

الفصل الرابع

تأليف الاختبارات

خطوات عمل الاختبار

١ - تحديد الأغراض : قبل البدء بعمل اختبار ما يجب أن نعرف بالتحديد ماذا يراد قياسه ؟ وما هو الغرض الذي يراد الوصول إليه من تطبيق هذا الاختبار ؟ ومن هم الأفراد المراد اختبارهم ؟ من أي سن ؟ ومن أي مستوى تعليمي ؟ وما هي الطريقة التي يراد بها تطبيق الاختبار ؟ هل سنتلق أسئلته شفها أم تحريريا ؟ وهل سيجرى بطريقة جمعية أم فردية ؟ وكيف سيظهر المفحوص عن إجاباته ؟ إلى غير ذلك من الاعتبارات التي تتحدد بها أغراض الاختبارات والنواحي المراد قياسها .

٢ - مادة الاختبار : وفي ضوء هذه الاعتبارات يمكن تحديد نوع المادة المناسبة التي سيتمكون منها الاختبار ، والمجال الذي يصح أن يخصص فيه اختبار محتوياته وأسئلته فنأخذ في حصر الموضوعات التي يراد اختبارها . . . وتتبع ذلك بعمل تصنيف أو تقسيم لهذه الموضوعات وتفصيل كل نقطة . ويحسن هنا عمل جدول يوضح أقسام الاختبار وموضوعات كل قسم أو استعمال نظام بطاقات التصنيف والترتيب في تبويب أجزاء الاختبار .

٣ - وضع الأسئلة : يوضع لكل من الفقر السابق تحديدها عدد من الأسئلة مع مراعاة التوازن بين النواحي التي يقيسها الاختبار ؛ ويراعى في وضع الأسئلة اعتبارات كثيرة منها : -

- (أ) الابتعاد عن الأسئلة الملتوية .
(ب) الابتعاد عن نقل عبارات معينة من الكتب في صيغة أسئلة .
(ج) أن يكون كل سؤال مستقلا عن الآخر بحيث لا تبنى الإجابة عن سؤال معين على الإجابة عن سؤال سابق .
(د) الابتعاد عن الأسئلة البديهية أو الدافئة .
(هـ) أن تكون اللغة صحيحة وواضحة ومشوقة .

ويراعى في تصميم الأسئلة محاولة صياغتها في أصح الصور التي تضمن الوصول إلى الفرض من القياس ، كأن يكون بعضها من النوع الذي تطلق فيه الحرية للإجابة ، وبعضها مما يضمن فيه اختيار إجابة واحدة من مجموعة اجابات معطاه ، وبعضها من تكميل الجمل إلى غير ذلك

٤ — تنظيم ترتيب الأسئلة : ثم ترتب الأسئلة بحسب تقدير سهولتها وصعوبتها . على أن هذا الترتيب لا يصح أن يكون نهائيا إلا بعد التجريب وتحليل النتائج . ويراعى في ترتيب الأسئلة تنوع مادتها وأسلوبها حسب الصورة التي تناسب تصميم الاختبار .

٥ — وضع تعليمات الأسئلة : يوضع في كل سؤال التعليمات المناسبة للإجابة . من حيث مكان الإجابة وطريقة إجابتها . وإذا كان للاختبار أقسام متشابهة فيصح أن يبدأ كل قسم منها بمثال توضيحي يساعد على فهم التعليمات والمطلوب من الأسئلة .

٦ — تجهيز مفتاح الاختبار : يوضع نموذج للإجابة عن كل سؤال والدرجة التي يصح اعطاؤها عند تصحيحه . مع عمل التعديلات التي تظهر الحاجة إليها في بعض الأسئلة بالنسبة لتسهيل أسلوب التصحيح .

٧ — تجريب الاختبار : يطبق الاختبار على مجموعة صغيرة من الأفراد على سبيل التجربة ، وتعمل الاحصاءات التحليلية للنتائج ، وهذه الخطوة العملية تفيد كثيرا في توضيح مفهوم الأسئلة ودرجة صعوبتها وما تحتمله لفتها من تأويل .

٨ — تعديل الاختبار : تعمل التعديلات المناسبة من حيث صياغة الأسئلة وإعادة ترتيبها ، وما يتطلبه التعديل من حذف أو إضافة — على أن تتكرر عملية تجريب الاختبار وإعادة تعديله عدة مرات حتى يصل إلى درجة يمكن الاطمئنان عليها .

٩ — تطبيق الاختبار : يقصد الوقوف على مدى صلاحية وثبات نتائجه ، ومدى تحقيقه لشروط الاختبار الجيد . ويحتاج ذلك لمجهود ووقت كبير وهذا هو ما يطلق عليه « تقنين الاختبار » .

١٠ — عمل معيار للاختبار : وهذا يتضمن تحديد المستويات المناسبة للأعمار المختلفة ، والدرجات التي يحصل عليها متوسط الأفراد لسكل عمر — ويصح أن تكون هذه المعايير Norms في صورة جداول أو رسوم بيانية .

شروط الاختبار الجيد

الاختبار الجيد باختصار هو الذي يصلح لأداء الغرض الذي وضع من أجله ، وهذا يتم بالشروط الآتية :

١ — أن يقيس الاختبار ما هو مفروض أن يقيسه ، وهذا هو ما يعبر عنه « بصحة الاختبار »

٢ — أن يكون هذا القياس دقيقا بحيث يعطى نفس النتائج بطريقة ثابتة ، وهذا هو ما يعبر عنه « بثبات الاختبار » .

٤ — أن يكون الاختبار « موضوعيا » بمعنى أنه يناسب مستوى الأفراد المراد تطبيق القياس عليهم .

٥ — أن يكون الاختبار صالحا للتمييز بين الأفراد ذوي المستوى العالي والمستوى الضعيف ، وهذا هو ما يعبر عنه « التمييز » .

٥ — أن يكون الاختبار شاملا بحيث يغطي معظم نقط الموضوع المراد الاختبار فيه ، وبحيث تكون أسئلته عينة ممثلة للموضوع كله ، وهذا هو ما يعبر عنه « بالشمول » .

٦ — أن يكون من السهل استعماله وتطبيقه ، من حيث الإجراء والتصحيح ومعالجة النتائج .

وسنتناول كل واحد من هذه الشروط بشيء من التفصيل :

١ — صلاحية الاختبار :

الاختبار الصالح هو الذي يقيس ما هو مفروض أن يقيسه تماما ، وهذا ينطبق على السؤال الواحد ، فصلاحية سؤال ما في الاختبار تتوقف على مدى قياسه للناحية المفروض أنه وضع لقياسها . فإذا كان الاختبار يقيس التحصيل فينبغي ألا تكون الأسئلة مشبهة بعامل الذكاء العام بحيث تكون النتائج توضح قوة الذكاء لا قوة التحصيل ، وإذا كان المراد قياس الذكاء فلا يصح أن يتأثر الاختبار بعامل اللغة أو المعلومات الثقافية .

ولكن هناك عوامل تكاد تكون ثابتة تؤثر في صحة الاختبار مثل : القدرة على القراءة ، نقص الحصول اللغوي ، ضعف الذاكرة ، بطء الرد والاستجابة . . . فإذا لم يكن الاختبار نقيًا من هذه العوامل وأمثالها فإن

الدرجة الممنوحة عن الاختبار لا تعبر عن المطلوب قياسه تماماً ، ويكون الاختبار معيباً من ناحية صلاحيته .

وقد يكون الاختبار صالحاً للقياس على مجموعة معينة من الأفراد ، ولكنه يعتبر غير صالح بالنسبة لمجموعة أفراد مختلفين عن المجموعة الأولى ، فقد يكون الاختبار في الحساب مثلاً صالحاً لقياس القدرة الحسابية في فصل معين يقوم بالتدريس له مدرس معين ، ولكنه لا يعتبر صالحاً لفصل آخر يدرس له مدرس آخر لتدخل عامل المدرس هنا .

وإذن يجب أن تكون صلاحية الاختبار مقيدة بظروف معينة مثل حالة التلاميذ وظروف إجراء الاختبار .

ولهذا نجد أن معظم الاختبارات التحصيلية المطبوعة لا تصالح صلاحية كاملة للقياس في المدارس المختلفة ، ولهذا يفضل دائماً أن تعمل كل مدرسة ، — بل كل مدرس — لفصوله اختبارات خاصة ، لأنها بلا شك ستكون أكثر صلاحية من هذه الناحية .

ولتقدير مدى صلاحية الاختبار يصح مقارنة نتائجه بنتائج اختبار آخر سبق أن ثبتت صلاحيته لقياس نفس النواحي والأغراض بالضبط ، كما يحدث عند محاولة عمل مسطرة أو مقياس جديد فيصح التأكد من صحة صلاحيته للقياس بأن تقارن النتيجة بنتائج قياس مسطرة سبق تدريبها وسبق صلاحيتها (مع ملاحظة الفرق في الدقة في الحالتين طبعاً) .

إذن يجب أن يكون لدينا أساس سابق كاختبار معروف صلاحيته لنقارن به الاختبار الجديد . . . وبحساب معامل الارتباط بينهما نحصل على ما يسمى « معامل صحة الاختبار » ، وإذا تعذر وجود مقياس أو اختبار سابق فيصح المقارنة بالنقط المحددة في بداية عمل الاختبار وأن نلجأ إلى بعض الخبراء

في هذه النواحي ليفتونا في مدى صلاحية الاختبار لما هو مفروض أن يقيسه ،
وعمل التعديلات المناسبة. ولكن هذه طريقة بدائية وتعتمد على العامل الذاتي .
أما صلاحية أسئلة الاختبار ذاتها سؤالا بسؤالا فيمكن تحسينها بأن
نسال أنفسنا هل يصلح كل سؤال لقياس ما هو مفروض أن يقيسه .

٢ - ثبات الاختبار :

ثبات الاختبار دليل الوثوق بنتائجه ، واستمرار هذه النتائج - في حالة
أجرائه أكثر من مرة - بحيث لا يحدث تغير في الدرجات التي يأخذها كل
تلميذ إذا أجرى الاختبار مرتين على نفس التلاميذ . على شرط ألا يحدث
تمرين أو تعلم بين المرتين طبعاً .

ومن الواضح أن عامل ثبات الاختبار مرتبط بصحة الاختبار ، فإذا كان
الاختبار ذا صلاحية عالية فلا بد أن يكون ذا معامل ثبات مرتفع أيضاً ، فإذا
كان الاختبار يقيس فعلاً ما هو مفروض أن يقيسه فستكون نتائجه ثابتة طبعاً ،
ولكن يصح أن يكون معامل الثبات عالياً ومع ذلك يكون الاختبار غير
صديق في صلاحيته وصحته .

ولسكى يكون الاختبار ذا صحة جيدة يجب أن يكون معامل ثباته مرتفعاً ،
أي أنه يجب أن يكون الاختبار ذا معامل ثبات عالى لسكى يكون صحيحاً .
ولكن قد يكون الاختبار ذا ثبات عالٍ ولكنه لا يكون ذا معامل صحة مرتفع
إذ ليس من الضروري أن يقيس ما يراد قياسه .

وهناك عوامل كثيرة تتدخل في صحة الاختبار مثل القراءة والسرعة
والفهم . الخ . وهذه عوامل ثابتة ، أما العوامل التي تؤثر في ثبات الاختبار

فهو ظروف متغيرة في كل مرة من مرات الإختبار وتختلف النتائج مختلفة مثل :
حالة التلاميذ النفسية ، زمن اجراء الإختبار ، طريقة الإجراء ، طريقة التصحيح
وما قد يكون فيها من رأى شخصي ، الظروف المحيطة بإجراء الإختبار من
حرارة وهدوء الخ .

طرق تعيين ثبات الإختبار :

١ - طريقة الصور المتكافئة : يعمل لكل إختبار صورتين متكافئتين
من حيث عدد الأسئلة ونوعها وصعوباتها ، ثم تجرى كل منهما على نفس
التلاميذ بنفس التعليمات وفي ظروف مشابهة ، وتقارن النتائج بحيث لا يكون
بينهما وقت طويل . فإذا كان معامل الثبات عالياً يحصل الطالب على أعلى
الدرجات في كلا الصورتين . والطالب الذي يحصل على درجات قليلة يحصل
على ذلك أيضاً في الصورة الثانية . وبحساب معامل الارتباط بين التمتحنين
يتمحدد معامل الثبات .

٢ - إذا لم يستطع المدرس عمل صورتين للإختبار يصحح أن يجرى
الإختبار مرتين ، على أن يكون تكرار إجراء الإختبار بنفس الطريقة وفي
نفس الظروف وعلى نفس الطلبة - ويعطى وقت مناسب بين المرتين بحيث
لا يحدث تعلم أو تذكر لتفاصيل الإختبار وذلك لمنع عامل الذاكرة وبحيث
يكون مستوى التحصيل لا يزال كما هو . وتقارن النتائج .

٣ - الطريقة الثالثة هي مقارنة نصف الإختبار مثل أخذ الأسئلة الفردية
١ ، ٣ ، ٥ والأسئلة الزوجية ٢ ، ٤ ، ٦ . . . ونظرياً تعتبر هذه مقارنة بين
صورتين متكافئتين للإختبار حيث يفترض تساوى الأسئلة الفردية مع
الزوجية في درجة الصعوبة مثلاً - ولكن لأن هذه الطريقة تقارن نصف

الإختبار بنصفه الآخر فتشتمل معادلة النمويض طول الإختبار وهي : -

$$r = \frac{r^2}{r + 1} \text{ (سيرمان براون) .}$$

ومن الواضح أنه كلما كان الإختبار طويلا وعدد أسئلته أكثر كلما ساعد ذلك على ثبات نتائجه ، كما أنه كلما كان عامل التضمين مستتبدا كلما زاد معامل الثبات . وكذلك وضوح التعليمات والدقة في اجراء الأختبار يساعد على زيادة معامل الثبات .

٣ - يجب أن يكون الأختبار موضوعيا :

للموضوعية ناحيتان :

- (ا) ناحية تتعلق بفهم التلميذ لكل سؤال وتفسيره للمطروب .
- (ب) ناحية تتعلق بطريقة التصحيح وتقدير المدرس للاجابة . وكل من الناحيتين مهم جدا وله أثر كبير في تحسين صحة الإختبار وثباته .

(أولا) ناحية التصحيح :

فأما ناحية التصحيح فالواجب أن لا يختلف المصححون في التقدير بحيث إذا قام أكثر من شخص بتصحيح ورقة واحدة فإنهم يعطونها نفس الدرجة . ولهذا يجب أن يعد مفتاح الإجابات بحيث لا يدع مجالاً للخلافات على طريقة التصحيح .

ومن الواضح أن الامتحانات التقليدية من الفروع الذي يتطلب إجابة طويلة لا يتوافر فيها هذا الشرط وهو الموضوعية في التصحيح ، إذ تعطى مجالاً للخلافات في التقدير لعدم وجود قاعدة ثابتة للتصحيح . ويتوقف هذا على تقدير الصعوبة والسهولة ومدى التشدد أو التساهل في إعطاء الدرجة . ويمكن عمل تجربة

بسيطة بإعطاء ورقة اجابة واحدة لأكثر من مصصح ليقدرها الدرجة ونسأهم على الأساس الذي أعطوا عليه الدرجة وسنجد خلافات كثيرة لعدم وجود أساس موحد للتقدير . ولتدخل العوامل الذاتية .

(ثانيا) موضوعية فهم التلميذ للسؤال :

١ يجب أن يكون هناك تفسير واحد فقط يمكن للتلميذ الذي يعرف الإجابة الصحيحة أن يفهمه من السؤال ، ولا يصح أن يوضع السؤال بحيث يقرأ التلميذ فيه تأويلات غير مقصودة من السؤال . إذ أن هذا سيؤثر بطبيعة الحال على صحة الاختبار ، معنى ذلك أنه يجب عندما يسأل التلميذ نفسه ، ماذا يقصد واضع السؤال ؟ وماذا ينتظر أن تكون الإجابة ؟ ينبغي ألا يكون هناك لبس أو التواء في السؤال أو خداع أو صعوبة في فهم المطلوب .

وحتى إذا حاول المدرس المتحزن أن يحافظ على موضوعية السؤال ، فقد يكون من الضروري تجريب الاختبار على عدد من الطلاب ودراسة الإجابات وتحليل ما فهمه كل تلميذ من كل سؤال ، ومناقشة بعض الأسئلة التي تعذر فهم المقصود بها وتستحق التعديل .

٤ — يجب أن يكون الاختبار مميزا :

ينبغي للاختبار الجيد أن يكون صالحاً لقياس الفروق الدقيقة بين الأفراد ، وأن ينتقى من بينهم المتفوقين والضعاف ، وهذه قاعدة ضرورية أيضاً ، وتتدخل في ترتيب التلاميذ حسب درجاتهم ، وبالتالي تؤثر في ثبات الاختبار .

ولتحقيق هذا الشرط يجب أن يراعى ما يأتي : —

(١) يجب أن يكون هناك مدى واسع بين السهل والصعب بحيث نحصل

على توزيع معقول للدرجات بين أحسن درجة والدرجة الأقل من المتوسط .

(ب) يجب أن يكون هناك أسئلة في كل مستوى من مستويات الصعوبة بحيث نجد تلاميذ يأخذون درجات مختلفة ويكون التوزيع معقولا ومتمشيا مع تدرج الاختبار من السهل إلى الصعب في مسافات متساوية بقدر الامكان .

(ج) يجب أن يفرق كل سؤال بين التلميذ الضعيف والتلميذ القوي ، ويجب ألا يترك التلميذ القوي سؤالا من الأسئلة ، لأن معنى هذا أنه لا يصلح للتصنيف بينه وبين التلميذ الضعيف . وبالعكس لا يصلح أن يتمكن تلميذ ضعيف من إجابة أسئلة يعجز عن إجابتها تلاميذ أقوياء . إذ أن القياس سوف لا يكون في اتجاه إيجابي هنا . بل يكون قياسا سلبيا أي يقيس الضعف لا القوة . ومن أمثلة ذلك حالة السؤال الغامض الذي يحتاج فهمه إلى تفكير ودقة بحيث تهدف إجابته لغرض بعيد غير مقصود بينما يكون المقصود هو الفرض السطحي الذي يفهمه التلميذ الضعيف لأول وهلة . إذ أن مثل هذا السؤال لا يقيس الذكاء أو القوة وإنما يقيس الضعف والقابلية للاستهواء والسطحية في النظر إلى السؤال — أي الاتجاه السلبى .

ولا بد من تحليل النتائج لكل سؤال لاستخراج عدد الإجابات الصحيحة والخطئة والمتروكة ، وللحكم على صلاحية كل سؤال يؤخذ مثلا عشرون ورقة إجابة من حصلوا على أعلى الدرجات في الاختبار كله — ويؤخذ أيضا عشرون ورقة أخرى من حصلوا على درجات ضعيفة في الاختبار كله . وتفرغ النتائج من هذه الأوراق في جدول كالآتي يبين فيه لكل سؤال من أسئلة الاختبار كله عدد الإجابات الصحيحة والخطئة والمتروكة من واقع الأوراق ذات المستوى العالى . وعدد الإجابات الصحيحة والخطئة والمتروكة لكل سؤال أيضا من واقع الأوراق الضعيفة . ومن مقارنة الأرقام التي نصل إليها في هذا الجدول

يمكن الحكم بسهولة عما إذا كان السؤال مميزاً أم لا . فالاختبار المميز هو الذي
تتمشى نتائجه مع النتيجة العامة للاختبار كله ، وفيما يلي مثال توضيحي لذلك .

جدول لتسجيل الأسئلة لدراسة قوة كل سؤال على التمييز :

رقم السؤال	٢٠ ورقة نتائجها مجموع عالي		٢٠ ورقة نتائجها ضعيفة		
	الخطأ	المصححة	الخطأ	المصححة	
١	٤	١٦	٨	١٠	٢
٢	٠	٢٠	١٨	٠	٢
٣	٦	١٠	٥	١١	٤
٤	٠	١٨	٢	١٦	٢
٥	١٥	٥	٢	١٥	٢
٦	١٥	٣	١٨	١	١

ومن هذا الجدول يتضح أن :

سؤال ١ — مميز ولكن الإجابات الصحيحة لضعف التلاميذ تحتاج

للبحث .

سؤال ٢ — مميز تماماً .

سؤال ٣ — مشكل ويجب تغييره .

سؤال ٤ — معظم الإجابات صحيحة فما فائدة وجوده إذن ؟

سؤال ٥ — يجب تغيير السؤال . لأن نتائجه عكسية .

سؤال ٦ — غير مميز .

وإذن يمكن اتخاذ الأسس الآتية للحكم على صلاحية كل سؤال بعد

تحليل النتائج : —

- (ا) مستوى السهولة والصعوبة .
(ب) معامل الارتباط بين نتائج كل سؤال والاختبار كله .
(ج) صلاحية السؤال للتمييز بين الأقوياء والضعاف من المختبرين .

٥ — أن يكون الاختبار شاملاً :

أى يجب أن يكون عينة ممثلة لجميع نقاط المنهاج ، ليس من الضروري وجود كل نقطه ، ولكن يجب أن يكون شاملاً بحيث يحقق صلاحية وصحة الاختبار .

٦ — يجب أن يكون الاختبار سهل التطبيق والتصحيح :

يجب ألا يضيع الاختبار وقتاً طويلاً على المفحوصين ، وأن يكون تصحيحه سريعاً وميسوراً — وأن يكون مشوقاً بحيث يقبل المختبرون على الإجابة عليه باهتمام .

وبجانب ذلك ينبغي أن يكون له معايير منتظمة سواء كان ذلك بالالتجاء إلى الثوابت أو الأرباعيات أو غير ذلك بحيث يمكن معرفة مستوى النتائج للشخص المتوسط من كل عمر معين .