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 الذى يراد الوصول إليه من تطبيق هذا الاختبار ؟ ومن هم الأفراد المراد اختبارهم ؟ من أى سن ؟ ومن أى مستوى تعليمى ؟ وما هى الطريقة التى يراد بها تطبيق الاختبار ؟ هل ستلقى أسئلته شفهيا أم تحريريا ؟ وهل سيجرى بطريقة جمعية أم فردية ؟ وكيف سيعبر المفحوص عن إجاباته ؟ إلى غير ذلك من الاعتبارات التى تتحدد بها أغراض الاختبارات والنواحى المراد قياسها .

٢ — مادة الاختبار : وفى ضوء هذه الاعتبارات يمكن تحديد نوع المادة المناسبة التى سيتكون منها الاختبار ، والمجال الذى يصح أن يحصر فيه اختيار محتوياته وأسئلته فنأخذ فى حصر الموضوعات التى يراد اختبارها . . ونتبع ذلك بعمل تصنيف أو تقسيم لهذه الموضوعات وتفاصيل كل نقطة . ويحسن هنا عمل جدول يوضح أقسام الاختبار وموضوعات كل قسم أو استعمال نظام بطاقات التصنيف والترتيب فى تبويب أجزاء الاختبار .

٣ — وضع الأسئلة : يوضع لكل من النقط السابق تحديدها عدد من الأسئلة مع مراعاة التوازن بين النواحى التى يقيسها الاختبار ؛ ويراعى فى وضع الأسئلة اعتبارات كثيرة منها : —

(ا) الابتعاد عن الأسئلة الملتوية .

(ب) الابتعاد عن نقل عبارات معينة من الكتب فى 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 التافهة .

(هـ) أن تكون اللغة صحيحة وواضحة ومشوقة .

ويراعى فى تصميم الأسئلة محاوله صياغتها فى أصلح الصور التى تضمن الوصول إلى الغرض من القياس ، كأن يكون بعضها من النوع الذى تطلق فيه الحرية للإجابة ، و بعضها مما يتعين فيه اختيار إجابة واحدة من مجموعة اجابات معطاه ، و بعضها 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 فى ترتيب الأسئلة تنوع مادتها وأسلوبها حسب الصورة التى تناسب تصميم الاختبار .

٥ — وضع تعليمات الأسئلة : يوضع فى كل سؤال التعليمات المناسبة للإجابة من حيث مكان الإجابة وطريقتها . و إذا كان للاختبار أقسام متشابهة فيصح أن يبدأ كل قسم منها بمثال توضيحى يساعد على فهم التعليمات والمطلوب من الأسئلة .

٦ — تجهيز مفتاح الاختبار : يوضع نموذج للإجابة عن كل سؤال والدرجة التى يصح اعطاؤها عند تصحيحه . مع عمل التعديلات التى تظهر الحاجة إليها فى 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 فى توضيح مفهوم الأسئلة ودرجة صعوبتها وما تحتمله لغتها 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 : بقصد الوقوف على مدى صلاحيته 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 التى يحصل عليها متوسط الأفراد لكل عمر — ويصح أن تكون هذه المعايير Norms فى صورة جداول أو رسوم بيانية .

شروط الاختبار الجيد

الاختبار الجيد باختصار هو الذى يصلح لأداء الغرض الذى 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 » .

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

٤ — أن يكون الاختبار صالحا للتمييز بين الأفراد ذوى المستوى العالى والمستوى الضعيف ، وهذا هو ما يعبر عنه « التميز » .

٥ — أن يكون الاختبار شاملا بحيث يغطى 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 بشىء من التفصيل :

١ — صلاحية الاختبار :

الاختبار الصالح هو الذى يقيس ما هو مفروض أن يقيسه تماما ، وهذا ينطبق على السؤال الواحد ، فصلاحية سؤال ما فى الاختبار تتوقف على مدى قياسه للناحية المفروض أنه وضع لقياسها . فإذا كان الاختبار يقيس التحصيل فينبغى ألا تكون الأسئلة مشبعة 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 فى صحة الاختبار مثل : القدرة على القراءة ، نقص المحصول اللغوى ، ضعف الذاكرة ، بطء الرد والاستجابة . . . فإذا لم يكن الاختبار نقيا من هذه العوامل وأمثالها فإن

الدرجة المعطاة عن الاختبار لا تعبر عن المطلوب قياسه تماما ، ويكون الاختبار معيبا من ناحية صلاحيته .

وقد يكون الاختبار صالحا للقياس على مجموعة معينة من الأفراد ، ولكنه يعتبر غير صالح بالنسبة لمجموعة أفراد مختلفين عن المجموعة الأولى ، فقد يكون الاختبار فى الحساب مثلا صالحا لقياس القدرة الحسابية فى فصل معين يقوم بالتدريس له مدرس معين ، ولكنه لا يعتبر صالحا 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 تصلح صلاحية كاملة للقياس فى المدارس المختلفة ، ولهذا يفضل دائما أن تعمل كل مدرسة ، — بل كل مدرس — لفصوله اختباراته الخاصة ، 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 النواحى والأغراض بالضبط ، كما يحدث عند محاولة عمل مسطرة أو مقياس جديد فيصح التأكد من صحة صلاحيته للقياس بأن تقارن النتيجة بنتائج قياس مسطرة سبق تدريجها وسبق صلاحيتها (مع ملاحظة الفرق فى الدقة فى الحالتين طبعا) .

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 فى بداية عمل الاختبار وأن نلجأ إلى بعض الخبراء

فى هذه النواحى ليفتونا فى مدى صلاحية الاختبار لما هو مفروض أن يقيسه ، وعمل التعديلات المناسبة . ولكن هذه طريقة بدائية وتعتمد على العامل الذاتى .

أما صلاحية أسئلة الاختبار ذاتها سؤالا بسؤال فيمكن تحسينها بأن نسأل أنفسنا هل يصلح كل سؤال لقياس ما هو مفروض أن يقيسه .

٢ — ثبات الإختبار :

ثبات الإختبار دليل الوثوق بنتائجه ، واستمرار هذه النتائج — فى حالة اجرائه أكثر من مرة — بحيث لا يحدث تغير فى الدرجات التى يأخذها كل تلميذ إذا أجرى الإختبار مرتين على نفس التلاميذ . على شرط ألا يحدث تمرين أو تعلم بين المرتين طبعا .

ومن الواضح أن عامل ثبات الإختبار مرتبط بصحة الإختبار ، فإذا كان الاختبار ذا صلاحية عالية فلابد أن يكون ذا معامل ثبات مرتفع أيضا ، فإذا كان الإختبار يقيس فعلا ماهو مفروض أن يقيسه فستكون نتائجه ثابتة طبعا ، ولكن يصح أن يكون معامل الثبات عاليا ومع ذلك يكون الاختبار غير صادق فى صلاحيته وصحته .

ولكى يكون الإختبار ذا صحة جيدة يجب أن يكون معامل ثباته مرتفعا ، أى أنه يجب أن يكون الإختبار ذا معامل ثبات عالى لكى يكون صحيحا . ولكن قد يكون الإختبار ذا ثبات عال ولكنه لايكون ذا معامل صحة مرتفع إذ ليس من الضرورى أن يقيس ما يراد قياسه .

وهناك عوامل كثيرة تتدخل فى صحة الإختبار مثل القراءة والسرعة والفهم . . الخ . وهذه عوامل ثابتة . أما العوامل التى تؤثر فى ثبات الإختبار

فهى ظروف متغيرة فى كل مرة من مرات الإختبار وتجعل النتائج مختلفة مثل : حالة التلميذ النفسية ، زمن اجراء الإختبار ، طريقة الإجراء ، طريقة التصحيح وما قد يكون فيها من رأى شخصى ، الظروف المحيطة بإجراء الإختبار من حرارة وهدوء الخ .

طرق تعيين ثبات الإختبار :

١ — طريقة الصور المتكافئة : يعمل لكل اختبار صورتين متكافئتين من حيث عدد الأسئلة ونوعها وصعوباتها ، ثم تجرى كل منهما على نفس التلاميذ بنفس التعليمات وفى ظروف متشابهه ، وتقارن النتائج بحيث لا يكون بينهما وقت طويل . فإذا كان معامل الثبات عاليا يحصل الطالب على أعلى الدرجات فى كلا الصورتين . والطالب الذى يحصل على درجات قليلة يحصل على ذلك أيضا فى الصورة الثانية . وبحساب معامل الارتباط بين النتيجتين يتحدد معامل الثبات .

٢ — إذا لم يستطع المدرس عمل صورتين للاختبار يصح أن يجرى الإختبار مرتين ، على أن يكون تكرار إجراء الإختبار بنفس الطريقة وفى 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 هى مقارنة نصفى الإختبار مثل أخذ الأسئلة الفردية ١ ، ٣ ، ٥ والأسئلة الزوجية ٢ ، ٤ ، ٦ . . . ونظريا تعتبر هذه مقارنة بين صورتين متكافئتين للاختبار حيث يفترض تساوى الأسئلة الفردية مع الزوجية فى درجة الصعوبة مثلا — ولكن لأن هذه الطريقة تقارن نصف

الإختبار بنصفه الآخر فتستعمل معادلة لتعويض طول الإختبار وهى : —

$$\text{ص} = \frac{2\text{ر}}{1 + \text{ر}}$$ (سبيرمان براون) .

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

٣ — يجب أن يكون الأختبار موضوعيا :

للموضوعية ناحيتان :

(ا) ناحية تتصل بفهم التلميذ لكل سؤال وتفسيره المطلوب .

(ب) ناحية تتصل بطريقة التصحيح وتقدير المدرس للاجابة . وكل من الناحيتين مهم جدا وله أثر كبير فى تحسين صحة الإختبار وثباته .

(أولا) ناحية التصحيح :

فأما ناحية التصحيح فالواجب أن لا يختلف المصححون فى التقدير بحيث إذا قام أكثر من شخص بتصحيح ورقة واحدة فإنهم يعطونها نفس الدرجة .

ولهذا يجب أن يعد مفتاح الإجابات بحيث لا يدع مجالا للخلافات على طريقة التصحيح .

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

بسيطة بإعطاء ورقة إجابة واحدة لأكثر من مصحح ليقدروا الدرجة ونسألهم على الأساس الذى أعطوا عليه الدرجة وسنجد خلافات كثيرة لعدم وجود أساس موحد للتقدير . ولتدخل العوامل الذاتية .

(ثانيا) موضوعية فهم التلميذ للسؤال :

يجب أن يكون هناك تفسير واحد فقط يمكن للتلميذ الذى 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 ؟ ينبغى ألا يكون هناك لبس أو التواء فى السؤال أو خداع أو صعوبة فى فهم المطلوب .

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

٤ — يجب أن يكون الاختبار مميزا :

ينبغى للاختبار الجيد أن يكون صالحاً لقياس الفروق الدقيقة بين الأفراد ، وأن يفرق من بينهم المتفوقين والضعاف ، وهذه قاعدة ضرورية أيضاً ، وتتدخل فى ترتيب التلاميذ حسب درجاتهم ، وبالتالى تؤثر فى ثبات الاختبار .

ولتحقيق هذا الشرط يجب أن يراعى ما يأتى : —

(ا) 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 فى كل مستوى من مستويات الصعوبة بحيث نجد تلاميذ يأخذون درجات مختلفة ويكون التوزيع معقولا ومتمشيا مع تدرج الاختبار من السهل إلى الصعب فى مسافات متساوية بقدر الامكان .

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

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

يمكن الحكم بسهولة عما إذا كان السؤال مميزاً أم لا . فالاختبار المميز هو الذى تتمشى نتائجه مع النتيجة العامة للاختبار كله . وفيما يلى مثال توضيحى لذلك .

جدول لتحليل الأسئلة لدراسة قوة كل سؤال على التمييز :

رقم السؤال	٢٠ ورقة نتيجتها مجموع عالى			٢٠ ورقة نتائجها ضعيفة		
	عدد الإجابات الصحيحة	الخاطئة	المتروك	الصحيحة	الخاطئة	المتروك
١	١٦	٤	٠	١٠	٨	٢
٢	٢٠	٠	٠	٠	١٨	٢
٣	١٠	٦	٤	١١	٥	٤
٤	١٨	٠	٢	١٦	٢	٢
٥	٥	١٥	٠	١٥	٣	٢
٦	٣	١٥	٢	١	١٨	١

ومن هذا الجدول يتضح أن :

سؤال ١ — مميز ولكن الإجابات الصحيحة لضعاف التلاميذ تحتاج للبحث .

سؤال ٢ — مميز تماما .

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 .

(جـ) صلاحية السؤال للتمييز بين الأقوياء والضعاف من المختبرين .

٥ — أن يكون الاختبار شاملا :

أى يجب أن يكون عينة ممثلة لجميع نقط المنهاج ، ليس من الضرورى وجود كل نقطه ، ولكن يجب أن يكون شاملا بحيث يحقق صلاحية وصحة الاختبار .

٦ — يجب أن يكون الأختبار سهل التطبيق والتصحيح :

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

وبجانب ذلك ينبغى أن يكون له معايير منتظمة سواء كان ذلك بالالتجاء إلى المئويات أو الارباعيات أو غير ذلك بحيث يمكن معرفة مستوى النتائج الشخص المتوسط من كل عمر معين .