المحاضرة الثالثة: الخصائص السيكو- مترية للإختبارات النفسية

حتى تصبح الإختبارات النفسية قابلة للتطبيق على العينات التي بنيت لأجلها والنتائج التي تقدمها ذات قيمة قياسية وتشخيصية، عليها أن تتسم ببعض الخصائص الإحصائية والسيكو- مترية التي سنتطرق لأهمها في هذه المحاضرة.

1. عدم التحيز (الموضوعية):

بما أنه لا يوجد أي إختبار خال من العوامل الحضارية؛ فعلى مؤلف الإختبار أو من يستخدمه أن يطبق الإختبار على مجموعات مختلفة من الأفراد الذين سيطبق عليهم هذا الإختبار فيما بعد.

2. صدق الإختبار la crédibilité:

يكون الإختبار صادقا عندما تقيس بنوده ما وضعت لقياسه، فليس من الضروري أن يحدد الباحث النفساني بدقة في بداية بحثه، السمة التي يأمل في أن يقيسها هذا الإختبار، إذ يكفي أن تكون له فكرة عامة عن هذه السمة وعلاقتها بالمفاهيم النظرية أو المواقف العملية، لأن تحديد معنى السمة سوف يزيد كلما جُرِب هذا الإختبار في سلسلة من البحوث المنفصلة:

- صدق الإختبار ليس خاصية مطلقة بمعنى أن الإختبار إما أن يكون صادقا وإما أن يكون غير ذلك.

- الصدق خاصية نسبية، فالإختبار يصدق في قياس ما يقيسه بقدر معين.

- قد يختلف صدق الإختبار من مجتمع لآخر (الفروق الحضارية) ومن فترة زمنية لأخرى.

- يجب مراجعة إجراءات الصدق بين فترة وأخرى لإعادة تقدير تأثير المتغيرات الزمنية التي يمكن أن تؤثر بشكل مباشر في المفاهيم التي تتضمنها بنود الإختبار.

أنواع الصدق:

- **الصدق الظاهري (السطحي) Face validity** : لا يعتبر هذا الصدق صدقا بالمعنى الدقيق للكلمة ولا يشير إلى ما يقيسه الإختبار فعلا، بل إلى ما يبدو ظاهريا أنه يقيسه سواء في نظر المفحوصين أنفسهم أو في نظر غيرهم من الأشخاص غير المختصين، ومع ذلك تصر **أنستازي Anastasi** على ضرورة توفير صفة الصدق الظاهري للإختبار حتى يكون أكثر فاعلية في المواقف العملية ولضمان تعاون المفحوصين في الموقف الإختباري. ويمكن دراسة درجة الصدق الظاهري لاختبار ما من خلال تقييم التوافق بين تقديرات المحكمين، فبحسب توافق هذه التقديرات يكون مؤشر ضعف أو صدق الإختبار وتقييمهم لما إذا كانت هذه الفقرات تتعلق بالشيء الذي نريد قياسه، ثم يقوم الفاحص بعد ذلك بحساب تكرارات استجابات هذه المجموعة من المحكمين ويختار البنود التي اتفق عليها أكثر عدد منهم.
- **صدق المحتوى (الصدق المنطقي):** يتوقف صدق المحتوى على مدى قياس أداة معينة لمجال محتوى معين، فعلى الباحث التأكد من أن فقرات هذا الإختبار ومضمونها تقيس ما يجب عليها أن تقيسه. ففي النهاية، يقيس كل اختبار عينة من السلوك، فمجموعة البنود التي يضمها يفترض أنها عينة من المجتمع الأصلي الأوسع من البنود التي كان يمكن أن يضمها، ولذلك يتوجب على باني الإختبار ضمان تمثيل السمة أو ميدان السلوك المقاس بصورة أمينة وصحيحة.

– يتطلب هذا النوع من الصدق وصفا دقيقا للمحتوى أو المجال المقاس وتحليله إلى عناصره ومكوناته حتى يتم التأكد من أن البنود الموضوعة تغطي الجوانب الأساسية لهذا المحتوى وبالنسب الصحيحة.

– ليس ثمة طريقة كمية يمكن استعمالها للتحقق من صدق المحتوى أو تعبير عددي يمكن استعماله للدلالة عليه، بل يعتمد على التحليل المنطقي والمقارنة أو المطابقة بين محتوى الإختبار ومحتوى الموضوع المقاس لذلك يسمى عادة بالصدق المنطقي.

- **الصدق المرتبط بالمحك:** ترتكز معرفته على مقارنة نتائج الإختبار مثلا بمحكات des critères ومعرفة حجم الترابط بين هذا الإختبار والإختبار المعتمد كمحك، مثال: مستوى أداء الأفراد في نشاطات أخرى مثل التحصيل الجامعي أو الأداء على اختبار آخر.

العوامل التي تؤثر في صدق الإختبار:

– طول الإختبار: يزداد صدق الإختبار بزيادة مكوناته سواء كانت عبارات أو بنودا أو أسئلة.

– ثبات الإختبار: يتأثر الصدق بقيمة الثبات.

– ثبات المحك: يزداد الصدق تبعا لزيادة ثبات المحك ويتأثر بالقيمة العددية للمحك.

– التباين: يتأثر الصدق بتباين درجات الإختبار، فزيادة أو نقصان الفروق الفردية تؤثر على الصدق.

ولزيادة صدق الإختبار المراد بناءه يجب على المختص:

– المعرفة الجيدة بالموضوع أو التخصص الذي يكتب فيه.

– أن تكون بنود الإختبار في علاقة بالصفة المراد قياسها وألا تقيس جوانب أخرى لا علاقة لها بهذه الصفة.

– يجب أن تركز البنود على الجوانب المهمة من الصفة دون السطحي والهامشي منها.

– أن يكون الإختبار ملما بكافة أبعاد الصفة ومجالاتها.

3. ثبات الإختبار la stabilité:

يقصد بالثبات مدى تشابه المعطيات المتحصل عليها من تطبيق الإختبار عبر الزمن، أي أنه لو أعيد تطبيق الإختبار نفسه على الفرد الواحد في فترة زمنية لاحقة فإنه يعطي شيئا من الإستقرار في النتائج.

لا يكون الإختبار الثابت مستقرا بالضرورة، ويُعرَف الثبات إحصائيا بحساب معامل الإرتباط بين الدرجات التي حصل عليها الأفراد في المرة الأولى وبين نتائج الإختبار في المرة الثانية.

❖ **طرق حساب معامل ثبات الإختبار:**

أ. طريقة الصور المتكافئة: ويقدم فيها نفس الإختبار بعد فترة ولكن ببنود مختلفة، حيث يكون الإختلاف دائما في صياغة هذه الأخيرة وليس في محتواها.

ب. طريقة إعادة الإختبار: التي يعاد فيها تطبيق الإختبار مرة أخرى على أفراد المجموعة نفسها ولكن بعد فترة زمنية ملائمة (لا يجب أن تقل عن أسبوع ولا تزيد عن ستة أشهر)، ثم نقوم بحساب معامل الإرتباط بين الدرجات التي حصل عليها أفراد العينة في المرة الأولى والثانية.

ت. طريقة التجزئة النصفية: تتطلب تصميم اختبار بقسمين متكافئين، بحيث يحتوي القسم الأول على الدرجات الفردية 1، 3، 5، 7 والقسم الثاني على الدرجات الزوجية 2، 4، 6، 8. ثم نستخدم معاملات الإرتباط بين درجات كل فرد على البنود الفردية والزوجية ونستخدم بعد ذلك معادلة سبيرمان–بروان Spearman Brown.

❖ **العوامل المؤثرة في ثبات إختبار:**

– طول الإختبار: كلما زاد عدد بنود الإختبار كلما أدى ذلك إلى ارتفاع معامل الثبات.

– زمن الإختبار: يتأثر ثبات الإختبارات الموقوتة بالزمن المحدد لها.

– غموض الأسئلة: تقلل البنود الغامضة والطويلة من الثبات وتزيد تلك الواضحة والموضوعية منه.

– صعوبة الفقرات ودقتها: إن وجود فقرات صعبة في الإختبار لا يستطيع أغلب الأفراد أو جميعهم الإجابة عنها تؤثر في ثبات الإختبار، كما أن وجود فقرات سهلة في الإختبار يستطيع الجميع الإجابة عنها له نفس الأثر على ثبات الإختبار، لذلك يجب حذف الأسئلة الصعبة والسهلة.

– حالة الفرد: حالته النفسية والصحية والتعب والتوتر ومدى تدربه على الموقف الإختباري.

– ظروف إجراء الإختبار: يؤثر تغيير الظروف الخاصة بإجراء القياس من إختبار لآخر على نتائج المقياس.

– موضوعية الإختبار: موضوعية تصحيح الإختبار من بين العوامل المؤثرة في ثبات الإختبار.

– التخمين: قد يلجأ بعض المفحوصين في حالة عدم تأكدهم من الإجابة الصحيحة إلى التخمين مما يؤدي إلى خفض ثبات الإختبار.

– تباين العينة وتجانسها: ينقص ثبات الإختبار بنقصان التباين ويزداد تبعا لزيادته، فالسهولة والصعوبة تؤديان إلى التجانس وبالتالي إلى خفض الثبات، أما الأسئلة المتدرجة في الصعوبة فتؤدي إلى رفع ثبات الإختبار.

4. حساسية الإختبار (تمييز الإختبار) la sensibilité:

يقصد بحساسية الإختبار مدى قدرته على اكتشاف التغيرات (أو الإختلافات) ما بين الأفراد أو لدى نفس الفرد في قدرة أو سمة معينة، ويستمد الإختبار قدرته التمييزية من القدرة التمييزية لكل واحد من بنوده.

تشير القدرة التمييزية للبند الإختباري إلى قدرته على التفريق أو التمييز بين المفحوصين الذين يظهرون أداءا حسنا وبين أولئك الذين يكون أداؤهم ضعيفا في الإختبار ككل. ويكون البند مميزا جيدا إذا ما ترابط الأداء على هذا البند ترابطا موجبا وعاليا مع الأداء في الإختبار ككل، فالبنود التي لا تميز بين المفحوصين أو تترابط ترابطا سلبيا مع الأداء الإختباري ككل تقلل التباين في توزيع درجات الإختبار الكلي وتضعف بالتالي مستوى الصدق والثبات.

يمكن استخدام معامل الإرتباط بيرسون Pearson الذي يستعمل لقياس درجة الترابط بين أداء الطالب على البند وأداءه على الإختبار ككل.

يجب الإبقاء على الفقرات أو البنود ذات التمييز العالي فقط، وهذا لا يعني أن الأفراد دوي المستوى الضعيف لا يجيبون على تلك البنود أو الفقرات بل يجب أن تكون نسبة المجيبين عليها من دوي المستوى العالي أعلى من نسبة ذوي المستوى الضعيف بصورة واضحة. كما أن البنود أو الفقرات التي لا يجيب عنها جميع المفحوصين مهما كانت مستوياتهم لا فائدة منها لأنها لا تستطيع التمييز بينهم.

يمكننا قياس معامل تمييز الإختبار أو بند منه من خلال:

- تطبيق الإختبار على عينة من الأفراد والحصول على الدرجة التي تحصل عليها كل فرد من هذه العينة.
- ترتيب الدرجات المتحصل عليها من طرف أفراد العينة من أعلاها إلى أدناها.
- تحديد مجموعة الدرجات العالية بأخذ ما نسبته 27% من مجموعها الأصلي، وكذلك الأمر بالنسبة للدرجات المنخفضة، بحيث تشكل المجموعتين ما مجموعه 54% من المجموع الكلي للمُختبَرين. فيما تمثل النسبة المتبقية (46%) الدرجات المتوسطة للمختبَرين.

عندما يتعلق الأمر **بالإختبارات البدنية والمهارية** يمكننا مقارنة المجموعتين عن طريق المتوسطات والإنحرافات المعيارية للإختبارات باستعمال إختبار (ت) ومقارنة القيمة المحسوبة مع القيمة المجدولة، فإذا كانت القيمة المحسوبة أكبر من القيمة المجدولة نتحدث عن وجود فروق معنوية تدل على قدرة الإختبار التمييزية والتفريقية.

أما إذا تعلق الأمر بالإختبارات النفسية والمعرفية فيمكن استخراج عدد الأفراد الذين أجابوا إجابة صحيحة على الفقرة في كل من المجموعتين، ثم نقوم بحساب النسبة المئوية لدرجة تمييز الفقرة وفق المعادلة التالية:

عدد الإجابات الصحيحة للفئة العليا – عدد الإجابات الصحيحة للفئة الدنيا ÷ عدد أفراد إحدى الفئتين × 100

- الفقرة التي تحصل على نسبة مئوية سالبة (−30%) هي فقرة **غير مُميزة**.
- الفقرة التي تحصل على نسبة مئوية موجبة (أقل من 40%) هي فقرة ذات **قدرة تمييزية ضعيفة**.
- إذا تراوحت النسبة بين (40 – 60%) هي فقرة ذات **قدرة تمييزية متوسطة**.
- الفقرة التي تزيد نسبتها عن 60% هي **فقرة جيدة التمييز**.
- كلما اقتربت نسبة الفقرة من 100% كلما كانت **قدرة الفقرة على التمييز ممتازة**.

كما يمكن أن نقيس معامل التمييز وفق الطريقة التالية:

عدد الإجابات الصحيحة عن الفقرة في المجموعة العليا – عدد الإجابات الصحيحة عن الفقرة في المجموعة الدنيا ÷ عدد أفراد إحدى المجموعتين

وقد وضع **إيبل Ebel** (1979) مجموعة القواعد التالية لتحديد إمكانية قبول أو رفض الفقرة في ضوء معامل تمييزها:

- إذا كان معامل التمييز أكبر من 0.40 فالفقرة ذات تمييز عال وممتاز.
- إذا كان معامل التمييز بين 0.30 – 0.39 فالفقرة ذات تمييز جيد.
- إذا كان معامل التمييز بين 0.20 – 0.29 فالفقرة ذات تمييز جيد إلى حد ما(فقرات جيدة تحتاج إلى تحسين).
- إذا كان معامل التمييز أقل من 0.19 فالفقرة ضعيفة وينصح بحذفها.

5. صعوبة وسهولة بنود إختبار:

يمكن لوجود بنود صعبة أو سهلة في الإختبار أن يؤثر على ثباته، كما أن البنود شديدة السهولة أو شديدة الصعوبة لا تميز بين المفحوصين ولذلك يتحتم علينا حذفها، قبل ذلك علينا تحديد تلك البنود الصعبة والسهلة عن طريق حساب معاملي السهولة والصعوبة:

أ. معامل سهولة البند:

يعتبر حساب معامل السهولة لكل بند من بنود الإختبار الخطوة الأولى في عملية تحليل بنود الإختبار الجديد وهو يتطلب:

- إعداد جدول بالإجابات الخاصة بكل بند في الإختبار.
- حساب معامل سهولة البند الذي يكون بـ: **تقسيم** عدد الذين أجابوا إجابة صحيحة عن السؤال **على** عدد من حاول الإجابة عليه من المفحوصين × 100.

كما يمكن حساب معامل السهولة كما يلي:

عدد الإجابات الصحيحة للبند ÷ مجموع الإجابات الصحيحة والخاطئة

وهكذا:

- تعتبر أفضل البنود أو الأسئلة من حيث القدرة التمييزية هي البنود التي يصل معامل سهولتها إلى 0.50.
- يفضل الإبقاء على البنود التي تتراوح معاملات سهولتها ما بين 0.40 و0.60.
- يمكن الإحتفاظ ببعض البنود السهلة ليبدأ بها المفحوصون وذلك لرفع دافعيتهم.

ب. معامل صعوبة البند:

للحصول على معامل الصعوبة **نطرح** معامل السهولة من واحد، فإذا كان معامل السهولة هو 0.40 فإن معامل صعوبته هو: 1- 0.40 = **0.60**

كما يمكن حساب معامل الصعوبة كما يلي:

عدد الإجابات الخاطئة عن البند ÷ مجموع الإجابات الصحيحة والخاطئة

6. المعايير والدرجات المشتقة:

تستخدم المعايير والدرجات المشتقة أو المحولة في الإختبارات المقننة، سواء كانت متعلقة بالإستعدادات أو التحصيل أو الشخصية، وهي درجات صممت خصيصا لمقارنة درجات الفرد بمعايير المجموعة.

تكون درجات المفحوص في اختبار معين غير كافية لوحدها ولا معنى لها، إذ يجب أن يكون هناك معيار يحدد معنى هذه الدرجة أو هذه النسبة المئوية ويدلنا على مركز الشخص بالنسبة للمجموعة.

يشتق المعيار من عينة التقنين التي تمثل المجتمع الأصلي المدروس، فتكون الدرجات المُحولة للدرجات الخام لأفراد هذه العينة مصدرا "للمعيار"، أي أننا نقارن درجة المفحوص في اختبار ما بالعودة إلى ما يكافئه في عينة التقنين.

تعتبر عملية استخراج معايير الإختبار أو عملية التعيير (أو المعايرة أو التقنين) la standardisation المرحلة الأخيرة من مراحل إعداد وتطوير الإختبار، ويطلق على العينة التي يطبق عليها الإختبار بهدف استخراج معاييره إسم عينة التعيير أو عينة التقنين.

هناك 3 طرق لتحويل الدرجات الخام إلى درجات يمكن مقارنة بعضها بالبعض الآخر:

– طرق مبنية على حساب ترتيب التلميذ مثلا بالنسبة لمجموعته في الصف أو بالنسبة لعينة التقنين كالمئينات والأعشاريات.

– طرق مبنية على التحويل الخطي للدرجات الخام كالدرجات المعيارية.

– طرق مبنية على التحويل المساحي للدرجات الخام كالدرجات التائية، المعيار الجيمي،...

❖ **أنواع المعايير:**

يميز المختصون في القياس النفسي والتربوي بين 4 أنواع من المعايير:

1. **معايير العمر Age norms:** تعبر عن القيمة الكمية المتوسطة للسمة المقاسة لدى أبناء العمر الواحد، وتظهر أهميتها عند قياس السمات التي تنمو مع العمر كالذكاء.

2. **معايير الصف Grade norms:** لا تختلف معايير الصف أو معايير الفريق الدراسي من حيث جوهرها عن المعايير العمرية، وتعتمد مستوى المتوسط أساسا في الحكم على أداء الفرد وتتيح مقارنته بهذا المستوى.

3. **المعايير المئينية Percentil norms:** حيث نستخدم المئين الذي يحدد مباشرة مركز الفرد في المجموعة معتمدا في ذلك على تحديد النسبة المئوية للحالات (أو الدرجات) التي تقع تحته في التوزيع، وقد أطلق على المعيار المئيني هذه التسمية لأنه يحول التوزيع التكراري إلى أجزاء من المائة أو أجزاء مئوية ويقسم مستويات الأفراد إلى مائة مستوى.

4. **معايير الدرجة المعيارية Standard score norms:** هي المسافة التي تبعد فيها الدرجة الخام عن المتوسط الحسابي معبر عنها في وحدات من الإنحراف المعياري وتسهل الدرجة المعيارية مقارنة المجموعات مع بعضها البعض حتى وإن اختلفت متوسطاتها:

الدرجة المعيارية= الدرجة الخام – المتوسط الحسابي ÷ الإنحراف المعياري

فإذا كانت الدرجات الخام التي حصل عليها 3 مفحوصين في اختبار المفردات هي 33، 25، 29 وكان متوسط الدرجات في المجموعة التي طبق عليها هذا الإختبار هو 29 وانحرافها المعياري 4، فإن الدرجة المعيارية التي تقابل الدرجة الخام للمفحوص الأول هي: 33 – 29 ÷ 4 = 4 ، والتي تقابل الدرجة الخام للمفحوص الثاني هي:

25 – 29 ÷ 4 = – 1، والتي تقابل الدرجة الخام للمفحوص الثالث هي: 29 – 29 ÷ 4 = 0. وهذا يعني أن الدرجة التي حصل عليها المفحوص الأول مثلا تبعد بمقدار +1 درجة معيارية (1 سيغما) عن المتوسط.

من الواضح أن الدرجات المعيارية المعادلة للدرجات الخام التي تقع تحت المتوسط تحمل إشارات سالبة وأن الدرجات المعادلة للدرجات الخام التي تقع فوق المتوسط تكون موجبة، في حين أن الدرجة المعيارية المعادلة للدرجات الخام التي تقع في نقطة الوسط تعادل الصفر.

تتراوح الدرجات المعيارية التي تقع على منحنى التوزيع الطبيعي أو الإعتدالي والذي تعتمده الإختبارات النفسية والتربوية كما هو معلوم من + 3 إلى – 3 مرورا بالصفر والذي يشير إلى نقطة الوسط تماما في التوزيع.

5. الدرجات المعيارية المعدلة:

أ. الدرجة التائية: وتصلح للإستخدام في المقاييس النفسية والتربوية كافة ويشيع استخدامها في مقاييس الشخصية، وتنسب الدرجة التائية T–Score أو المعيار التائي إلى **إدوارد ثورندايك** و**لويس ترمان** وترتكز على توزيع افتراضي للدرجات متوسطه (50) وانحرافه المعياري (10) وتحسب كما يلي:

الدرجة التائية (ت)= الدرجة المعيارية × 10 + 50

ب. الدرجة الموزونة: يشيع استخدام هذا النوع من الدرجات في اختبارات الذكاء التي اعتمدت المعايير غير الطولية ومنها اختبار **وكسلر**، وهي درجة معيارية معدلة تستهدف توحيد أوزان درجات الإختبارات الفرعية التي قد يضمها الإختبار الكلي الواحد، أي جعلها من وزن واحد بغض النظر عن النهاية العظمى لدرجات كل من هذه الإختبارات، وتقوم الدرجة الموزونة على توزيع متوسطه (10) وانحرافه المعياري (3) وتحسب باستخدام المعادلة التالية:

الدرجة الموزونة= الدرجة المعيارية × 3 + 10

ت. الدرجة الجيمية: وتصلح لقياس الفروق الكبيرة بين الأفراد، حيث تقلص تلك الفروق إلى عدد ضئيل من المستويات من خلال تقسيم التوزيع إلى 11 قسما أو مستوى يبدأ تدريجه من 0 إلى 10، وتنسب هذه الدرجة الجيمية G–Score إلى **جيلفورد** وتقوم على إخضاع الدرجات المعيارية لتوزيع جديد متوسطه (5) وانحرافه المعياري (2) ومعادلتها:

الدرجة الجيمية= الدرجة المعيارية × 2 + 5

المراجع:

- إمطانيوس نايف، م، **بناء الإختبارات والمقاييس النفسية والتربوية وتقنينها**، الإعصار للنشر والتوزيع، الأردن، 2016.

- شاكر الجبلي، س، **أساسيات بناء الإختبارات والمقاييس النفسية والتربوية**، مؤسسة علاء الدين للطباعة والتوزيع، سوريا، 2005.

- طالب الريماوي، ع، **بناء وتصميم الإختبارات والمقاييس النفسية والتربوية**، دار أمجد للنشر والتوزيع، الأردن، 2017.