دروس وحدة: بناء الاختبارات

إعداد الأستاذ: عيمر موراد

السنة الثانية علم النفس

أعمال موجهة

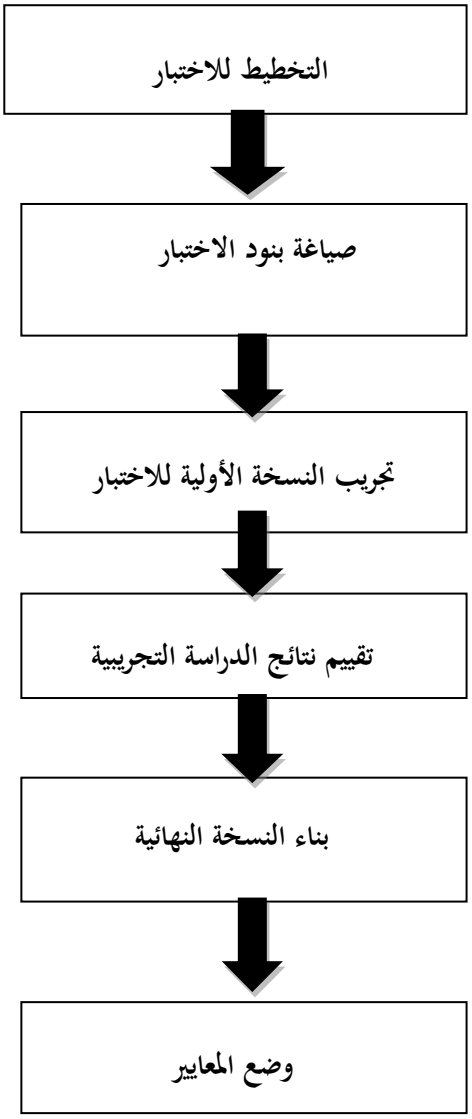
السنة الجامعية 2019-2020

خطوات بناء الاختبار

في بعض الأحيان قد يعجز الباحث عن إيجاد مقياس مناسب للسمة أو الخاصية المراد قياسها، أو الذي يتلاءم مع خصائص أفراد مجتمعه، ومن ثم يصبح لزاما عليه أن يقوم بتصميم مقياس جديد.

وتعتمد عملية تصميم المقاييس في المقام الأول على القيام بعدة خطوات متسلسلة تؤدي في النهاية إلى تجنب كثير من الأخطاء وتتيح إمكانية إعداد مقياس جيد. و هي خطوات متشابهة بالنسبة لكل الاختبارات، وإنما تختلف الاختبارات باختلاف أهدافها ومحتوى بنودها .

وتتمثل هذه الخطوات في:



مع الإشارة إلى ضرورة مراعاة القيام بها بنفس التسلسل.

أولا: التخطيط للاختبار

تعد خطوة من أهم الخطوات، حيث يسعى الباحث إلى تحديد الغرض من المقياس ومبررات تصميمه، خصائص مجتمع وعينة الدراسة، الشكل الذي سيكون عليه الاختبار... بشكل واضح و ذي دلالة.

حيث تسمح هذه المرحلة ب:

أ- تحديد الهدف من الاختبار بوضوح.

ب- تحديد الإطار النظري للسمة المعنية بالقياس.

ج- تقديم تعريف إجرائي للسمة المراد قياسها.

د- تحديد مجتمع الدراسة وحجم العينة وكيفية اختيارها في حالة تعذر الحصر الشامل.

هـ- اختيار شكل المقياس.

و- تقدير تكلفة بناء الاختبار.

ويستعين الباحث في هذه المرحلة ب:

-مراجعة الكتابات حول السمة موضوع القياس.

-حصر المقاييس المشابهة التي صممت لقياس الخاصية نفسها أو سمات مشابهة.

-الاستعانة بآراء الخبراء.

ثانيا: صياغة بنود الاختبار

يقوم الباحث بجمع وإعداد بنود المقياس. ويلجأ مصمم المقياس في هذه المرحلة إلى عدة أساليب للحصول على البنود اللازمة لبناء المقياس منها :

أ. مراجعة المقاييس المشابهة

ب . الاستبيان الاستطلاعي خاصة في المقاييس ذات النهاية المغلقة.

➢ عدد البنود

من الضروري أن يقترح المصمم بنودا أكثر بكثير مما يتوقع أن يحتويه الاختبار (ضعف العدد الذي يحتاجه الباحث)، حيث من المتوقع أن يتم الاستغناء عن عدد من البنود المقترحة أثناء عرض الاختبار على الخبراء وتجريبه وتحليل بنوده ودراسة خصائصه السيكوميترية.

➢ شكل البنود

قد تختلف طبيعة و شكل البنود و مقاييس الإجابة من اختبار إلى آخر، و يختار الباحث البنود التي تتماشى مع طبيعة السمة موضوع الدراسة، طبيعة مجتمع الدراسة، سهولة التطبيق...

فمثلا يغلب على اختبارات الذكاء بنود: 1_ التناظر (Analogies)؛ 2-إلغاء إحدى الإجابات (Odd-man-out items)؛ 3-بنود إكمال الجمل (Sequences). (Sequences). بينما تعتمد بنود اختبارات القدرات والتحصيل في الغالب على: 1-اختيار إجابة واحدة من إجابتين (true–false item) ؛ 2-اختيار إجابة واحدة من عدة إجابات.

وأما اختبارات الشخصية بالمعنى العام فيغلب فيها استعمال سلالم ليكرت التي تكون ثلاثية، خماسية،سباعية ... لسهولة تطبيقها وتصحيحها.

➢ بعض القواعد العامة لصياغة بنود الاختبارات

من القواعد العامة التي يجب أن يلتزم بها الباحث عند صياغة بنود اختباره:

-جعل الاختبار قصيرا قدر الإمكان مع الحرص على عدم التأثير على مصداقيته.

-استعمال لغة بسيطة و مباشرة، مع تقديم تعاريف إن اقتضت الضرورة.

-جعل البنود البسيطة والسهلة في بداية الاختبار.

- جعل كل بند خاص بفكرة واحدة فقط. فالبنود التي تشمل أكثر من واحدة قد تربك المفحوص و تجعله لايدري كيف يجيب.

-تجنب استعمال عبارات ك: عادة ما، كثيرا ما... وتعويضها بأخرى أكثر دقة.

-تحديد الزمن إن أمكن، فنكتب سنة 1990 مثلا عوض نهايات القرن الماضي .....

-التأكد من أن العلامات في سلم الإجابات لا تتداخل.

-تجنب صياغة البنود بطريقة توجه إجابة المفحوص أو توحي إليه بالإجابة المفضلة أو المنتظرة.

و على العموم، يرى kent 1993 أن الحصول على إجابات صادقة و مفيدة يتوقف على توفر ثلاثة شروط:

- يجب أن يفهم المفحوص بنود الاختبار فهما مناسبا ومساويا للمفحوصين الآخرين.
- يجب أن يكون المفحوص قادرا على تقديم الإجابة.
- يجب أن يكون المفحوص راغبا في تقديم الإجابة.

➢ عرض المقياس على المتخصصين في المجال

هي خطوة يلجأ إليها الباحث قبل إخضاع نسخته التجريبية للدراسة الميدانية. حيث يستعين بمجموعة من الزملاء الأخصائيين لمراجعة البنود، ذلك لأنه في كثير من الأحيان قد تبدو بعض البنود واضحة بالنسبة للباحث مصمم الاختبار، في حين أنها أقل وضوحا بالنسبة للآخرين.

ويعد عرض المقياس على المتخصصين والخبراء في المجال خطوة هامة تحقق عديد من الفوائد من أهمها التأكد من مدى مناسبة البنود وقدرتها على قياس الخاصية أو السمة طبقا للتعريف الإجرائي والهدف من المقياس والإطار النظري الخاص بالسمة أو الخاصية موضوع القياس والفئة المستهدفة.

- **وضع التعليمات، المقدمة، الخاتمة، وعنوان الاختبار**

تنقسم تعليمات المقياس إلى قسمين رئيسيين هما:

أ-تعليمات المطبِّقين: وهم الذين يقومون بتطبيق المقياس، وتتضمن شرحا وافيا للمقياس والخاصية التي يتم قياسها، وإجراءات التطبيق بالتفصيل، والزمن، وطريقة تسجيل الاستجابات وتصحيحها، والمواقف التي يحتمل مواجهتها أثناء التطبيق، وحدود الشرح والتوضيح المسموح به للمفحوصين.

ب-تعليمات المفحوصين: وتتضمن عدة محاور منها:

1-فكرة مبسطة عن المقياس والهدف من وراء تطبيقه.

2-طريقة الاستجابة والزمن المحدد إن وجد.

3-تقديم بعض النماذج المحلولة إن تتطلب الأمر.

وفي النهاية ينتهي الباحث بصياغة مقدمة الاختبار، الخاتمة، وعنوان الاختبار يشرح من خلالها الهدف من الدراسة، مصادر التمويل إن وجدت، أهمية إجابات المفحوصين، كيفية استغلال المعلومات المقدمة و سرية التعامل معها... ما يشجع المفحوص و يرغبه في المساهمة في المشروع.

ثالثا: تجريب النسخة الأولية للاختبار

يتم تجريب الاختبار على عينة ممثلة للمجتمع الأصلي (أنظر الدرس حول اختيار العينة) والذي سوف يجري عليه الاختبار بعد ذلك. وقد يقوم المصمم إلى تقسيم هذه المرحلة إلى قسمين لأسباب عملية. قد تسمى الأولى التجربة الاستطلاعية (**Preliminary Tryout**) والأخرى الأساسية (**Actual Tryout**).

- **المرحلة الأولى**

وفيها يقدم الباحث الاختبار لعينة قليلة العدد (15-20 فردا) ولكنها ممثلة للمجتمع الأصلي.وإذا كانت هذه المرحلة تسمح بتصحيح كثير من الأخطاء، إلا أنها لا تهدف إلى التأكد من الخصائص السيكوميترية للاختبار. بل يسعى المصمم من خلالها إلى:

1-التأكد من وضوح التعليمات للمفحوصين.

2- تقدير الزمن الذي يستغرقه المقياس.

3-الاستقرار على الترتيب الأمثل للفقرات.

4-التأكد من وضوح البنود بالنسبة للمفحوصين (لم يتم تجاهل بند من طرف كل المفحوصين مثلا) ذلك أن فهمهم للبنود قد يختلف عن فهم المصمم والمحكمين.

- **المرحلة الثانية**

هي المرحلة التي يتم فيها تحليل بنود الاختبار والتأكد من صدقه وثباته وسهولة بنوده وقدرتها على التمييز. وهي مرحلة مهمة جدا يستعمل فيها المصمم عينة كبيرة (400 مفحوص حسب guilford 1954 أو 10 إلى 15 بند بالنسبة لباحثين آخرين) ممثلة للمجتمع الذي صمم له الاختبار.

رابعا: تقييم نتائج الدراسة التجريبية

تسمح البيانات التي تم جمعها من خلال تطبيق الاختبار على عينة ممثلة للمجتمع الذي صمم الاختبار من أجله بتحليل بنود الاختبار ودراسة صدقه وثباته.
لكن قبل ذلك، قد يحتاج الباحث إلى تخليص نتائج الاختبار من أثر التخمين.

➢ تخليص نتائج الاختبار من أثر التخمين

قد تعتمد بنود الاختبار المصمم على اختيار إجابة واحدة من عدة إجابات، وهي من أكثر الأنواع استخداماً في اختبارات القدرات والتحصيل، حيث تقدم للمفحوص مجموعة من الإجابات تتضمن الإجابة الصحيحة وعلى المفحوص أن يختار إحداها. ويتأثر هذا النوع من البنود بالتخمين. ولذلك يجب أن تصحح الدرجات إحصائيا من خلال المعادلة التالية مثلا.

$$\text{ن ص} = \text{ص} - \frac{\text{خ}}{\text{ب} - 1}$$

حيث:
ن ص= النسبة المصححة
ص= الإجابات الصحيحة
خ= الإجابات الخاطئة
ب= عدد البدائل

➢ تحليل البنود Item Analysis

يركز الباحث في هذه المرحلة على حساب معاملي الصعوبة والتمييز(فريمان Freeman1962). ففعالية وصلاحية أي اختبار تتوقف على نوعية بنوده. فصدق البنود، ثباتها وارتباطها فيما بينها هو ما يحدد جودة المقياس.

فالباحث يسعى من خلال هذه الخطوة إذن إلى التأكد من مدى ملاءمة بنود الاختبار من حيث درجة صعوبتها، وكذا قدرتها على التمييز بين المفحوصين. حيث يهدف إلى تحسين صدق وثبات الاختبار الكلي.

- **معامل الصعوبة item difficulty**

هو أحد مؤشرات جودة البنود، حيث يسمح حساب معامل الصعوبة بـ:
-التحقق من مدى ملاءمة البند. حيث يحتفظ بالبنود بين 20% و90 %، وتكون البنود بين 25% و75% مقبولة، وتكون مناسبة جدا كلما اقتربنا من 50%.
-وضع البند في المكان المناسب وضمان التدرج من حيث الصعوبة.
ويتم حساب معامل الصعوبة بعدة طرق، لعل أبسطها المعادلة التالية:

$$م\ ص = \frac{ص}{ن} \times 100$$

حيث:
م ص = معامل الصعوبة
ص = عدد الناجحين
ن = عدد أفراد العينة

- **معامل التمييز Item Discrimination**:

تعتبر قدرة البند على التمييز دليلا على صدقه. وصدق الاختبار الكلي إنما يتوقف على صدق كل بنوده وقدرتها على التمييز. وهناك طرق كثيرة لحساب معامل التمييز نذكر منها المعادلة التالية:

$$م\ ت = \frac{ع - د}{ن}$$

حيث:
م ت: معامل التمييز
ع: هي الإجابات الصحيحة في المجموعة العليا
د: هي الإجابات الصحيحة في المجموعة الدنيا
ن : عدد أفراد إحدى المجموعتين

ويتم تحديد المجموعة العليا والدنيا بإتباع الطريقة التالية:

1- ترتيب المفحوصين تصاعديا حسب الدرجات على الاختبار الكلي.

2- يقسم المفحوصون إلى مجموعتين: عليا تمثل 27% من الذين تحصلوا على الدرجات العليا ، ودنيا تمثل 27% من الذين تحصلوا على الدرجات الدنيا . وبذلك يكون عدد أفراد المجموعة العليا = عدد أفراد المجموعة الدنيا= 27% (مع إمكانية استعمال نسب أخرى 50% في كل مجموعة مثلا).

3- حساب عدد الإجابات الصحيحة عند المجموعة العليا وعدد الإجابات الصحيحة عند المجموعة الدنيا بالنسبة لكل بند.

- **تحليل البدائل**

يرتبط صدق الاختبارات متعددة الإجابات بنوعية البدائل المقدمة. وبالمقابل، فان حساب معملا الصعوبة والتمييز لا يسمح بالتأكد من فعاليتها وقدرتها على التشويش. وبالتالي يسعى المصمم في هذه الحالة إلى تحليل البدائل للتأكد من أنها:

- غير صحيحة في أي حال من الأحوال
- واقعية حيث لا يسهل تجاهلها

ويكون ذلك من خلال التأكد من أن:

- كل البدائل تم الإجابة عليها من طرف بعض المشاركين على الأقل، إذ تعتبر الإجابات الخاطئة التي لم يتم اختيارها تماما سهلة أو غير واقعية.
- التأكد من أن عدد الأفراد الذين أجابوا على البدائل عوض الإجابة الصحيحة أكبر عند المجموعة الدنيا مقارنة بالمجموعة العليا.

✓ **حساب معاملا الصدق والثبات**

- **طرق حساب ثبات الاختبار**

يرتبط مفهوم الثبات بما يسمى إحصائياً بخطأ القياس المتضمن في كل درجة من درجات الاختبار. فإذا حصل نفس الفرد على نفس الدرجة أو ما يقرب منها في نفس الاختبار في مناسبات مختلفة أو في مجموعات من الأسئلة المتكافئة أو المتماثلة فإننا نصف الاختبار أو المقياس في هذه الحالة بأنه على درجة عالية من الثبات. ويحسب معامل الثبات بطرق عديدة. نذكر منها:

_طريقة إعادة تطبيق الاختبار Test-retest method

-تطبيق الاختبار على نفس لعينة مرتين.

-يحسب معامل الارتباط بيرسون بين نتائج التطبيقين، فتدل الدرجة العالية على ثبات عال والعكس صحيح.

-يسمى كذلك معامل الاستقرار أي مدى استقرار الدرجات عبر الزمن.

من أبرز عيوبه: -التأثر بعامل التعلم والنضج، -التكلفة في الجهد والوقت، -صعوبة ضبط الظروف في التطبيقين.

_طريقة الصورتان المتكافئتان Equivalent or Parallel Forms Method

– لتجاوز صعوبات إعادة الاختبار، يصمم الباحث نسختين متشابهتين (متكافئتان).

–يحسب معامل الارتباط بيرسون بين نتائج التطبيقين، فتدل الدرجة العالية على ثبات عال والعكس صحيح.

من أبرز عيوبه: –صعوبة إيجاد التكافؤ، –الجهد، الوقت والتكاليف المترتبة عن بناء نسختين للاختبار.

_طريقة التجزئة النصفية Split – Half Method

يطبق الاختبار على عينة واحدة مرة واحدة فقط ثم تقسم درجات العينة إلى نصفين.

–عادة ما يقسم الباحث الاختبار بجعل الأعداد الفردية في مجموعة والزوجية في مجموعة أخرى حتى يتفادى تأثير درجة صعوبة وسهولة البنود وتأثير التعب والملل على ثبات الاختبار.

–تجدر الملاحظة إلى أن الارتباط المحسوب هو بين نصفي طول الاختبار. وعليه يتعين على الباحث إجراء تصحيح حتى نحصل على معامل ثبات الاختبار ككل. وعادة ما تستعمل معادلةسبيرمان – بروان التالية لتصحيح قيمة معامل الارتباط:

$$\text{الثبات} = \frac{r \times 2}{r + 1}$$

حيث:

r = قيمة الارتباط بين نصفي الاختبار

_طريقة التناسق الداخلي Internal Consistency

من أكثر المعادلات استخداماً لقياس الاتساق الداخلي هي معادلة ألفا –كرونباخ Alpha Cronbach. وتستخدم هذه المعادلة في المقاييس والاختبارات متعددة الاختيارات وليست الثنائية. حيث تستعمل في هذه الأخيرة معادلة كيودر وريتشارد سون KR20 Kuder–Richardson

- **العوامل المؤثرة على ثبات الاختبار**. منها:

– وضوح بنود الاختبار وتعليماته وطريقة التطبيق.

–عوامل شخصية كالإجهاد والتعب والملل والتوتر والانفعال والتذكر وغيرها.

– الوقت الفاصل بين الاختبارين.

-أثر النضج والتدريب والتعلم.

- طول الاختبار. بمعنى انه إذا زاد عدد فقرات الاختبار زاد معامل الثبات

ويمكن حساب تأثير الطول باستعمال معادلة سبيرمان – بروان التالية:

$$r_n = \frac{Kr}{r \times (1 _ K) + 1}$$

حيث:

r= قيمة ثبات النسخة الأصلية للاختبار

$\frac{K}{}$ = عدد المرات التي يضاعف أو يخفض بها عدد بنود النسخة الأصلية للاختبار.

- **طرق حساب صدق الاختبار**

✓ يقصد بالصدق مدى قدرة المقياس أو الاختبار على قياس السمة التي أعد لقياسها أو بمعنى أخر هل يقيس الاختبار أو المقياس فعلا ما أعد لقياسه؟.

✓ يتأثر صدق الاختبار بأخطاء القياس المنتظمة والعشوائية، عكس الثبات الذي يتأثر بالعشوائية فقط، وبالتالي يمكن أن يكون الاختبار ثابتا وغير صادق، ولكن العكس غير صحيح.

✓ تتعدد أساليب التأكد من صدق المقياس، جعلتها الجمعية الأمريكية لعلم النفس 1966 في 3 أصناف (مقدم عبد الحفيظ 1993): صدق المحتوى، صدق المحك، وصدق التكويني (البناء).

_الصدق الظاهري أو السطحي Face Validity: يعتبر من اقل الطرق أهمية واستخداما، وغالباً ما يحدده المتخصصين. فهو يمثل المظهر الخارجي للاختبار . ويقوم على فكرة مدى مناسبة البنود لما يقيسه الاختبار ولمن يُطبق عليهم. ويبدو هذا في وضوح البنود ومدى علاقتها بالقدرة أو السمة التي يقيسها الاختبار.

_صدق المحتوى Content Validity: وهو يقوم كذلك على تحديد المختصين لمدى تمثيل بنود الاختبار للمحاور المختلفة للقدرة التي يقيسها والتوازن بين هذه المحاور.

_صدق المحك Criterion–Related Validity: هو مدى ارتباط الاستبيان بمحك خارجي ثبت صدقه في قياس السمة التي يقيسها الاستبيان.

ويمكن التمييز بين الصدق التلازمي الذي يعتمد على إيجاد معامل الارتباط بين الاختبار الجديد واختبار آخر سبق إثبات صدقه، والصدق التنبئي الذي يعتمد على فكرة أن الاختبار الصادق هو الذي لديه القدرة على التنبؤ بالنتائج في مواقف مستقبلية. فنتائج اختبار صادق لقياس سمة معينة (الذكاء مثلا) في بداية السنة الجامعية ترتبط إيجابا بأداء وسلوكيات الطلبة في اختبار سمة أخرى أثبتت الدراسات ارتباطها بالسمة الأولى (التحصيل الدراسي مثلا) في نهاية السنة الجامعية.

_الصدق الذاتي Intrinsic Validity: ويمثل العلاقة بين الصدق والثبات. ويحسب بإيجاد الجذر التربيعي لثبات الاختبار.

_صدق البناء Construct Validity: ويقوم على حساب عدد معاملات، كمعامل الاتساق الداخلي (الارتباطات بين كل بند والمحور الذي ينتمي إليه، بحيث يحتفظ فقط بالبنود الموجبة والدالة)، الصدق العاملي (حيث يحتفظ فقط بالبنود التي تكون درجة تشبعها مقبولة).

خامسا: بناء النسخة النهائية

تضم النسخة النهائية البنود التي تم الاحتفاظ بها بعد التأكد من مستوى صعوبتها وقدرتها على التمييز وصدقها وثباتها.
كما يتم تحديد التعليمات بدقة، سلم الإجابة وطريقة التصحيح.
وفي الأخير، قد يحتاج الباحث بعد الانتهاء من بناء الاختبار إلى وضع معايير تسمح بمقارنة نتائج فرد أو أفراد في اختبار ما بنتائج مجموعة كبيرة تقاسمهم نفس الخصائص. وهي خطوة مهمة في بناء الاختبار وإعداده للاستخدام والتطبيق ،وهي تؤهل الاختبار للاستخدام في عينات ومجموعات أخرى غير تلك العينة أو المجموعة التي استخدم فيها أول مرة .