون متحرراً من التحيز (أي لا ينبغي أن يتأثر وضع درجات مقياس المحك بأي عوامل غير الأداء الفعلي الخاص بالمحك).



وللصدق المحكي نوعان هما:

الصدق التلازمي (محكات أداء راهنة).

الصدق التنبؤي (محكات أداء مستقبلية).

الصدق البنوي (Construct Related Validity)

يركز على وجود مؤشر أو دليل على أن درجات أداة القياس تقيس المفهوم النظري أو السمة المراد قياسها. والمفهوم النظري هو تصور فكري متعلق بالمفهوم وأبعاده أو هو أفضل ما يتوصل إليه العلماء في وقت ما لتوضيح سمة معينة. **مثال:** النضج الاجتماعي لا يمكن قياسه مباشرة، إلا أن العديد من أنواع السلوك التي نعتقد أنها تقيس المفهوم يمكن وصفها أو قياسها. **مثل:** مساعدة الذات، التوجيه الذاتي، المهنة، الاتصال، العلاقات.

طرائق التأكد من الصدق البنوي

دراسة الفروق

دراسة الفروق بين مجموعات مختلفة على مقياس معين ثم المقارنة بين المجموعة العليا والمجموعة الدنيا. وعند وجود فروق دالة إحصائية بين المجموعتين يمكن عدّ هذا أحد أدلة الصدق البنوي للاختبار (صدق المقارنة الطرفية).



معاملات الارتباط

دراسة السمة من خلال استعمال أكثر من أداة طريقة. فالارتباطات بين السمات دليل على الصدق التقاربي.
مثل: سمة الفهم اللفظي تُقاس من خلال: تقديرات الفاحصين. اختبارات ورقية. اختبارات فردية.
فعند دراسة مجموعة من السمات (الفهم اللفظي. الاستدلال العددي. المعلومات العامة) بطريقة واحدة. وهى تقديرات الفاحصين. ووجد ارتباط بين هذه السمات فهذا دليل على الصدق التمايزي.

التحليل العاملي

هو طريقة إحصائية لدراسة معاملات الارتباط المتبادلة بين مجموعة من درجات الاختبارات بهدف تحديد عدد العوامل (المفاهيم البنائية) المطلوبة لتفسير هذه معاملات الارتباط المتبادلة؛ فالمرء يبدأ بعدد كبير من المقاييس المختلفة ويفحص معاملات الارتباط بينها ويجد العوامل التي يرتبط بعضها مع بعض. وهذا يعمل على تقليص البنود إلى عدد أصغر من العوامل التي يجرى قياسها. وتعد أساساً لها « صدق الاتساق الداخلي».

التجريب

يعتمد على التدخل التجريبي لإحداث تغيرات في درجات الاختبار كوسيلة لتعرّف مدى تأثر الأفراد بالمعالجات أو متغيرات معينة؛ مما يساعد في تأكيد بعض التفسيرات المتعلقة بنتائج الاختبار أو رفضها.

التحليل المنطقي

يعتمد على الفحص الدقيق لمحتوى الاختبار والأداء الذي يتطلبه وإحداث تكامل بين نتائج الفحص والنظرية التي يستند إليها الاختبار (التحكيم).

العوامل المؤثرة في صدق الاختبار

عوامل متعلقة بالمتعلم

- اضطراب المتعلم وخوفه عند أداء الاختبار للحصول على نتيجة لا تمثل مقدرته العقلية.
- الغش يؤثر سلباً في الصدق إذ يعطي صورة مغايرة عن مقدرات المتعلمين.
- عدم تحديد المتعلم للإجابة بصورتها الصحيحة.

عوامل متعلقة بالاختبار

- اللغة غير الصحيحة وعدم ارتباط بنود الاختبار يؤثر سلباً في فهم المتعلم.
- غموض الأسئلة يؤثر في استيعاب المتعلمين.
- صعوبة بنود الاختبار يؤثر سلباً في الصدق.

عوامل تتعلق بإدارة الاختبار

- عوامل بيئية، مثل: درجة الحرارة، والبرودة الشديدة، ضيق المكان.



- المراقبة الشديدة تؤثر في أداء المتعلمين مما يؤثر في صدق الاختبار.
- الضوضاء في قاعة الاختبار تؤثر في نفسية المتعلمين: مما يؤثر في الصدق.

طرائق حساب صدق الاختبار

أ - طرائق غير إحصائية

هي طرائق الحكم على صدق الاختبار من خلال المحتوى الدراسي حيث نتأكد من أن بنود الاختبار تمثل جميع المقرر من ناحية والأهداف من ناحية أخرى. فعلى سبيل المثال لو أردنا أن نتأكد من صدق اختبار في القياس والتقويم لطلاب كلية التربية، فإننا نتأكد من أن جميع الأسئلة تمثل المحتوى والأهداف التي نريد أن نحققها.

ب - الطرائق الإحصائية

هي الطرائق الحكم على صدق الاختبار من خلال تحليل نتائج المتعلمين فيه باستعمال مقاييس عديدة منها مقاييس (النزعة المركزية) إذ تهتم هذه المقاييس بمعرفة تكتل درجات التحصيل حول المنحنى، وتشمل القيم الوسطية الحسابية التي تستخرج من حاصل قسمة مجموع درجات المتعلمين على عددها. إضافة إلى معاملات الارتباط التي تحدد شدة العلاقة واتجاهها بين الاختبار والحكم، كذلك اختبارات الدلالة الإحصائية **مثل: اختبار (t-test).**





◦ الثبات (Reliability)

يعني الثبات أن الفرد يحافظ على الموقع نفسه تقريباً لمجموعته، عند تكرار قياسه. ويبقى على حاله تقريباً بالقدر الذي يتمثل فيه بقيمة صغيرة للخطأ المعياري في القياس. أو بمعامل ثبات مرتفع.

يجب أن يكون الاختبار التحصيلي على درجة عالية من الدقة والإتقان والاتساق والاطراد فيما يزودنا به من بيانات على مستوى المتعلم إذ إن مفهوم الثبات من المفاهيم الأساسية التي يجب أن تتوفر في الاختبار لكي يكون صالحاً للاستعمال. ففي كل اختبار يوجد قدر من أخطاء القياس. وقد تكون الأخطاء قليلة أو كثيرة مما يؤثر في نتائج القياس. وتسمى هذه الأخطاء بأخطاء الصدفة. إذ إنه لا يوجد اختبار سواء كان تحصيلياً أو عقلياً أو نفسياً يحصل على درجة ثبات كاملة لأن من غير الممكن التخلص من الأخطاء والشوائب في الاختبار؛ لأننا عندما نريد استخراج معامل الثبات نحسبه بناءً على ارتباط درجات المرة الثانية بالصورة نفسها أو بصورة متكافئة من الاختبار. وهنا لا بد أن يكون معرضاً للخطأ. وهذا الخطأ هو أن هؤلاء الأفراد أنفسهم يتغيرون. وعلينا أن نعرف مدى تغيرهم حتى نعرف سبب تغير درجاتهم في المرات المختلفة. فقد لا يكون التغير في الاختبار بل فيمن يطبق عليهم إذ إن الدرجة التي يحصل عليها الفرد في أي اختبار لا تعبر عن الأداء الحقيقي فقط؛ إنما تمثل الأداء الحقيقي للفرد مع عوامل أخرى. بمعنى أن درجة الفرد على الاختبار



تعبّر عن التباين الحقيقي للفرد وتباين الخطأ. وبذلك فإن درجة الاختبار تتضمن قدرًا من الخطأ. قد يؤدي إلى زيادة الدرجة الحقيقية أو إلى نقصان في الدرجة عما يستحقه الفرد المقاس؛ مما يؤدي إلى تقدير أداء الفرد أقل أو أكثر من الواقع الذي يستحقه. ويتراوح معامل الثبات بين درجتين (الصفري و 1) ويعد الصفري أدنى معامل ثبات. أما الدرجة (1) فتمثل أعلى معامل ثبات. ومن الصعب الوصول إلى معامل ثبات يساوي (1) بسبب الأخطاء العديدة التي تتعلق بنتائج الاختبار التي لا تخضع للضبط العلمي والتحكم الدقيق. وكذلك الحالة النفسية أو الجسمانية للفرد والحالة الفيزيائية وغيرها مما يؤثر مباشرة في نتائج الثبات.

طرائق قياس الثبات

1 - إعادة الاختبار نفسه (Test-Retest Method) والإجراءات التي يمكن للفاحص أن يستعملها في استخراج معامل الثبات تتمثل في أن يقوم بإعطاء الاختبار الذي يريد معامل ثباته، ثم يصححه ويسجّل نتائجه. وبعد فاصل زمني يتراوح بين بضع ساعات وبضعة أيام يعيد إعطاء الاختبار نفسه للمتعلمين أنفسهم. وضمن ظروف متشابهة ثم يصححه بحسب القواعد نفسها ويسجّل نتائجه. وأخيراً يحسب معامل الارتباط بين درجات المتعلمين في المرة الأولى. ودرجاتهم في المرة الثانية. ونسمى معامل الارتباط الذي نحصل عليه بهذه الطريقة عامل الثبات.



2 - **طريقة الصور المتكافئة (Equivalent Forms Method)** تعد الصور المتكافئة للاختبار نماذج بنيت وفقاً لمواصفات واحدة، ولكنها تتألف من عينات مستقلة، من منطقة سلوك محددة، ولذلك فإن اختبارين متكافئين للقراءة، يجب أن يتضمنا أسئلة لها الصعوبة نفسها، ويسأل فيها النوع نفسه من الأسئلة.

3 - **طريقة الثبات النصفي (Split- Half Method)** قد يكون من الصعب على الفاحص أن يطبق اختبارين متكافئين على المتعلمين، أو قد يتعذر عليه فحص المتعلمين مرتين في الاختبار نفسه؛ لذلك نلجأ إلى تقسيم الاختبار إلى نصفين، يفترض أنهما متكافئين. ومن الممكن تجميع نصفي الاختبار على أساس تفحص دقيق للمحتوى والصعوبة لكل فقرة، وهذا جهد منظم لموازنة المحتوى ومستوى الصعوبة في النصفين، ولكن الطريقة الأبسط التي يكثر استعمالها، هي وضع الأسئلة ذات الأرقام الفردية في النصف الأول والزوجية في النصف الثاني، وحساب الارتباطات بين النصفين، هو معامل الثبات في هذا الاختبار، ويلاحظ أن التجزئة مرتبطة فقط بتصحيح الاختبار، أما تطبيقه فنجره مرة واحدة.

4 - **طريقة اتفاق المقومين (Rates Agreement Method)** في هذه الطريقة يحسب معامل ثبات الاختبار عن طريق حساب معامل الارتباط بين تقويم المقومين للمجموعة نفسها من المتعلمين، وقد يلجأ إلى مثل هذه الطريقة حين يصعب استعمال الطرائق الأخرى في حساب معامل الثبات.



5 - **طريقة الخطأ المعياري (Standard Error Measurement)** في هذه الطريقة يطبق الاختبار أكثر من مرة على العينة، وبحسب الخطأ المعياري لدرجات العينة، فكلما كان الخطأ المعياري كبيراً كان معامل الثبات منخفضاً والعكس صحيح.

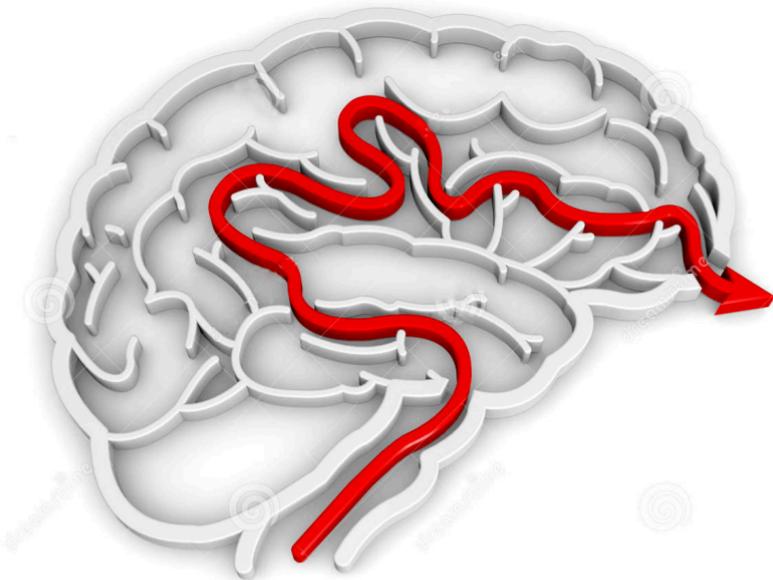
6 - **طريقة كرونباخ والمعادلات المشتقة منها:** تعتمد على مدى تجانس درجات المقياس المستعمل، وهي تكشف عن تشتت درجات التوزيع الممثل بدرجات المتعلمين على المقياس المطبق، وتعتمد على حساب الارتباطات بين الدرجات لمجموعة الثبات على جميع البنود الداخلة في الاختبار، وكأننا قسمنا الاختبار إلى عدد من الأجزاء يساوي عدد بنوده، وتستعمل هذه المعادلة مع أي نوع من أنواع الأسئلة الموضوعية أو المقالية.

7 - **معادلة كودر ريتشاردسون 20 (Kuder-Richardson 20)** تعد من أفضل الطرائق في إيجاد معامل الاتساق الداخلي للاختبار، إذ تعتمد على قياس تجانس بنود الاختبار، ولهذا فهي تنتمي إلى أساليب تحليل التباين (Variance Analysis)، كما أن هذه الطريقة تتعامل مع نسبة عدد المتعلمين الذين أصابوا في إجاباتهم عن كل سؤال والذين لم يجيبوا عن تلك الأسئلة، وتستعمل مع أي نوع من أنواع الأسئلة الموضوعية.

8 - **معادلة كودر ريتشاردسون 21 (Kuder-Richardson 21)** تستعمل هذه المعادلة في حالة إيجاد معامل ثبات أي اختبار خصلي أو غيره، إذ تعتمد على درجة تباين استجابات المتعلمين على بنود الاختبار ككل.



وكذلك على عدد بنوده. فكلما كان عدد البنود قليلاً أنخفض جَانس الاختبار. وإذا كان عددها كبيراً ارتفع جَانس الاختبار. ولهذا يفضل استعمال معادلة كودر ريتشاردسون في حالة الاختبارات التي يزيد بنودها على (25) فقرة.





العوامل المؤثرة في الثبات

- 1 - **طول الاختبار:** إن عدد أسئلة الاختبار عامل مؤثر في درجة ثبات الاختبار. فكلما زاد عدد البنود أدى ذلك إلى ارتفاع معامل الثبات.
- 2 - **زمن الاختبار:** يتأثر ثبات الاختبارات الموقوتة بالزمن المحدد لها. وبذلك يزداد الثبات تبعاً لزيادة الزمن حتى يصل إلى الحد المناسب للاختبار فيصل إلى نهايته العظمى ثم يقل بعد ذلك كلما زاد الزمن عن ذلك الحد.
- 3 - **صوغ الأسئلة:** إن الأسئلة الغامضة والطويلة تقلل من الثبات. والأسئلة الواضحة والموضوعية والقصيرة تزيد من الثبات. وهذا يتطلب من الباحث الدقة في اختيار الألفاظ والعبارات ونوعها عند صوغ الأسئلة حتى يجعلها قادرة على الوصول إلى الثبات الحقيقي.
- 4 - **صعوبة البنود ودقتها:** إن وجود بنود صعبة في الاختبار لا يستطيع أغلب الأفراد أو جميعهم من الإجابة عنها تؤثر في الاختبار من حيث ثباتها إلا أن حذفها لا يؤثر في درجة المتعلم. كما أن وجود بنود سهلة يستطيع جميع أفراد العينة الإجابة عنها تؤثر في معامل ثبات الاختبار إلا أن حذفها لا يؤثر في الاختبار؛ لذلك ينبغي حذف الاسئلة الصعبة أو السهلة لرفع ثبات الاختبار.
- 5 - **حالة الفرد:** يتأثر الثبات بحالة الفرد النفسية والصحية ومدى تدريبه على الموقف الاختباري؛ فالمرض والتعب والتوتر الانفعالي قد يؤدي إلى نقصان الثبات.



6 - ظروف إجراء الاختبار: أي تغيير في الظروف الخاصة بإجراء القياس من اختبار لآخر يؤثر في نتائج المقياس ويحد من عوامل الخطأ الذي يؤثر في الثبات.

7 - موضوعية الاختبار: تعد موضوعية التصحيح من بين العوامل المؤثرة في ثبات الاختبار ولاسيما في الاختبارات التي تعتمد على تقدير المصحح كاختبارات المقال التحصيلية واختبارات الإبداع إذ إن تباين التصحيح يؤثر في زيادة تباين الخطأ وبالتالي نقصان معامل الثبات.

8 - التخمين: يلجأ بعض من المتعلمين في حالة عدم تأكدهم من الإجابة الصحيحة إلى التخمين مما يؤدي إلى خفض ثبات الاختبار إذ تؤدي زيادة تأثير التخمين إلى نقص الثبات.

9 - تباين العينة وجانستها: يرتبط الثبات بتباين الاختبار إذ ينخفض الثبات عندما يقل التباين ويزداد تبعاً لزيادة التباين. فلو افترضنا أن درجات ثلاث متعلمين في التطبيق الأول كانت على النحو الآتي: (10-11-13) فمن المحتمل في التطبيق الثاني أن يحصل المتعلم صاحب الدرجة (10) على الدرجة (13) وصاحب الدرجة (13) على الدرجة (11) بسبب جانس الدرجات وتقاربها، في حين لو كان توزيع الدرجات متبايناً كالاتي: (15-25-42) فمن غير المحتمل أن يحصل المتعلم الذي كانت درجته (15) في التطبيق الأول على الدرجة (45) في التطبيق الثاني.



علماً أن سهولة البنود وصعوبتها تؤثر في تباين الدرجات، إذ إن الأسئلة السهلة والصعبة تؤدي إلى التجانس وبالتالي خفض الثبات والأسئلة المتدرجة في الصعوبة تؤدي إلى رفع الثبات.

ما الحد الأدنى لثبات مقبول لأداة معينة؟

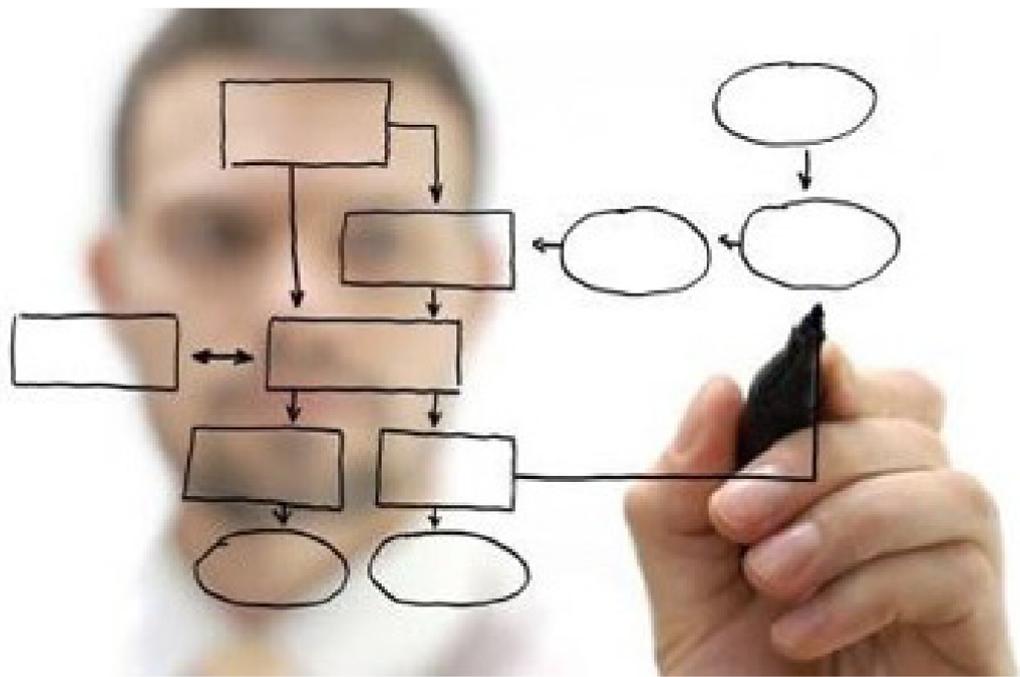
◆ يعتمد هذا على السمة المقاسة وعلى استعمال النتائج المترتبة ولاسيما الأشخاص الذين يطبق عليهم الاختبار.

◆ يتوقف المستوى المقبول على كمية خطأ القياس التي يمكن لمستعمل الاختبار أن يتحملة في التطبيق المقترح.



العلاقة بين الصدق والثبات

إن الصدق يتضمن الثبات. وهو مظهر من مظاهره. كما أنه أعم وأشمل من الثبات إذ إن ثبات الاختبار ودقته لا تدل على صدقه بل من الممكن أن يكون الاختبار ثابتاً ودقيقاً. وإن درجة الفرد عليه لا تتغير كثيراً من إجراء لآخر. ولكنه ليس صادقاً بمعنى أن يقيس دقة وثبات عامل آخر غير العامل الذي صمم الاختبار لقياسه. ومن جهة أخرى فإن كل اختبار صادق هو ثابت بالضرورة. ولكن لا يمكن القول إن كل اختبار ثابت هو صادق بالضرورة إذ إن الاختبار الصادق يقيس فعلاً ما أعد لقياسه فإن درجته معبرة عن الأداء الحقيقي أو المقدرات الفعلية للفرد. وتعبر عن الوظيفة المقيسة بكل دقة؛ وبالتالي تكون ثابتة في الوقت نفسه. أي إن الاختبار الثابت ليس بالضروري أن يكون صادقاً لأنه قد يقيس وظيفة أخرى عدا الوظيفة المخصص لقياسها أو يقيس وظيفة أخرى إلى جانب الوظيفة التي خصص لها.



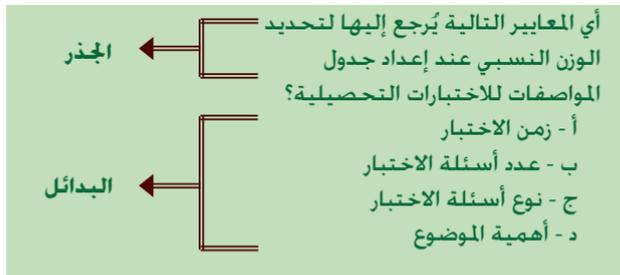


مواصفات اختبار الأسئلة متعددة الخيارات

⦿ تعريف اختبار الأسئلة متعددة الخيارات ومواصفاته

يتألف البند المعباري لاختبار الأسئلة متعددة الخيارات من جزأين أساسيين: المشكلة (الجذع *Stem*) وقائمة من الحلول المقترحة (البدائل *Alternatives*). يمكن أن يصاغ الجذع في هيئة سؤال أو في عبارة تقريرية ناقصة. وقائمة من البدائل تشتمل على إجابة واحدة صحيحة أو أفضل بديل (إجابة). إضافة إلى عدد من البدائل الخاطئة (المشتتات *Distractors*).

الهدف من المشتتات عرضها كحلول مقبولة للمشكلة بفعل أولئك الذين لم يحققوا الهدف المقاس بالبند. وبالمقابل يجب أن تبدو المشتتات إجابات غير ممكنة لمن حقق الهدف. وتظهر لهم فقط الإجابة الصحيحة كحلٍّ مقبول للمشكلة. **مثال:**





يعد هذا النوع من الاختيارات أفضل أنواع اختبارات الموضوعية؛ فهي تقيس أهدافاً عقلية عليا. ويصعب على الاختيارات الموضوعية الأخرى قياسها. كما أنها تلائم جميع الأهداف التي يمكن تقويمها في اختبارات المقال باستثناء التي تتطلب قدرات التعبير والتفسير. إلا أن بناءها يحتاج إلى خبرة لدى الفرد الذي يقوم بوضعها إلى جانب تمكنه من المادة العلمية المعدة. ويستعمل هذا النوع في قياس الأهداف الأكثر تعقيداً (المستويات المعرفة العليا) والمختلفة.

© خصائص اختبار الأسئلة متعددة الخيارات

أ - الثبات

يعد اختبار الأسئلة متعددة الخيارات من أفضل صيغ الاختبارات الموضوعية فيما يتعلق بالثبات. ويعد أقل قابلية للتخمين مقارنة بالصيغ الأخرى كصيغة أسئلة الخطأ والصواب. وبذلك فهو قادر على إنتاج درجات ثبات عالية إضافة إلى أن تقدير الدرجة أكثر تحديداً من صيغ أسئلة الإجابات القصيرة. وهذه الميزة تجعل بنود اختبار الأسئلة متعددة الخيارات لا تتأثر باختلاف تقدير المصححين. كما هو الحال في الاختبارات المقالية. كما أنها غير قابلة للتأثر بعوامل الكتابة كالإلتفاف على الإجابة الصحيحة أو الإيهام بالفهم. مع ذلك فإن اختبار الأسئلة متعددة الخيارات تتأثر إلى حد ما بعامل التخمين الذي يقل تأثيره كلما زاد عدد الأسئلة. كما هو موضح بالجدول الآتي:



عدد الأسئلة التي يحتوي كل منها على (4) بدائل	احتمال الحصول على (70%) أو أكثر من الدرجة الكلية باستعمال التخمين (تقريباً).
2	1 من 16
5	1 من 64
10	1 من 285
15	1 من 8670
20	1 من 33885
25	1 من 942651

ب - الفاعلية

الخاصية الثانية لاختبار الأسئلة متعددة الخيارات هي الفاعلية ويقصد بها: أن بنود هذا الاختبار تتميز بسهولة وسرعة تصحيحها ولاسيما باستعمال المصحح الآلي؛ مما يساعد على إصدار نتائج الاختبار في وقت قصير.



© مزايا اختبار الأسئلة متعددة الخيارات

- 1 - يمكن من قياس جميع الأهداف التربوية من فهم، وتحليل، وتركيب وتفسير، وتقويم.
- 2 - يقلل من التخمين والصدفة.
- 3 - يمكن الثقة بنتائجها؛ لأنها تقرر درجة المتعلم بموضوعية كاملة.
- 4 - تدفع المتعلم إلى التفكير بروية قبل تحديد الإجابة.
- 5 - أكثر صدقاً وثباتاً من بقية الاختبارات الموضوعية.
- 6 - تساعد في تشخيص أخطاء المتعلمين أو سوء الفهم لديهم من خلال إجاباتهم الخاطئة.
- 7 - تلزم المتعلم بمراجعة أكبر قدر من المقرر المطلوب.
- 8 - يمكن تحليل نتائجها بسهولة.

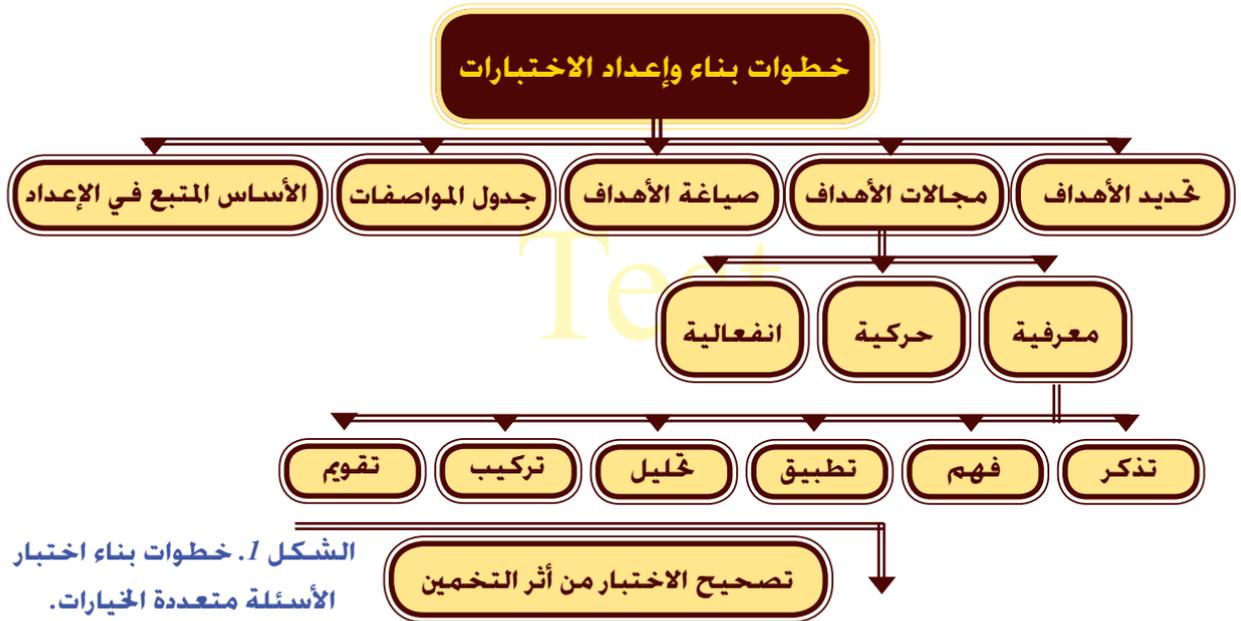
© عيوب إعداد اختبار الأسئلة متعددة الخيارات

- ◆ صعوبة الإعداد ويصعب الحصول على البدائل الجذابة.
- ◆ معرضة للكلمات غير المناسبة مثل أسئلة الإجابة القصيرة.





© خطوات بناء وإعداد الاختبارات



الشكل 1. خطوات بناء اختبار
الأسئلة متعددة الخيارات.



◉ التخطيط لبناء بنود اختبار الأسئلة متعددة الخيارات

يعد التخطيط لبناء الاختبار الخطوة الأولى التي تبنى عليها مواصفات الاختبار التي تتضمن طوله ونوع المستويات المعرفية التي يقيسها، وتتطلب هذه الخطوة الإلمام بعاملين. الأول التصنيف المعرفي المزمع استعماله، والثاني جدول الأوزان النسبية للمحتوى المطلوب تمثيله:

◈ العامل الأول التصنيفات المعرفية

يسترشد معدو الأسئلة عموماً بتصنيف معرفي معين لضمان شمول الأسئلة وتمثيلها لمستويات معرفية متعددة. ويعد تصنيف بلوم الموضح أدناه من أكثر التصنيفات المعرفية استعمالاً.



تصنيف بلوم (Bloom)

المستوى	التعريف
المعرفة	تذكر حقائق أو معلومات أو أشياء سبق تعلمها. وهي أدنى مستويات المجال المعرفي
الفهم	المقدرة على إدراك أو استيعاب معنى المقرر أو النص الذي يدرسه المتعلم
التطبيق	استعمال ما تم تعلمه من قوانين أو قواعد في مواقف جديدة
التحليل	جزئة أو تحليل المقرر إلى عناصرها أو مكوناتها
التركيب	جمع عناصر أو أجزاء لتكوين كل متكامل أو نمط غير موجود أصلاً
التقويم	إصدار أحكام حول قيمة الأفكار أو الأعمال وفق معايير أو محكات



تصنيف بلوم المنقح

طور أندرسون بالتعاون مع كراثول العمل السابق لبلوم (Anderson and Krathwohl, 2000) ضمن فريق ضم مجموعة من الخبراء في مجال علم النفس المعرفي والمناهج وطرائق التدريس والاختبارات التربوية والقياس والتقويم وخرجوا بالتنقيح الموضح بالشكل التالي:

الشكل 2. تصنيف بلوم المنقح.

يتعرف. يكتب قائمة.
يصف. يسترجع. يسمي. يوجد

يشرح الأفكار أو المفاهيم.
يفسر. يلخص. يعيد الصياغة. يصنف

يستعمل المعلومات في مواقف
مشابهة. يطبق. يستعمل. ينفذ.

يجزء المعلومات إلى أجزاء صغيرة لتحري الفهم
واستكشاف العلاقات. يقارن. ينظم. يحلل مكون إلى
أجزائه وفق قوانين معينة. يستجوب. يستنتج

يبرر قرار أو مجموعة من الأفعال. يفحص.
يصوغ الفروض. ينتقد. يجرب. يحكم

يولد معلومات جديدة. ينتج.
يصمم. يبني. يخطط. يخترع





وقد تركزت أهم التعديلات كما يتضح من الشكل الموضح أدناه في المستويات العليا وإبدال المستويات إلى صيغ فعلية مباشرة.

الشكل 3. تصنيف بلوم.





تطبيقات تصنيف بلوم في بناء الأسئلة

يستعمل تصنيف بلوم على نحو كبير للتخطيط للتدريس وبناء الأهداف التعليمية. كما يستعمل لبناء الأسئلة. وقد أدى استعماله في بعض الاختبارات الدولية مثل التيمز (TIMSS) إلى تطوير تطبيقاته. وظهر مصطلح عمليات التفكير العليا التي تمثل المستويات الثلاث العليا على التصنيف. وتتضمن عمليات التفكير العليا توظيف المعارف والأفكار في مواقف جديدة. باستعمال مجموعة من العمليات كالتركيب والتعميم والشرح وفرض الفروض والتوصل بذلك إلى حل مشكلة ما أو بناء أفكار جديدة أو تفسير ظاهرة أو موقف. وقد استعمل معدو الاختبار في العام 2003 توزيعاً جديداً للتصنيف للاعتماد عليه في بناء الاختبار إلا أن معدي الأسئلة واجهوا صعوبة في التمييز بين بعض المجالات المعرفية لما بينها من تداخل ولاسيما معرفة الحقائق. والإجراءات. واستعمال المفاهيم. لذلك اقتصر استعمال التصنيف على اختبار 2003. وقد تضمن المستويات الآتية:

- معرفة الحقائق. والإجراءات. والمفاهيم (*Knowing Facts, Procedures, and Concepts*).
- الفهم وتطبيق المعارف (*Understanding and Applying Knowledge*).
- عمليات التفكير (*Reasoning*).

وقد أمكن تطوير هذا التصنيف لاحقاً ليكون أكثر سهولة لعدي الأسئلة. واعتمد في خطة الاختبار لعام 2007، ويتضمن:

- المعرفة
- التطبيق
- التفكير



■ مستوى المعرفة (The Level of Knowledge)



تشير المعرفة إلى المعلومات المبنية على الحقائق والمفاهيم والأدوات والإجراءات. وتعد المنطلق الأساسي للعمليات العقلية العليا. فكلما كان المتعلم يمتلك معلومات قادر على استدعائها، وكلما كان لديه مدى واسعاً من المفاهيم والمبادئ سبق استيعابها. كان لديه قاعدة صلبة للانخراط في مدى واسع من مواقف حل المشكلات وتطوير فهم لمواقف جديدة. فالمعلومات الدقيقة والغزيرة تمكّن المتعلم من الانخراط على نحو فاعل في أنشطة عقلية معقدة. وبالمقابل دون استعمال حيلة معرفية تمكّن من تذكر اللغة والحقائق الرئيسية ومجموعة الأرقام والرموز والعلاقات سوف يجد المتعلم صعوبة في التفكير البناء. تشتمل الحقائق على المعارف الفعلية التي تشكل حجر الأساس لأنماط التفكير والمفاهيم التي تمكّن المتعلمين من تشكيل علاقة بين عناصر المعرفة واستعمالها خارج النطاق الذي تعلمها فيه للحكم على مدى صحة التقريرات والطرائق والإجراءات التي تشكل جسراً بين المعارف الأساسية والمقدرة على استعمال معطيات تخصصية معينة لحل مشكلات اعتيادية تواجه أغلب الناس في الحياة اليومية. استعمال الإجراءات يستوجب في الأساس ذكر مجموعة من الأعمال وكيفية تنفيذها. ويحتاج المتعلمون لكي يكونوا فاعلين ودقيقين أن يستعملوا مجموعة متنوعة من الإجراءات والأدوات لحل فئة كاملة من المشكلات وليس مجرد مشكلات فردية. **مثال:**



أمثلة	الأفعال
يتذكر تعريفاً أو مصطلحاً. يحدد خصائص أو مكونات. يبين حقائق أو عمليات. يحدد حالة تقريرية حول الحقائق أو العلاقات أو الإجراءات والمفاهيم.	يتذكر (Remember)
يقدم أو يحدد تعريفاً لمصطلح ما. يميز ويستعمل بنود ورموز واختصارات ووحدات ومقاييس في محتوى ذي علاقة.	يعرف (Define)
يسترجع معلومات من رسم بياني. جداول أو ما شابهه. يقرأ أدوات مقياس بسيط.	يسترجع (Retrieve)
يستعمل أدوات قياس. يستعمل وحدات للقياس على نحو صحيح. ويقدر المقاسات.	يقيس (Measure)
يصنف الموضوعات، الأشكال، الأرقام، والتعبير بناء على خصائص عامة. يتخذ القرار الصحيح حول فئة العلاقات وترتيب الأرقام والموضوعات بناء على صفة مميزة.	يصنف / يرتب (Classify / Order)
يصف موضوعات معينة وعمليات تبين معرفة خصائص. بناء. وظائف وعلاقات.	يصف (Describe)
يدعم أو يوضح حالة تقريرية لحقائق أو مفاهيم مع ذكر أمثلة مناسبة. يحدد أو يقدم أمثلة محددة يبين معرفته بمفاهيم عامة. يحدد أو يقدم مثالاً معيناً لتوضيح معرفة أو مفهوم عام.	يبين مع الأمثلة (Illustrate with Examples)
يظهر معرفة باستعمال أدوات معينة. أجهزة. إجراءات. أدوات قياس.	يستعمل أدوات وإجراءات (Use Tools and Procedures)
يرتب. يحدد اسماً. يعرض. يسمي. يكرر. يحدد. يسترجع. يقر. يحسب.	مزيد من الأمثلة



■ مستوى التطبيق (Application Level)

Application

يمثل هذا المستوى التطبيق المباشر للفهم والمعارف في المواقف المعروضة. ولقياسه يطلب من المتعلم أن يقارن، ويطابق، ويصنف، بهدف تفسير معلومات علمية في ضوء مفاهيم أو مبادئ معينة سبق تعلمها، أو إيجاد حل أو تقديم شروحات، وكذلك يطلب منه إيضاح العلاقات أو المعادلات أو الصيغ في محتوى معروف لتعليم وتدرّيس تلك المفاهيم. كما يشمل هذا المستوى حل مشكلات اعتيادية على نحو وصفي أو رقمي، وكذلك المقدرة على استعمال رسم بياني أو نموذج لتوضيح البناء والعلاقات ويعكس المعرفة العلمية للطالب. **مثال:**





أمثلة	الأفعال
اختيار عملية أو طريقة أو استراتيجية فعالة أو مناسبة لحل مشكلة.	يختار (Select)
يعرض معلومات رياضية أو بيانات في رسم بياني. جدول أو خارطة.	يقدم (Represent)
ينتج نموذج مثل تكوين معادلة أو رسم تخطيطي لحل مشكلة اعتيادية. أو لبيان فهمه لمفهوم أو علاقات أو عمليات.	يقدم نموذج (Model)
يتبع وينفذ تعليمات معينة وفق مواصفات معطاة. يرسم شكلاً أو صورة.	يطبق
يحل مشكلة اعتيادية كذلك التي تواجهها الفئة التي ينتمي إليها. ويستعمل بيانات من جدول أو رسم بياني أو خارطة لحل مشكلة اعتيادية يتضمن تطبيقاً مباشراً أو توضيحاً لمفهوم.	يحل مشكلة معينة
يحدد أو يصف التشابه والاختلافات بين مجموعة من المكونات والعمليات. يميز ويصنف وفقاً لبيانات معطاة.	يقارن. يطابق. يصنف
يربط المعلومات التي يتضمنها مفهوم معين بالموقف الملاحظ أو الخصائص التي أمكن استنتاجها.	يربط
يفسر معلومات في أي من الصيغ المعطاة في نص أو جدول أو شكل أو غيره في ضوء المفاهيم والمبادئ ذات العلاقة.	يفسر البيانات
يقدم أو يحدد تفسيراً لظاهرة ملاحظة أو طبيعية يعكس مدى فهمه لمضامين مفهوم أو مبدأ أو نظرية أو قانون معين.	يشرح
يلخص. يفرق. يعرض. يسرد. يصوغ. يتصور. يناقش. يصف. يراجع. يترجم. يميز.	مزيد من الأمثلة



■ مستوى عمليات التفكير (Reasoning)

Reasoning



يشمل التفكير أنماطاً من العمليات العقلية تتضمن المقدرة على التفكير المنطقي والمنظم، ويتضمن التفكير الحدسي والاستقرائي المبني على أنماط أو نماذج وقوانين يمكن أن تستخدم للوصول إلى حل لمسألة غير اعتيادية (*Non Routine Problem*) وتعرف المشكلات غير الاعتيادية: بأنها تلك المسائل التي لا تعد مألوفاً للمتعلمين وتتطلب مستوى معرفياً أعلى من المستويات التي تتطلب إيجاد حلول للمشكلات الروتينية. حتى وإن كانت المعارف والمهارات المطلوبة لحلها سبق تعلمه. تتطلب المشكلات تفكيراً بعدة أساليب لجدة الموضوع أو تعقيدات الموقف أو ربما لأن الحل يتطلب عدة مراحل معتمداً في ذلك على معرفة وفهم موضوعات مختلفة مرتبطة بمجال معين. وبالرغم من أن العديد من السلوكيات المدرجة في قائمة التفكير تستند على عمليات التفكير في المشكلة سواء بسيطة أم معقدة، فإن كل منها يقدم في حد ذاته مخرجات تعليمية مهمة مع إمكان التأثير في عمليات التفكير لدى المتعلم عموماً. على سبيل المثال التفكير يتضمن مقدرة على الملاحظة وتكوين تخمينات، ويتضمن أيضاً استنتاجات منطقية مبني على افتراضات وقواعد معينة، وتبرير النتائج. من ذلك يتضح أن التفكير يتضمن مهاماً معقدة مرتبطة بموضوع معين، فمثلاً الهدف الأساسي للعلوم إعداد المتعلم للانخراط في تفكير علمي لحل مشكلة، بناء تفسيرات، استخلاص نتائج، اتخاذ قرار، وتعميم معرفته لمواقف جديدة. إضافة إلى التطبيقات المباشرة لمفاهيم العلوم المتمثلة في تطبيقات المجال، بعض



مواقف حل المشكلات المتضمنة لمحتوى غير مألوف أو معقد يتطلب من المتعلم التفكير وفق مبادئ علمية للوصول إلى نتيجة. ربما يشمل الحل جزئية أو تحليل المشكلة إلى أجزاء يتضمن كل منها مفهوماً أو علاقة علمية. وربما يطلب من المتعلمين تحليل مشكلة لتحديد المبادئ أو المفاهيم المتضمنة. وبتكر ويشرح أساليب حل المشكلات. تنقيح حل من مجموعة من المداخل أو استراتيجيات. تطوير مقدرة على الأخذ في الحسبان استراتيجيات بديلة. استخراج نتائج من بيانات علمية وحقائق. وتقديم أدلة على عمليات التفكير الاستقرائي (inductive) والاستنباطي (deductive) وفهم العلاقة بين السبب والنتيجة. **مثل:**

أمثلة	الأفعال
يحلل (Analyze)	يحدد ويصف أو يستعمل العلاقات بين متغيرات أو موضوعات في موقف رياضي. يحلل شكلاً هندسياً لتوضيح حل مشكلة. يحلل مشكلة لتحديد العلاقات. والمفاهيم. ومراحل حلها. يطور ويشرح استراتيجيات حل المشكلات.
يعمم (Generalize)	يوسع المدى لتكون نتائج التفكير وحل المشكلات مكنة التطبيق بإعادة استعمال النتيجة في مستوى أعم وتطبيقات أكثر عمومية.
يركب / يدمج (Synthesize / Integrate)	يركب أو يوحد مجموعة من الإجراءات للوصول إلى نتائج أكثر. عمل ارتباط بين عناصر مختلفة من المعارف. دمج المفاهيم من أحد التخصصات في حل مسائل في تخصصات أخرى.
يبير (Justify)	يقدم تبريراً للحالات التي يرى أنها صحيحة أو خاطئة مع دعمها بأدلة منطقية أو رياضية.



أمثلة	الأفعال
يحل مشكلات غير اعتيادية (Solve Non-routine Problems)	حل مشكلة غير مألوفة ذات علاقة بتخصصة أو مرتبطة بها في واقع الحياة. وتطبيق إجراءات تخصصية في مواقف غير مأوفه قد تتخذ عدة خطوات لحل مشكلة.
يصوغ الفروض / يتنبأ (Hypothesize Predict)	يوحد المعارف مع ما اكتسبه من الخبرة أو الملاحظة لصوغ سؤال يمكن الإجابة عنه بالبحث والاستقصاء. صوغ فروض. ويحلل معلومات علمية وي طرح توقعات مبنية على فهم مبادي علمية.
يصمم / يخطط (Design / Plan)	يخطط أو يصمم بحث للإجابة عن تساؤلات علمية أو اختبار فروض. ويصف خصائص البحث الجيد أو الأساليب العلمية التي يحتاج إليها في تدريسه كأسلوب حل المشكلات والاستقصاء.
يستخلص نتائج (Draw Conclusions)	يستنبط أو يستقري من معلومات معطاة. يقدم استنتاجات صحيحة بناء على أدلة أو فهم لمبادئ ومفاهيم معينة. يستخلص نتائج تجيب عن أسئلة أو على فروض معينة. يبين فهماً للعلاقة بين السبب والنتيجة.
يقوم (Evaluate)	يوازن بين المميزات والعيوب لاتخاذ قرار حول عمليات وأدوات ومصادر بديلة. يأخذ عند الحكم على موضوع ما العوامل العلمية والاجتماعية المصاحبة. يقوم تفسيرات بديلة واستراتيجيات لحل المشكلات. يقوم النتائج في ضوء البيانات العلمية التي تدعمها.
مزيد من الأمثلة	يحلل. يحسب. يختار. يباين. يستنتج. يجرب. يسأل. يخمن. يرتب. ينقد. يفحص. ينظم. يختبر. يرتب. يقارن. ينشئ. يخرع. يناقش. يفترض. يجهز. يركب. يجمع. يبني. يطور. يصيغ. يدير. يخطط. يقترح. يدعم. يقوم. يحكم. يحدد. ينظم. يجادل. يكتب. يبرر. يعاير.





العامل الثاني من عوامل التخطيط للاختبار

إعداد جدول المواصفات (Specifications table)

تعريف جدول المواصفات

«قائمة تربط بين الأهداف والمحتوى من ناحية، وعدد بنود الاختبار التي تمثلها من ناحية أخرى». فهو بذلك مخطط تفصيلي يحدد محتوى الاختبار، ويربط محتوى المقرر الدراسي بالأهداف التعليمية السلوكية، ويبين الوزن النسبي الذي يعطيه المدرس لكل موضوع من الموضوعات المختلفة، والأوزان النسبية للأهداف المعرفية السلوكية في مستوياتها المختلفة.

أهمية جدول المواصفات

يكشف عن تمثيل الاختبار لمحتوى المطلوب للمادة، ويعطي حكماً دقيقاً عن صلاحية الاختبار وخصائص المعلمين، ويعطي مؤشراً واضحاً في قياس الأهداف المراد تحقيقها.

- 1 - يحقق صدق المحتوى لأنه يلزم الفاحص توزيع أسئلته على مختلف أجزاء المقرر وعلى جميع الأهداف.
- 2 - يساعد على الاهتمام بجميع مستويات الأهداف وعدم التركيز على المستويات الدنيا فقط.
- 3 - يشعر المتعلم أن وقته لم يضيع سدى في المذاكرة والاستعداد للاختبار إذ إن الاختبار قد غطى جميع أجزاء المقرر ولم يتناول جزءاً يسيراً منها، وبذلك أخذ جهده في الحسبان.



- 4 - يعطي كل جزء من المقرر وزنه الحقيقي للزمن الذي استغرقه في تدريسه وكذلك بحسب أهميته.
- 5 - الشمول المطلوب في كل اختبار تحصيلي؛ مما يتيح تغطية جميع عناصر المحتوى أو الموضوعات التي تم تدريسها.

◉ خطوات بناء جدول المواصفات

يشتمل جدول المواصفات على بعدين، الأول أفقي، يمثل الأهداف التعليمية السلوكية، والثاني رأسي يمثل موضوعات المقرر الدراسي. وهناك بعض المهام التي ينبغي القيام بها قبل الشروع في بناء جدول المواصفات وهي:

Test

1 - تحديد موضوعات المقرر الدراسي

يمكن بناء جدول المواصفات على مستوى وحدة دراسية، أو منهج تعليمي بأكمله، ولعمل ذلك يُقسم المنهج إلى وحدات تعليمية، وتُقسم الوحدات إلى دروس، وتُقسم الدروس إلى موضوعات. فالموضوعات هي الجزء الأصغر.

2 - تحديد الوزن النسبي لموضوعات المقرر الدراسي عن طريق

- حساب عدد الصفحات لكل موضوع أو درس في المحتوى الخاص بالمقرر الدراسي، ويمثله عادة الكتاب التعليمي أو كتاب المقرر.



■ تقدير عدد الساعات التدريسية أو الحصص الدراسية التي يُدرس كل موضوع أو درس في المقرر الدراسي. يُحسب الوزن النسبي لأهمية الموضوع بقسمة عدد الحصص أو عدد الصفحات على إجمالي عدد الحصص أو عدد الصفحات ويضرب الناتج في مائة.

الوزن النسبي لأهمية الموضوع = عدد حصص الموضوع / العدد الكلي لخصص المقرر × 100

جدول الوزن النسبي لأهمية موضوعات المقرر الدراسي

المحتوى	الموضوع (1)	الموضوع (2)	الموضوع (3)	الموضوع (4)	المجموع
عدد الحصص	1	1	1	1	4
الوزن النسبي (%)	25	25	25	25	100

3 - تحديد الوزن النسبي لأهداف المقرر الدراسي

■ يجري حصر الأهداف التعليمية السلوكية لموضوعات المقرر الدراسي ضمن مرحلة التحليل في خطوات التصميم التعليمي. إذ يعدّ تصنيف بلوم (Bloom) للأهداف في كتابة المشهور (Education Objectives Taxonomy) من أشهر التصنيفات في مجال تعرّف الأهداف التعليمية وتحديدّها؛ فهناك ثلاث مجالات للأهداف التعليمية هي: المجال المعرفي (Cognitive Domain)، والمجال الوجداني (Affective Domain)، والمجال الحركي النفسي (Psychomotor Domain).



■ يُحدد عدد الأهداف في الموضوع الواحد لموضوعات المقرر الدراسي التي أمكن تحديدها في الخطوة السابقة. ويُحسب وزنها بالطريقة نفسها. إذ يُحدد الوزن النسبي لأهداف كل موضوع بقسمة عدد أهداف كل موضوع على العدد الكلي لأهداف المقرر الدراسي. ويضرب الناتج بمئة.

الوزن النسبي لأهمية أهداف الموضوع = عدد أهداف الموضوع / العدد الكلي لأهداف المقرر الدراسي × 100

جدول الوزن النسبي لأهداف موضوعات المقرر الدراسي

المحتوى	الموضوع (1)	الموضوع (2)	الموضوع (3)	الموضوع (4)	المجموع
عدد أهداف الموضوع	3	3	4	3	13
الوزن النسبي (%)	23	23	31	23	100

■ تحديد الوزن النسبي للأهداف السلوكية بمستوياتها المختلفة باستعمال المعادلة

الوزن النسبي للأهداف في مستوى معين = عدد أهداف المستوى / العدد الكلي لأهداف المقرر الدراسي × 100
على أن يراعى التقريب لأعداد صحيحة.

جدول الوزن النسبي لمستويات الأهداف

مستويات الأهداف	تذكر	فهم	تطبيق	تحليل	تركيب	تقويم	المجموع
عدد أهداف الموضوع	8	6	5	3	2	1	25
الوزن النسبي (%)	32	24	20	12	8	4	100



4 - تحديد عدد الأسئلة

■ يُحدد العدد الكلي لأسئلة الاختبار في ضوء الزمن المتاح للإجابة، ونوع الأسئلة، وعمر المتعلم، وغيرها من المتغيرات المؤثرة.

■ يُحدد عدد الأسئلة لكل موضوع من موضوعات المقرر الدراسي في كل مستوى من مستويات الأهداف وفقاً للمعادلة:

عدد أسئلة الموضوع = العدد الكلي للأسئلة × الوزن النسبي لأهمية الموضوع × الوزن النسبي لأهداف الموضوع

⊙ تطبيق على بناء جدول المواصفات

تتكون وحدة تعليمية لأحد المناهج الدراسية من ثلاثة موضوعات رئيسية. يجري تدريسها في ثماني حصص موزعة على الموضوعات كالتالي: الموضوع الأول (3) حصص، الموضوع الثاني (3) حصص، الموضوع الثالث (2) حصتان.

وتحتوي الوحدة على (25) هدفاً سلوكياً: تذكر (8)، فهم (6)، تطبيق (5)، تحليل (3)، تركيب (2)، تقويم (1).

المطلوب بناء جدول مواصفات

خطوات بناء جدول المواصفات لهذه الوحدة

1 - يُبنى جدول كالتالي، ويُحدد الوزن النسبي لكل موضوع بقسمة عدد حصص كل موضوع على إجمالي عدد الحصص ثم نضرب الناتج بمائة مع التقريب فيكون على النحو التالي:



جدول المواصفات به الوزن النسبي للموضوعات

مستويات الأهداف الموضوع	التذكر	الفهم	التطبيق	التحليل	التركيب	التقويم	المجموع
الأول (38)							
الثاني (37)							
الثالث (25)							
المجموع (100%)							

2 - يُحدد الوزن النسبي للأهداف في كل مستوى بقسمة عدد أهداف المستوى على العدد الكلي للأهداف وضرب الناتج بمئة. فيصبح شكل الجدول على النحو التالي:

جدول المواصفات به الوزن النسبي للموضوعات والأهداف

مستويات الأهداف	التذكر	الفهم	التطبيق	التحليل	التركيب	التقويم	المجموع
الموضوع	32	24	20	12	8	4	(100%)
الأول (38)							
الثاني (37)							
الثالث (25)							
المجموع (100%)							



3 - حساب عدد الأسئلة في كل مستوى لكل موضوع من المعادلة التالية؛ على أن يُقَرَّبَ العدد لأعداد صحيحة على نحو متوازن.

عدد أسئلة الموضوع = العدد الكلي للأسئلة × الوزن النسبي لأهمية الموضوع × الوزن النسبي لأهداف الموضوع
فإذا كان عدد الأسئلة في الاختبار (50) سؤالاً (من نوع الاختبار متعدد الخيارات) يصبح جدول المواصفات على النحو التالي:

المجموع	التقويم	التركيب	التحليل	التطبيق	الفهم	التذكر	مستويات الأهداف
(%100)	4	8	12	20	24	32	الموضوع
20	1	2	2	4	5	6	الأول (38)
18	1	1	2	4	4	6	الثاني (37)
12	1	1	1	2	3	4	الثالث (25)
50	3	4	5	10	12	16	المجموع (%100)



قواعد كتابة اختبار الأسئلة متعددة الخيارات

التخطيط للبند

- 1 - ربط كل بند بمؤشر معين.
- 2 - التأكد من أن جذع البند يتناسب من حيث الصوغ اللغوي مع البدائل.
- 3 - تجنب صيغ النفي، وإذا كان لابد فيوضع تحتها خط أو تظلل.
- 4 - تجنب استعمال القيم المطلقة مثل دائماً وجميعاً و فقط وأبداً.
- 5 - حاشي إعطاء أي إشارة أو تلميح للإجابة الصحيحة.
- 6 - جعل كل بند مستقل يعطى معنى في حد ذاته.

كتابة جذع البند

تقديم عبارة تقريرية واضحة بغرض إكمالها أو سؤال للإجابة عنه بأحد الخيارات المطروحة.

كتابة البدائل

- 1 - التأكد أن إجابة واحدة تعد الأفضل أو هي الصحيحة.
- 2 - استعمال مشتتات جذابة ومعقولة تقدم أخطاء شائعة.
- 3 - تجنب استعمال «كل ما تقدم» أو «لا شيء مما سبق».



4. تنظيم البدائل في ترتيب منطقي حرفي أو رقمي.
- 5 - استعمال إجابة واحدة صحيحة وثلاثة مشتتات.
- 6 - جعل البدائل متشابهة في الطول ومختصرة قدر الإمكان.

دليل بناء البنود

1 - جذع السؤال ينبغي أن يعطي معنى واضحاً ويعرض مشكلة محددة

من الأخطاء الشائعة في كتابة اختبار الأسئلة متعددة الخيارات كتابة جذع قصير جداً لا يؤدي إلى معنى ولا يعرض مشكلة ذات علاقة بالبدائل المطروحة. وفي مثل هذه الحالة من الصعب معرفة المراد من الجذع لذا ينبغي أن يحتوي الجذع على الفكرة الأساسية للسؤال بدلاً من عرضها في البدائل.

2 - استعمال حالة النفي في الجذع فقط عند الضرورة

يواجه بعض المتعلمين صعوبة في فهم المعنى في عبارات النفي. والبعض يقرأ البند دون التوقف عند حروف النفي أو ينسى عكس العلاقة المنطقية مجال السؤال. مع أن جملة النفي يوصى بتجنبها إلا أنها قد تكون أحياناً مفيدة. من ذلك تقويم ما إذا كان المتعلم يمكن أن يحدد الخطر ويراعي قواعد السلامة عند العمل.



3 - تجنب التعقيد الزائد

ينبغي عدم تضخيم صعوبة البند بإدخال معلومات صعبة أو معقدة وغير ضرورية في جذع السؤال. على سبيل المثال إذا أردنا من المتعلم أن يحل مسألة تتعلق بتخفيف محلول مستعملاً مفهوم المولارية (*Molarity*) فليس من الضروري استعمال كسور لتعقيد الوصول للحل الصحيح. فالمثال التالي يحتوي معلومات تفصيلية مربكة بينما القيم في المثال الثاني تقيس الهدف نفسه، وتتلافى في الوقت ذاته المعلومات الصعبة وغير الضرورية.

4 - كتابة بدائل واضحة وموجزة

صوغ البدائل في لغة سهلة وواضحة ومختصرة تخفض حالة الارتباك عند المتعلم وتقلل من عبء فهم السؤال. ففي المثال السيء الذي يغلب عليه الحشو غير المرتبط بهدف السؤال يتطلب القراءة أكثر من مرة حتى يمكن فهمه. بينما في المثال التالي تُختصر البدائل لتعطي معنى واضحاً دون أن يؤثر ذلك في معناها.

5 - جعل البدائل مستقلة لا يرتبط أحدها بالآخر

تخلق البدائل المتداخلة حالة من عدم الفهم، وأحياناً أخرى تجعل من السهولة تحديد بعض المشتتات إضافة إلى ذلك إذا كان التداخل يحتوي على الإجابة الصحيحة فذلك يعنى وجود أكثر من إجابة صحيحة.



6 - جعل البدائل متماثلة في الصوغ

إذا كتبت الإجابة الصحيحة في صوغ معين والمشتتات كتبت على نحو مختلف فربما يعطى ذلك المتعلم إحياء بالإجابة.

7 - تجنب استعمال بعض الصيغ المحددة

بعض الكلمات مثل أبدأ ودائماً و فقط عادة ما تُضمن في المشتتات كعامل جذب ولكنها في الغالب تعطى إشارة للمتعلم باستبعادها من الإجابات المحتملة.

8 - يجب أن تُؤلف المشتتات إجابات معقولة ظاهرياً

المشتتات الجيدة ينبغي أن يختارها المتعلمون الأقل كفاءة؛ وبهذا تزداد مقدرة البند على التمييز. ولتحقيق ذلك يجب أن تكون متسقة منطقياً مع الجذع أو مثلة للأخطاء الشائعة بين المتعلمين.

الجدول أنواع المشتتات بناء على وظائفها

المعنى	التمييز	الوصف
يختارها على نحو أكبر الأقل خصيلاً	إيجابي	وظيفي (Functional)
يختارها الأقل والأعلى خصيلاً على نحو متساوٍ أو أن اختيارها ضعيف.	ضعيف/ أو لا يوجد	غير وظيفي
يختارها الأعلى خصيلاً أكثر من الأقل خصيلاً.	سالب	مختل الوظيفة (Dysfunctional)



عموماً، تتضمن بنود اختبار الأسئلة متعددة الخيارات من ثلاثة إلى أربع بدائل. يعتقد البعض أنه كلما ازداد عدد البدائل قلت الفرصة لتخمين الإجابة الصحيحة، وهذا الافتراض بطبيعة الحال صحيح إذا كانت كل البدائل فاعلة، والبنود النموذجية لاختبار الأسئلة متعددة الخيارات تحتوي على الأقل مشتتاً واحداً غير فاعل؛ أي مشتت يفشل في أداء وظيفته فإن البند يعمل كما لو أن هذا المشتت غير موجود. الحل إذاً هو تضمين المشتتات الفاعلة، إذا كان من الممكن التوصل إلى مشتتين أو ثلاثة فاعلة، فتجنب الإصرار على حشو مشتتات ضعيفة لمجرد الحفاظ على أعداد متساوية من البدائل لجميع البنود. توحيد البدائل أمر مصطنع ويخدم فقط إطالة الاختبار دون زيادة المعلومات المتحصلة بالاختبار.

9 - تجنب بدائل «كل ما تقدم» و «لا شيء من ذلك»

هذان البديتان يستعملان على نحو متكرر عندما يجد معد الأسئلة صعوبة في توليد العدد المناسب من البدائل أو المشتتات وتركيزه على الكم بدلاً من الكيف، ولسوء الحظ فاستعمال أي منهما يؤدي إلى خفض كفاءة البند، كما هو موضح بالجدول التالي:



البديل	الاستعمال	الفاعلية
"كل ما تقدم"	إجابة	يمكن تحديد الإجابة إذا اتضح أن إجابتين صحيحتين
	مشنت	يمكن استبعادها إذا اتضح أن أحد البدائل الأخرى غير صحيح
"لاشيء مما تقدم"	إجابة	الإجابة تقيس المقدرة على تعرّف الإجابات غير الصحيحة بدلاً من الصحيحة
	مشنت	المشنت قد لا يبدو مقبولاً ظاهرياً عند بعض المتعلمين

10 - تجنب الإيحاء غير المقصود بالإجابة

مثل أن تكون عبارات البدائل الصحيحة أطول من عبارات الخطأ دائماً. أو توجد كلمة في الجذع تتكرر في البديل الصحيح. أو ألا يتفق أحد البدائل نحوياً مع الجذع.

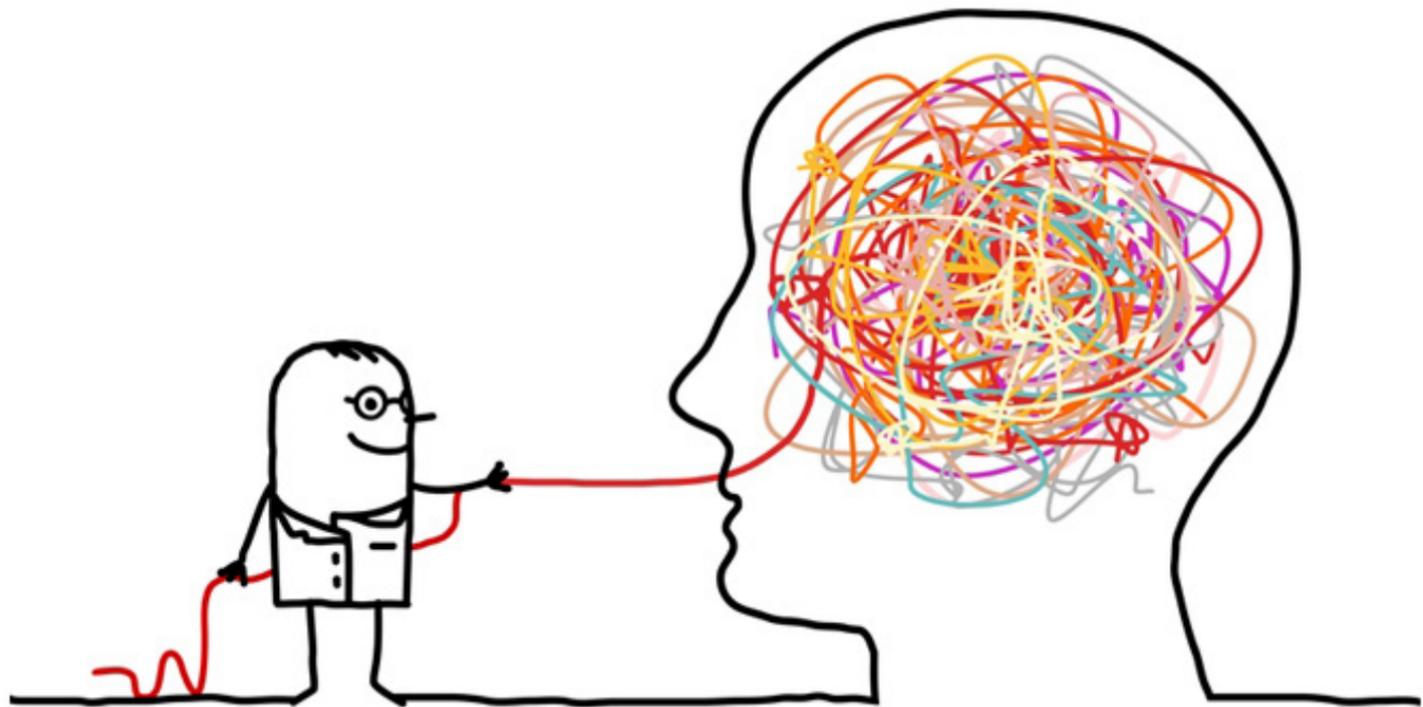
11 - تجنب استعمال الكلمات الصعبة غير الضرورية

إذا كانت الكلمات المستعملة صعبة فإن البند يقيس مقدرة على القراءة إضافة إلى الهدف الذي كتب من أجله البند. بالنتيجة ضعيفي مقدرة في القراءة ربما يحصلون على درجات ضعيفة لتؤكد عدم قدرتهم على تحقيق هدف البند. استعمال الكلمات الصعبة والتقنية تستعمل فقط عندما تكون مهمة لتحقيق هدف البند. وقد أوضح بعض الدراسات أن سهولة الكلمات المستعملة تسهم بنسبة (10%) من سهولة البند.

12 - إذا لم يكن هناك جواب متفق عليه فإن الأفضل استعمال (أي ما يلي) في جذع البند

ملخص لقواعد بناء اختبار الأسئلة متعددة الخيارات

بناء جذع السؤال	إجراءات عامة لكتابة البنود
<p>1 - وضع الجذع في صيغة سؤال أو في عبارة تقريرية (إكمال). 2 - عند استعمال صيغة الإكمال عدم ترك فراغ للإكمال عند بداية الجذع أو وسطه. 3 - التأكد من أن اتجاه الجذع واضح وأن الصوغ واضح إلى حد كافٍ بحيث يعرف المتعلم ما المطلوب منه على نحو محدد. 4 - تضمين الجذع المعلومات الضرورية لجعل المشكلة واضحة. وجنب إضافة معلومات دخيلة. 5 - تجنب استعمال جذور النفي قدر الإمكان. 6 - إدراج الفكرة الرئيسية في جذع السؤال.</p>	<p>1 - استعمل صيغة الإجابة الأفضل أو صيغة الإجابة الصحيحة. 2 - تجنب الصيغ المعقدة. 3 - وضع البدائل عمودياً وليس أفقياً. 4 - مراجعة البنود وتنقيحها. 5 - كتابة البنود بلغة صحيحة وواضحة. 6 - التقليل من وقت المتعلم لقراءة وفهم كل بند. 7 - تجنب استعمال البنود الخادعة التي تقود المتعلم أو تمنعه بالإجابات الخاطئة.</p>
بناء البنود (المحتوى)	بناء البدائل
<p>1 - بناء كل بند بناء على مؤشر محدد. 2 - التركيز على مشكلة واحدة. 3 - جعل العبارات بسيطة في لغتها قدر الإمكان. 4 - تجنب ربط البند بآخر. 5 - تجنب التركيز على معلومات دقيقة جداً عند بناء البند. 6 - تجنب النقل النصي من كتاب أو مذكرة. 7 - تجنب البنود المبنية على الآراء. 8 - تطوير بنود تقيس جميع مستويات المجال المعرفي. 9 - بناء البنود الممثلة لمحتوى مهم وليس لعناصر تافهة.</p>	<p>1 - تضمين البند أربعة خيارات. 2 - وضع البدائل في ترتيب منطقي رقمي أو حرفي. 3 - جعل البدائل مستقلة وغير متداخلة. 4 - جعل كل البدائل متجانسة في مضمونها. 5 - الحفاظ على طول متساو للبدائل. 6 - تجنب استعمال "كل ما تقدم" أو "لا أعرف". 7 - تجنب المشتتات التي يمكن أن تقدم إحاءاً بالإجابة للمتعلم. 8 - تجنب استعمال المحددات مثل دائماً وأبداً.</p>





التحليل الإحصائي للاختبارات

يختبر الفاحص الجيد اختباره أو يمتحن امتحانه بطريقتين: الأولى هي التحليل الإحصائي لنتائج الاختبار، وفيها يعامل الاختبار لوحده عند تحليل نتائجه لكي يكتشف إن كان الاختبار جيداً أم لا. ويجري في الطريقة الثانية تحليلاً للأسئلة، كلاً على حدة، تمهيدا لتحسينه، ثم يحسن الاختبار برمته، فالطريقة الأولى طريقة كليّة، أما الثانية فهي طريقة جزئية، تهتم بالعناصر أو الأسئلة. يقصد بتحليل بنود الاختبار استخراج معامِل الصعوبة، ومعامِل التمييز وتحديد فاعلية البدائل.

٥ فوائد التحليل الإحصائي للاختبارات

- 1 - من خلال تحليل بنود الاختبار يستطيع الفاحص تحديد مواطن الضعف والقوة عند المتعلمين، وتشخيص الأسباب سواء أكانت فنية في الاختبار أم تربوية في طريقة التدريس، فمثلاً: إذا وجدنا أكثر الأخطاء في البنود التي تقيس التطبيق فربما يعود ذلك إلى أن الفاحص لم يكن يعطيهم الوقت الكافي للتدريب وحل التمارين المتعلقة بالمقرر، وبناء على التغذية الراجعة من التحليل يكون العلاج.
- 2 - إن تحليل بنود الاختبار من قبل الفاحص يجعل منه كاتب بنود جيداً، فيتخلص بالمران والتدريب من كثير من العيوب التي تصيب البند سواء من حيث الصوغ أم من حيث الأخطاء الفنية الأخرى.



- 3 - إن بحث وتحليل بنود الاختبار مع المتعلمين كما هي خبرة تعليمية جيدة للفاحص وهي خبرة تعليمية جيدة للمتعلمين. لأنها تخلق لديهم نوعاً من الدقة والانتباه في فحص البدائل قبل اختيار الجواب.
- 4 - يستطيع الفاحص أن يحصل على بنود جيدة من حيث الصعوبة والتمييز فيحتفظ بها ويستفيد منها في الاختبارات القادمة.
- 5 - يقود تحليل بنود الاختبار إلى اختصار طول الاختبار دون أن يؤثر ذلك في ثبات الاختبار. لأن كلاً من البنود السهلة التي يجب عنها كل متعلم أو تلك الصعبة التي لا يجب عنها أحد بنود زائدة لا تحدث أي تأثير في ثبات الاختبار. لأن أي زيادة على الوسط الحسابي أو أي نقصان مائل لا يُغيّر في ترتيب الدرجات؛ وبالتالي لا يؤثر في ثبات الاختبار.
- 6 - إن تعرّف درجات الصعوبة يحدد مواقع البنود في الاختبار. فإذا كانت درجة صعوبة البند عالية (أي البند سهلة) تأتي هذه البند في أول الاختبار. وإذا كانت درجة صعوبتها منخفضة (أي البند صعبة) توضع في نهاية الاختبار. وهكذا توضع كل فقرة في الاختبار في المكان المناسب لها.
- 7 - إن تحليل بنود الاختبار لا يكشف عن درجة الصعوبة والتمييز فقط بل يكشف عن فعالية البدائل (الموهبات). فالبدل الذي يختاره أفراد المجموعة العليا أكثر من الدنيا يعاد النظر فيه. كما يعاد



النظر في كل بديل لم يختره أحد من المتعلمين. أو كانت نسبة من اختاره دون الحد المطلوب.
يشتمل تحليل الاختبارات على ثلاثة مفاهيم أساسية هي:



معامل السهولة أو الصعوبة

هو أحد مؤشرات الصلاحية للاختبارات. وهو نسبة عدد المتعلمين الذين أجابوا إجابة صحيحة عن السؤال أو البند إلى عدد المتعلمين الكلي. فكلما زاد عدد الذين أجابوا إجابة صحيحة عن السؤال كان دليلاً على سهولة السؤال. ويمثل معامل السهولة كسراً يتراوح بين الصفر - الواحد. **مثال:**



في امتحان الإحصاء بلغ عدد طلاب الصف (50) طالباً، وعدد الذين أجابوا إجابة صحيحة عن السؤال الثالث (20) طالباً، ما معامل سهولة السؤال الثالث؟

معامل السهولة = عدد الذين أجابوا إجابة صحيحة / عدد المتعلمين

$$0.40 = 50 / 20 =$$

المثال نفسه: الإجابات الصحيحة (40)، عدد المتعلمين (60)

$$0.67 = 60 / 40 =$$

ويمكن كذلك استعمال المجموعتين العليا والدنيا لاستخراج معامل السهولة كالتالي:

معامل السهولة = نسبة المتعلمين من المجموعة العليا الذين أجابوا إجابة صحيحة + نسبة المتعلمين من

المجموعة الدنيا الذين أجابوا إجابة صحيحة / 2

ويمكن أن يكون:

معامل السهولة = عدد الإجابات الصحيحة عن السؤال / عن الإجابات الصحيحة والخاطئة

ملاحظة: يكون البند الذي تتراوح نسبة سهولته (25 - 75%) مقبولاً، وكلما اقتربت النسبة من (50%)

تكون مناسبة.

معامل الصعوبة = عدد الذين أجابوا إجابة خاطئة / عدد المتعلمين



مثال: بعد إجراء امتحان على (100) متعلم، كانت إجابات (90) منهم عن السؤال الرابع خاطئة، ما معامل صعوبة السؤال الرابع؟

نسبة الصعوبة = عدد الذين أجابوا إجابة خاطئة / عدد المتعلمين * 100

$$10\% = 100 * 100 / 60 =$$

ملاحظة: معامل السهولة + الصعوبة = 1

مثال: إذا كان معامل السهولة للسؤال الأول في امتحان علم التدريب (0.75) أوجد معامل الصعوبة لذلك السؤال؟

Test

$$0.25 = 0.75 - 1 =$$

مثال: أجابت جماعة من المتعلمين عن اختبار لقلق المنافسة، والمكون من بنود صح وخطأ، فكانت إجاباتهم كما يوضحه الجدول التالي، احسب معامل السهولة والصعوبة لكل فقرة.

رقم البند	عدد الإجابات الصحيحة	عدد الإجابات الخاطئة	عدد الذين لم يحاولوا الإجابة
1	35	15	0
6	25	20	5
8	6	40	6

$$\text{معامل الصعوبة للبند (1)} = 100 * 50 / 15 = 30\%$$



$$\text{معامل السهولة} = 100 - 30 = 70\%$$

$$\text{معامل الصعوبة للبند (6)} = 100 * 45 / 20 = 44.4\%$$

$$\text{معامل السهولة} = 100 - 44.4 = 55.6\%$$

$$\text{معامل الصعوبة للبند (8)} = 100 * 46 / 40 = 86.9\%$$

$$\text{معامل السهولة} = 100 - 86.9 = 13.1\%$$

الجدول معايير الصعوبة

مستوى الصعوبة	مدى القيم
صعب جداً	صفر - 0.34
معتدل الصعوبة	0.35 - 0.59
معتدل السهولة	0.60 - 0.84
سهل جداً	0.85 - 1

ملاحظة: لو كان لدينا (10) أسئلة في اختبار ما. أو (10) بنود في استبانة. ما معامل السهولة أو الصعوبة الملائم؟ يجب أن يتراوح بين (0.30 - 0.80).
يتحقق معامل السهولة أو الصعوبة إحصائياً وفقاً لطبيعة الاختبار (بدني، مهاري، مقياس نفسي، معرفي، ...). نوع البيانات، وحجم العينة.



معامل التمييز

يرتبط معامل التمييز إلى درجة كبيرة بمعامل الصعوبة. فإذا كان الغرض من الاختبار هو أن يفرق بين القادرين وأولئك الأقل مقدرة من المتعلمين فإن السؤال المميز هو ما يقود إلى هذا الغرض. إذ إن مهمة معامل التمييز ينبغي أن تتمثل في تحديد مدى فاعلية سؤال ما في التمييز بين المتعلم ذي المقدرة العالية والمتعلم الضعيف بالقدر نفسه الذي يفرق الاختبار بينهما في الدرجة النهائية بصورة عامة.

هناك طرائق كثيرة لحساب معامل التمييز ونكتفى بذكر أحدها وتلخص في الآتي:

1 - ترتب أوراق المتعلمين تصاعدياً بحسب الدرجات. ونفترض أن عددها (100) ورقة.

2 - تقسم الأوراق إلى مجموعتين عليا ودنيا. وتمثل أعلى (27%) من الأوراق ذات الدرجات العليا. وأدنى (27%) منها ذات الدرجات الدنيا. وبذلك يكون:

عدد أفراد المجموعة العليا = عدد أفراد المجموعة الدنيا = 27

إذا كان العدد قليلاً يمكن تقسيم الأفراد إلى مجموعتين: أعلى (50%) وأدنى (50%)

3 - يُحدّد عدد المتعلمين الذين أجابوا عن السؤال الأول إجابة صحيحة من بين أولئك الذين حصلوا على الدرجات العليا. ونفترض أنهم كانوا (20) طالباً.



4 - يُحدّد عدد الذين أجابوا عن السؤال الأول إجابة صحيحة من بين أولئك الذين حصلوا على الدرجات الدنيا. ونفترض أنهم كانوا (12) طالباً.

5 - نطرح الخطوة (4) من الخطوة (3). فالفرق بين المجموعتين (20 - 12 = 8)

6 - يقسم الناتج من الخطوة (5) على عدد أفراد إحدى المجموعتين.

$$\text{معامل التمييز} = \frac{\text{س} - \text{ص}}{\text{ن}}$$

Test

حيث:

س: عدد طلاب الفئة العليا في التحصيل الذين أجابوا عن السؤال إجابة صحيحة.

ص: عدد طلاب الفئة الدنيا في التحصيل من أجابوا عن السؤال إجابة صحيحة.

ن: عدد أفراد إحدى المجموعتين.

$$\text{معامل التمييز} = \frac{12 - 20}{27} = 0.30$$

وهذا معامل تمييز مناسب. وإن كان لا يصل في دقته إلى معامل تمييز (0.60) أو (0.70). ويكون معامل التمييز مرتفعاً أو منخفضاً أو سلبياً.



مثال: نفرض أن مجموع الدرجات التي حصلت عليها الفئة العليا من المتعلمين للسؤال الأول في اختبار مادة الإحصاء الرياضي (70) درجة. ومجموع الدرجات التي حصلت عليها الفئة الدنيا من المتعلمين للسؤال نفسه (38) درجة. وعدد أفراد المجموعة العليا = (8) أفراد. وعدد أفراد المجموعة الدنيا = (8) أفراد. ودرجة السؤال المخصصة لهذا السؤال (10) درجات. احسب معامل التمييز.

$$0.4 = \frac{4}{10} = \frac{32}{80} = \frac{38 - 70}{8 * 10} = \text{معامل التمييز}$$

لا توجد قاعدة لقبول السؤال أو البند اعتماداً على التمييز. ولكن يمكن القول إن:

معامل التمييز	التقويم
أقل من 0.19	ضعيفة جداً وحذف
صفر إلى 0.19	ضعيفة التمييز وينصح بحذفها أيضاً
20 إلى 0.39	ذات تمييز مقبول وينصح بتحسينها (إما بتعديل جذع السؤال أو البدائل)
أعلى من 0.39	جيدة التمييز
أي بند معامل تمييزه سالب يحذف	

وعموماً، كلما زاد معامل التمييز كان ذلك أفضل.



فاعلية البدائل

■ في الأسئلة ذات اختبار الأسئلة متعددة الخيارات تكون الخيارات بدائل محتملة للإجابة عنها، ويوجد بديل واحد هو الإجابة الصحيحة. ويفترض أن البدائل الأخرى تمثل إجابات محتملة وأنها تجذب المتعلمين. ولا بد أن يحدث اختيار أي بديل من البدائل الخاطئة من قبل طالب واحد على الأقل أو بنسبة لا تقل عن (5%) من المتعلمين. على أن تكون غالبيتهم من الفئة الدنيا.

■ البدائل التي لا تجذب أحداً منهم أو القليل جداً منهم هي بدائل غير فعّالة ويفضل استبدالها.

■ حذف البدائل التي تجذب من طلاب المجموعة العليا أكثر مما تجذبه من طلاب المجموعة الدنيا. أي إنه يميز باتجاه معاكس لتمييز البند.

■ كلما كانت قيمة البديل سالبة كانت دليلاً على أن المشتت جيد.

تفسير معامل الصعوبة والتمييز

إن أفضل معامل الصعوبة في غياب عنصر التخمين هو (0.50) لأنه يكشف أكبر تباين للبند أي إن البند توضح التباين الواسع في قدرات المتعلمين.



* **تباين البند:** حاصل ضرب معامل الصعوبة بمعامل السهولة إذ إن معامل السهولة هو عدد الذين لم يجيبوا عن البند إجابة صحيحة على العدد الكلي للمتعلمين.

عندما تكون صعوبة البند تساوي (0.50) فإن هذا يدل على أن نصف الممتحنين قد أجابوا عن البند إجابة صحيحة والنصف الآخر أجاب إجابة خاطئة، وبناء على هذه النتيجة يمكن أن نتوقع الآتي:

- إن نصف المتعلمين الذين أجابوا إجابة صحيحة عن البند هم طلاب الطبقة العليا، ونصف المتعلمين الذين أجابوا إجابة خاطئة عن البند هم طلاب الطبقة الدنيا. وفي هذه الحالة تميز هذه البند يساوي (+1) وهو أفضل وأعلى تمييز للفقرة.
- إذا وجد عكس ذلك وهو أن المتعلمين الذين أجابوا إجابة صحيحة عن البند هم طلاب الطبقة الدنيا فإن تمييز البند يساوي (-1) وهو أسوأ وأدنى معامل تمييز للفقرة.
- إذا تساوى عدد الذين أجابوا عن البند إجابة صحيحة من أفراد الطبقة العليا مع أفراد الطبقة الدنيا فهذا يدل على أن معامل صعوبة البند يساوي (0.50)، ولكن تمييز البند في هذه الحالة قد يأخذ أي قيمة بين (+1 و -1)، لذلك فإنه من الضروري إيجاد معامل تمييز البند إضافة إلى معامل صعوبتها للحكم على فعاليتها.



عندما يكون عامل التخمين جزءاً مكوناً لدرجة المتعلم فإننا نحكم على فعالية البند في ضوء معامل الصعوبة اعتماداً على نتائج المعادلة الآتي:

$$\text{أفضل معامل صعوبة} = \frac{\text{أعلى احتمال للإجابة الصحيحة} - \text{احتمال الإجابة صحيحة بالصدفة}}{2}$$

معامل الصعوبة	عدد البدائل
0.75	2
0.67	3
0.62	4
0.60	5

نلاحظ من الجدول السابق أنه كلما زاد عدد البدائل قلت احتمالية التخمين. وبالتالي اقترب معامل الصعوبة الأفضل نحو القيمة (0.50)



ويمكن توضيح العلاقة بين معامل الصعوبة والتمييز على النحو الآتي:

أعلى معامل للتمييز	نسبة الإجابات الصحيحة (%)
صفر	100
0.2	90
0.4	80
0.6	70
0.8	60
1	50
0.8	40
0.6	30
0.4	20
صفر	صفر

يلاحظ أن معامل التمييز تزداد قيمته كلما اقترب معامل الصعوبة من (0.50).





تطبيقات

أمثلة عن أسئلة سيئة:

1 - مفاهيم مكافحة العدوى وسلامة المريض

المينو (نموذج عن سؤال سيء)

- ا. قوته على قتل الجراثيم تعادل الفيوسيدين.
- ب. قوته على قتل الجراثيم أقل من الفيوسيدين.
- ج. يعني تأمين وسط رطب للجرح.
- د. كل ما أعلاه صحيح.
- هـ. كل ما أعلاه خطأ.

Test

2 - العناية بالجروح وخياطتها

لسبر المقدرة على العناية بالجروح: الإنضار: (نموذج عن سؤال سيء)

- ا. باستئصال النسيج النخرة قطعة واحدة.
 - ب. بطريقة الضماد الرطب - الجاف.
 - ج. بتكرار الغسل بالماء الجاري.
- .a + b.d
.b + c.e



3 - نموذج في الثقافة الطبية العامة

- لتجنب العدوى أثناء الإصابة بجرح غائر في اليد، ينبغي عليك:
- التردد دائماً على المركز الصحي لأخذ تطعيمات ضد الكزاز.
 - تلقى علاج بالمضادات الحيوية ولكن فقط إذا كان الجرح مؤلماً.
 - التأكد من عدم دخول جسم غريب في الجرح.
 - عدم مسح الجرح ابداً بمواد كحوليه إلا إذا كان لا يزال ينزف.

Test

4 - نموذج في الإحصاء

ما قيمة المتوسط الحسابي للأعداد (3, 4, 5, 9, 9, 18)؟

- 80 .a
- 70 .b
- 60 .c
- 8 .d



5 - نموذج في القياس والتقويم

أي مما يأتي أحد تصنيفات المجال المعرفي؟

a. المقدرة على التفكير.

b. التفكير النقدي.

c. الحفظ.

d. كل ما تقدم.

e. لا شيء مما تقدم.

Test

6 - نموذج في المنهج العلمي

عند القيام ببحث تربوي، أي فهرس مما يأتي يعد الأفضل لتحديد المقالات المناسبة؟

a. الدليل للمصادر العلمية في التربية.

b. الفهرس الحديث للمجلات التربوية.

c. مراجع في التربية.

d. الموسوعة العالمية في التربية.



أمثلة عن أسئلة جيدة مع تصحيح بعض الأمثلة السيئة السابقة

1 - نموذج جيد لسؤال صيغ من أجل سبر المقدرة على تشخيص سرطان الكولون

ذكر (60) سنة يشكو من إمساك مفاجئ منذ شهرين يترافق مع براز مدمّى. ما هو التشخيص الأكثر احتمالاً:

a. بواسير.

b. شق شرجي.

c. التهاب كولون قرحي.

d. تشنج كولون.

e. سرطان كولون.

Test

2 - نموذج جيد لسؤال صيغ من أجل سبر المقدرة على تدبير المرضى

يعالج داء أديسون بالأدوية الآتية:

a. ستيروئيدات سكرية ومعدنية.

b. سيبروترون أسيتات.

c. أندروجينات.



d. ستيروئيدات جنسية.

e. سبيرونولاكتون.

3 - لتجنب العدوى في أثناء الإصابة بجرح غائر في اليد، ينبغي عليك:

a. الذهاب إلى المركز الصحي للتطعيم ضد الكزاز (*tetanus*).

b. تلقي علاج بالمضادات الحيوية إذا كان الجرح مؤلماً.

c. التأكد من عدم دخول جسم غريب في الجرح.

d. مسح الجرح بمواد كحولية إذا لم يكن الجرح ينزف.

4 - ما قيمة المتوسط الحسابي للأعداد (3، 4، 5، 9، 9، 18)؟

a. 5.

b. 7.

c. 8.

d. 9.



5 - ما أكثر المستويات تعقيداً في تصنيف المجال المعرفي؟

- a. المعرفة.
- b. التركيب.
- c. التقويم.
- d. التحليل.

6 - عند القيام ببحث تربوي، أي مما يأتي يعد الأفضل لتحديد المقالات المناسبة؟

- a. الدليل للمصادر العلمية في التربية.
- b. الفهرس الحديث للمجلات التربوية.
- c. مراجع في التربية.
- d. الموسوعة العالمية في التربية.



الملاحق

ملحق (1)

بطاقة كتابة البند الاختباري

الرقم:

التخصص:

الموضوع:

المعيار العام:

المعيار الفرعي:

رقم المؤشر:

المجال المعرفي للبند:

نص البند:

أ -

ب -

ج -

د -

هـ -

الإجابة الصحيحة: (أ ب ج د)
مستوى الصعوبة المتوقع: (80%) (60%) (40%) (20%)
الزمن المتوقع للإجابة: () دقيقة



ملحق (2)

مراجعة البنود

راجع كل بند وفقاً للأسئلة الآتية	
()	هل كتب البند لقياس هدف واحد؟
()	هل بُني البند على أساس مشكلة محددة بينت بوضوح في الجذع؟
()	هل تضمن الجذع البيانات اللازمة دون حشو زائد؟
()	هل صُيغ الجذع بطريقة مثبتة أو تم إبراز حرف النفي؟
()	هل كتبت البدائل بوضوح وإيجاز؟
()	هل البدائل غير متداخلة؟
()	هل البدائل متجانسة في المحتوى؟
()	هل البدائل خالية من إحصاءات بالاجابة؟
()	هل البدائل خالية من "كل ما ذكر" و "لاشيء مما ذكر"؟
()	هل تضمن البند إجابة واحدة صحيحة أو الإجابة الأفضل؟
()	هل وزعت الإجابات الصحيحة على نحو عشوائي بين البنود؟
()	هل البند والبدائل خالية من الأخطاء اللغوية والإملائية؟
()	هل تضمن السؤال بنود صعبة أو الغا؟



ملحق (3)

قائمة تحكيم البند الاختباري من نوع اختبار الأسئلة متعددة الخيارات
رقم البند () المقرر () رقم معدّ الأسئلة

م	خاصية البند	نعم	لا	التطوير المقترح
1	دقيق من الناحية العلمية			
2	يقيس المؤشر المحدد			
3	المجال المعرفي المحدد صحيح			
4	له إجابة واحدة صحيحة فقط			
5	مستوى صعوبته المتوقع معقول			
6	لا يوجد فيه خيز			
7	كتب البند لقياس هدف واحد			
8	بني البند على أساس مشكلة محددة بينت بوضوح في الجذع			
9	تضمن الجذع البيانات اللازمة دون حشو زائد			
10	الجذع صيغ بطريقة مثبتة أو نفي بارز			
11	كتبت البدائل بوضوح وموجز			
12	البدائل غير متداخلة			
13	البدائل متجانسة في المحتوى			



م	خاصية البند	نعم	لا	التطوير المقترح
14	البدائل خالية من إبهامات بالإجابة			
15	تضمنت البدائل "كل ما ذكر" و "لا شيء مما ذكر"			
16	تضمن البند إجابة صحيحة واحدة أو الأفضل			
17	البند والبدائل خالية من الأخطاء اللغوية والإملائية؟			
18	تضمن السؤال بنوداً صعبة أو ألغازاً			

ملاحظات أخرى عن البند:

.....

مستويات المجال المعرفي:
المعرفة التطبيق التفكير

الحكم النهائي على السؤال:
مقبول بدون تعديل مقبول بعد تعديل طفيف مقبول بعد تعديل جذعي غير مقبول
نص البند بعد التعديل:

.....

.....

مبررات رفض البند:

.....

.....





المراجع (References)

1. أحمد عبد الرحمن (2011). تصميم الاختبارات. دار أسامة. الأردن.
2. أبو زينة. فريد كامل (1998). أساسيات القياس والتقويم. ط2. الفلاح للنشر. الكويت.
3. أبو علام. رجاء محمود (1987). قياس وتقويم. دار القلم. الكويت.
4. الإمام. مصطفى محمود وآخرون (1987). التقويم والقياس. ج1. جامعة بغداد. العراق.
5. باهي. مصطفى (1999). المعاملات العلمية والعملية بين النظرية والتطبيق. الثبات. الصدق. الموضوعية. المعايير. ط1. مركز الكتاب للنشر. القاهرة. مصر.
6. الظاهر. زكريا محمد (1999). مبادئ القياس والتقويم في التربية. مكتبة دار الثقافة للنشر والتوزيع. عمان.
7. صلاح الدين محمود علام (2006). الاختبارات والمقاييس التربوية والنفسية. عمان. دار الفكر للنشر والتوزيع.
8. عبد الواحد الكبيسي (2007). القياس والتقويم تجديداً ومناقشات. عمان. دار الجرير للنشر والتوزيع.
9. عودة. أحمد سليمان (1985). القياس والتقويم في العملية التدريسية. المطبعة الوطنية. دار الأمل. إربد. الأردن.



10. علي الزامل، عبد الله الصارمي، علي كاظم (2009). مفاهيم وتطبيقات في التقويم والقياس التربوي. مكتبة الفلاح، الكويت.
11. الغريب، رمزية (1985). التقويم والقياس النفسي والتربوي. مكتبة الأجلو المصرية، القاهرة، مصر.
12. فرج، صفوت (1989). القياس النفسي، ط2، مكتبة الأجلو المصرية، القاهرة، مصر.
13. الفريق العلمي (1426 هـ). مواصفات البنود الاختيارية (من نوع الاختبار متعدد الخيارات) الجيدة ومعايير مراجعتها. برنامج اختبار الكفايات الأساسية للمعلمين لدول الخليج العربية. مكتب التربية العربي لدول الخليج.
14. البستنجي، محمود محمد (2010). كتاب القياس والتقويم للمعلم بين النظرية والتطبيق. المملكة العربية السعودية جدة. خوارزم العلمية للنشر والتوزيع.
15. النبهان، موسى (2004). أساسيات القياس في العلوم السلوكية. عمان: دار الشروق للنشر والتوزيع.
16. ملحم، سامي محمد (2000). القياس والتقويم في التربية وعلم النفس. دار المسيرة للنشر والتوزيع والطباعة، الأردن.



17. Cheung, D. Bucat, R. (2002). *How can we construct good multiple-choice Items?. June 20 – 21. Paper presented at the science and technology education conference, Hong Kong.*
18. Garden, R. A., Lie, S., Robitaille, D. F., Angell, C., Martin, M. O., Mullis, I. V. S., et al., (2007). *TIMMS Advanced 2008 Assessment Framework: IEA.*
19. Mullis, I. V. S., Martin, M. O., Foy, P. (2005). *IEA,s TIMSS 2003 International report on achievement in the mathematics cognitive domains findings from a developmental project, IEA.*
20. Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O, Sullivan, C. Y., Arora, A., Erberber, E., (2007). *TIMSS 2007 assessment framework: IEA.*
21. Mullis, I., Martin, M. (June 25 – 28). *TIMSS 2003 cognitive domains: Country difference in knowing, applying, and reasoning mathematically. Paper presented at the CCSSO national conference on large-scale assessment, San Francisco. Organization for economic co-operation development (OECD). (2003). The PISA 2003 assessment framework.*
22. Steven J., Burton, Richard R. Sudweeks, Paul F. Merrill, Wood, B., (1991). *How to prepare better multiple-choice test Items: guidelines for university faculty. Brigham Young University testing services and the department of instructional science.*
23. Zimmaro, D. M., (2004). *Writing good multiple-choice exams. Austin: Measurement and evaluation center.*