الجمهورية الجزائرية الديمقراطية الشعبية
People's Democratic Republic of Algeria
وزارة التعليم العالي والبحث العلمي
Ministry of Higher Education and Scientific Research
جامعة الشهيد الشيخ العربي التبسي – تبسة
Echahid Cheikh Larbi Tebessi University - Tebessa

Faculty of Humanities and Social Sciences
Department of psychology

كلية العلوم الانسانية والاجتماعية
قسم علم النفس

مطبوعة بيداغوجية في مـــــادة:

القيـــاس النفســـي

Psychometrics

موجهة لطلبة السنة الثانية علم النفس

إعداد الدكتور

زيـــاد رشيـــد

أستـــاذ محـــاضر قسم -أ-

جامعة تبسة
UNIVERSITE DE TEBESSA

السنـــة الجامعيـــة: 2025 ـ 2026

تقديـــــــــــم:

مطبوعة بيداغوجيـــــة محكمـــــة موجهة لطلبة السنة الثانية شعبة علم النفس، المطبوعة عبارة عن مجموعة من المحاضرات المنسجمة وفق محاور مادة القياس النفسي، بحسب البرنامج الرسمي المعتمد من طرف وزارة التعليم العالي والبحث العلمي.

عنوان الليسانس: علم النفس

السداسي: الأول **اسم الوحدة:** المنهجية

اسم المادة: القيـاس النفســي 1

الرصيد: 3 **المعامل:** 2

- أهداف التعليم:

1- تعرف الطالب على القياس النفسي وعلى علم النفس القياسي.

2- تمكين الطالب من التعرف على الخواص النفسية وطرق تقييمها.

3- التعرف على مستويات القياس النفسي ومعالجتها الإحصائية.

- المعارف المسبقة المطلوبة:

1- معارف في المجال الاحصائي الوصفي وأهداف استخدامه في اختصاصه.

2- القدرة على تصنيف الخواص + القدرة على تحديد مستوى القياس.

3- ادراكه لمعنى أداة قياس والتمييز بين أنواعها.

- القدرات المكتسبة:

1- إدراك ما هية القياس النفسي وأهداف استخدامه في اختصاصه.

2- القدرة على تصنيف الخواص + القدرة على تحديد مستوى القياس.

3- ادراكه لمعنى أداة قياس والتفريق بين أنواعها.

- مفردات المادة:

1- مدخل عام للقياس النفسي (مفهوم القياس، مفهوم علم النفس القياسي، المفاهيم ذات الصلة، التقدير، التقييم...الخ)

2- التطور التاريخي (حركة القياس النفسي).

3- مسلمات القياس النفسي وخصائصه.

4- أهداف وأهمية القياس النفسي.

5- مجالات القياس النفسي (القياس العقلي، قياس الشخصية).

6- مدارس القياس النفسي.

7- مفهوم الخاصية والسمة (أنواع السمات أو الخواص من حيث التكوين ومن حيث التقدير، طرق قياسها).

8- القياس الأسمي أو التصنيفي+ معالجته الإحصائية.

9- القياس الرتبي + معالجته الإحصائية.

10- القياس الفئوي + معالجته الإحصائية.

11- القياس النسبي + معالجته الإحصائية.

12- الاستخدامات الأساسية لأدوات القياس والتقييم النفسي.

13- تصنيف أدوات القياس والتقييم النفسي.

14- مشكلات القياس في علم النفس.

15- الاختبارات والاستخدام الاكلينيكي للاختبارات النفسية.

- طريقة التقييم: مراقبة مستمرة، امتحان، ...الخ

السداسي: الثاني **اسم الوحدة:** المنهجية

اسم المادة: القيـاس النفســي 2

الرصيد: 3 **المعامل:** 2

- أهداف التعليم:

1- التعرف على المفاهيم الأساسية في بناء الاختبارات.

2- التعرف على الاستخدامات الأساسية لأدوات القياس ومعايير تصنيفها.

3- التعرف على الخطوات العملية المتبعة لبناء الاختبارات.

4- التعرف على مفهوم الثبات والصدق وتطبيق الطرائق المختلفة لتقدير كل منهما.

5- التمييز بين التفسيرات المعيارية والمحكية المرجع للدرجات.

- المعارف المسبقة المطلوبة:

1- تاريخ انفصال علم النفس عن الفلسفة.

2- معرفة بعض الأدوات المستخدمة في القياس الطبيعي وفي القياس النفسي وفي الحياة اليومية.

3- فهم العناصر الرئيسية للنظرية الكلاسيكية في القياس النفسي.

4- معارف أساسية في الإحصاء الوصفي والاحصاء الاستدلالي.

- القدرات المكتسبة:

1- إدراك مفهوم القياس والتقييم ومكونات عملية القياس النفسي.

2- إدراك الافتراضات الأساسية لعلمية القياس والتقييم ومستوياتها.

3- إدراك مفهوم الثبات والصدق والفروق الأساسية بين التفسير المعياري والمحكي للدرجات.

- مفردات المادة:

1- المفاهيم الأساسية في بناء الاختبارات.

2- الخطوات الأساسية في بناء الاختبارات النفسية.

3- بناء جدول المواصفات.

4- بناء الفقرات الاختبارية وتحليل البنود.

5- احتساب الشروط السيكومترية (الصدق، الثبات، المعايير).

6- التفسير المرجعي والتفسير المحكي للدرجات.

7- اعداد دليل الاختبار.

8- اختبارات الشخصية.

9- اختبارات الذكاء.

10- اختبارات الاتجاهات والقيم والميول.

11- نماذج من بطاريات التقييم والقياس.

12- مدخل عام إلى النظرية القياسية الحديثة.

13- أسس ومبادئ وتوجهات النظرية القياسية الحديثة.

14- نماذج من النظرية القياسية الحديثة.

15- المسؤولية الأخلاقية في ممارسة القياس والتقييم النفسي.

- طريقة التقييم: مراقبة مستمرة، امتحان، ...الخ

الدكتور/ زيـاد رشـيد **تأشيرة رئيس القسم**

أستاذ محاضر قسم "أ"

فهـرس المحتويــات

مقدمة 09

المحــاضرة الأولى
مدخل للقيــاس النفســـي

تمهيد: 12
أولاً: مفهوم القياس في علم النفس 12
1- الرواد الأوائل وتجاربهم 12
2- تطور المنهج العلمي 13
3- نشوء القياس النفسي 13
ثانياً- تعريف القياس 13
ثالثاً: المفاهيم ذات الصلة (التقويم – التقييم - التقدير) 16
1- التقويم (Evaluation) 16
2- التقييم (Assessment) 17
3- التقدير (Rating) 18
4- الفرق بين القياس والتقويم والتقييم والتقدير 18

المحــاضرة الثانية
التطور التاريخي (حركة القياس النفسي)

تمهيد: 21
أولاً: بدايات ظهور حركة القياس النفسي 21
1- مفهوم المعادلة الشخصية 22
2- السيكوفيزيقا 23
ثانيًا: القياس والتأخر العقلي 27
1- بينيه وقياس الذكـــاء 28
ثالثاً: القياس النفسي في إنجلترا 29
رابعًا: القياس النفسي في أمريكا 30

المحـــاضرة الثالثة
مسلمات القياس النفسي وخصائصه

تمهيد: 32
أولاً- مسلمات تستند إليها نظريات القياس في علم النفس 32
1- قابلية أداء الفرد للقياس 33

2- أداء الأفراد له خصائص مميزة 34
3- اختلاف العلاقة بين الخصائص والأداء من فرد لآخر 36
4-الخطأ مكون أساسي من مكونات عملية القياس 37

المحـــاضرة الرابعة
مجـــالات القيـــاس النفســـي (القياس العقلي)

تمهيد: 41
أولاً: القياس العقلي (قياس القدرات العقلية) 41
1- القدرات العقلية (Mental abilities / Cognitive abilities) 41
2- تصنيف القدرات العقلية 41
ثانياً: مفهوم الذكـــاء 42
ثالثاً: مشكلة تعريف الذكـــاء 45
رابعاً: العوامل المؤثرة في الذكـــاء 47
خامساً: التفاعل النسبي بين الوراثة والبيئة 49
سادساً: نماذج ونظريات الذكاء 50
سابعاً- أدوات قياس الذكاء العام 57
ثامناً: تطبيقات القياس العقلي في التعليم والإرشاد والعيادات النفسية 65
تاسعاً: القياس العقلي في البيئة العربية (التحديات الثقافية والتقنين) 66
عاشراً: مفهوم الاستعداد وعلاقته بمفهوم القدرة 66
الحادي عشر: الإبتكاريـــة ومفهوم التفكير التباعـدي 69

المحـــاضرة الخامسة
مجـــالات القيـــاس النفســـي (قياس الشخصية)

تمهيد: 73
أولاً: الشخصيـــة (Personality) 73
1- مفهوم الشخصية 74
2- نظريات الشخصية 75
3- تصنيف مقاييس الشخصية 78
4- مناهج قياس الشخصيـــة 79
ثانيـــاً: قياس الميول من منظور تاريخي 82
1- تعريف الميول (Interest) 84
2- قياس الميول (Interest Measurement) 85
3- بعض المشكلات المتعلقة بالإستبيانات المقننة للميول 86

ثالثاً: الاتجـاهــات (Attitudes) 90
1- قياس الاتجاهات من منظور تاريخي 90
2- تصنيف مقاييس الاتجاهات 92
3- أساليب بنـاء موازين الاتجـاهــات 94
رابعاً: بعض المشكلات المتعلقة بقياس الجوانب غير المعرفية 106

المحـــاضرة السادسة
بنـــاء وتصميم الاختبـــارات النفسيـــة والتربويـــة

تمهيد: 109
أولاً: قواعد إرشادية في تطوير المقياس 109
الخطوة الأولى: تحديد ما يراد قياسه بدقة 109
الخطوة الثانية: توليد مجموعة من البنود 111
الخطوة الثالثة: تحديد شكل أو تنسيق القياس 116
الخطوة الرابعة: مراجعة تجمع البنود الأولية من قبل الخبراء 131
الخطوة الخامسة: النظر في إدراج البنود الداعمة للصدق 132
الخطوة السادسة: تطبيق البنود على عينة تطوير المقياس 132
الخطوة السابعة: تقييم البنود 133
الخطوة الثامنة: تحقيق الحد الأمثل لطول المقياس 134

المحـــاضرة السابعة
الخصائص السيكومترية للفقرات

تمهيد: 139
1- معامل صعوبة الفقرة Item Difficulty Index 139
2- معامل سهولة الفقرة Item Easiness Index 142
3- معامل تصحيح الصعوبة من أثر التخمين Guessing 143
4- قوة تمييز فقرات الاختبار Item Discrimination Index 145
5- طرق أخرى لتقييم تمييز فقرات الاختبار 149
6- المموهات (المشتتات) Distractions 150
7- معامل ثبات المفردة 153

المحـــاضرة الثامنة
التحقق السيكومتري من أدوات القياس

تمهيد: 156

1- خصائص الاختبارات والمقاييس الجيدة 156
2- الشروط العامة للاختبـار النفسـي والتربوي 157
3- الشروط التجريبيــة أو السيكومتريــة (Psychometric Conditions) 157
3-1- الصدق (Validity) 158
3-2- العوامل المؤثرة على الصدق 167
3-3- الثبـات (Reliability) 168
3-4- العوامل المؤثرة على ثبات الاختبار 184

المحــاضرة التاسعة
نمــاذج من بطاريــات التقييــم والقيــاس

تمهيد: 187
أولاً- نماذج من اختبارات القدرات العقلية 187
1- مقياس استانفورد – بينيه (Stanford-Binet Scale) 187
1-1- مفهوم العمر العقلي (Mental Age Concept) 189
2- مقياس وكسلر للذكــاء النسخة الرابعـة (Wechsler Intelligence Scale- WISC_1V) 190
3- اختبـار المصفوفـات المتتابعـة لرافن (Raven) 194
4- اختبار أوتس – لينون للقدرة العامة (Otis-Lennon Mental Ability Test (OLMAT)) 196
5- مقياس الينوي للقدرات السيكولغوية (Illinois Test of Psycholinguistic Abilites) 197

المحــاضرة العاشرة
مدخل عــام إلى النظريــة الحديثــة في القيــاس

تمهيد: 201
1- التطور التاريخي لنظريــة الاستجابـة للمفردة (IRT) 201
2- افتراضات نماذج نظريـة الاستجابـة للمفردة الاختباريـة أحاديـة البعد 203
2-1- أحاديـــة البعد Unidimensionality 203
2-2- الاستقـلال الموضعـي Local Independence 204
2-3- توازي المنحنيات المميزة للمفردات Item Characteristic Curve (ICC) 205
2-4- الحد الأدنـى من أثر التخمين Minimal guessing 208

المحــاضرة الحادية عشر
أسس ومبادئ وتوجهات النظرية الحديثة في القياس

تمهيد: 210
1- معنى الموضوعيـة في نموذج راش 210
2- الجانب السيكولوجي في نموذج راش 212

3- الاختبارات التحصيلية المبنية على نظرية الاستجابة للمفردة 213
4- صدق الاختبار وثباته في ضوء نموذج راش اللوغاريتمي الاحتمالي 213
5- بنوك الأسئلة (Item Banking) 214
6- الاختبارات التكيفية المحوسبة (Computerized Adaptive Testing - CAT) 218

المحـاضرة الثانية عشر
نمـاذج من النظريـة الحديثـة في القيـاس

تمهيد: 221
أولاً- أنواع نمـاذج الاستجابـة للمفردة أحاديـة البعد 221
1- النماذج الاستاتيكية (Static Models) 221
1-1- نمـاذج تتعلق بالمفردة ثنائية الدرجـة 222
1-1-1- النموذج آحادي المعلم ((One-Parameter Logistic Model (1PL)) 222
1-1-2- النموذج ثنائـي المعلم ((Two-Parameter Logistic Model (2PL)) 225
1-1-3- النموذج ثلاثـي المعلم ((Three-Parameter Logistic Model (3PL)) 227

المحـاضرة الثالثة عشر
المسؤولية الأخلاقيـة في ممارسة القيـاس والتقييم النفسـي

تمهيد: 231
أولاً- الاعتبارات الأخلاقية في القياس والتقويم 231
1- المعايير الأخلاقية في التقويم التربوي 232
2- قضايا العدالـة والمساواة 232
3- السريـة وحماية البيانات 233
ثانياً- أخلاقيات البحث في القياس والتقويم 234

الخاتمة 136
قائمـــة المصادر والمراجع 137

مقـدمـــــة:

تتناول هذه المطبوعة البيداغوجية في مادة القيـاس النفســي مجموعة من المحاضرات المتسلسلة التي تهدف إلى بناء فهم تدريجي ومتكامل لمجال القياس في علم النفس. تبدأ **المحاضرة الأولى** بمدخل إلى القياس النفسي من خلال توضيح مفهوم القياس في علم النفس، مع التطرق إلى الرواد الأوائل وتجاربهم، وتطور المنهج العلمي ونشوء القياس النفسي، إضافة إلى تعريف القياس والمفاهيم ذات الصلة مثل التقويم والتقييم والتقدير والتمييز بينها.

وتأتي **المحاضرة الثانية** لتعرض التطور التاريخي لحركة القياس النفسي، من خلال مفهوم المعادلة الشخصية والسيكوفيزيقا، ثم القياس والتأخر العقلي مع بينيه وقياس الذكاء، إضافة إلى تطور القياس النفسي في إنجلترا والولايات المتحدة الأمريكية.

أما **المحاضرة الثالثة** فتتناول مسلمات القيـاس النفسـي وخصائصه، من خلال إبراز قابلية الظواهر النفسية للقياس، وتميز أداء الأفراد، واختلاف العلاقات بين الخصائص والأداء، واعتبار الخطأ عنصرًا أساسيًا في عملية القياس.

وتعالج **المحاضرة الرابعة** مجالات القيــاس النفسـي في بعده العقلي، من خلال التطرق إلى القدرات العقلية والذكاء، تعريفه وإشكالاته ونظرياته، والعوامل المؤثرة فيه، إضافة إلى التفاعل بين الوراثة والبيئة، وأدوات قياس الذكاء وتطبيقاته، إلى جانب الاستعداد والإبداع والتفكير التباعدي.

وتخصص **المحاضرة الخامسة** لقياس الشخصيـة، من خلال تناول مفهوم الشخصية ونظرياتها، وتصنيف ومناهج قياسها، إضافة إلى قياس الميول والاتجاهات تاريخيًا وأساليب بنائها، مع الإشارة إلى المشكلات المرتبطة بقياس الجوانب غير المعرفية.

وتعرض **المحاضرة السادسة** أسس بناء وتصميم الاختبـارات النفسية والتربوية، بدءًا من تحديد ما يراد قياسه، وتوليد البنود، وتحديد شكل الاختبار، وتحكيم البنود، وتطبيقها على العينة، ثم تقييمها وتحديد الطول الأمثل للمقياس.

أما **المحاضرة السابعــة** فتتناول الخصائص السيكومترية للفقرات، كمعامل الصعوبة والسهولة، وتصحيح أثر التخمين، وقوة التمييز، والمموهات، وطرق تقييم التمييز، وثبات المفردة.

وتأتي **المحاضرة الثامنــة** لتعرض التحقق السيكومتري من أدوات القيـاس، من خلال خصائص الاختبارات الجيدة، والشروط العامة والتجريبية، إضافة إلى مفهومي الصدق والثبات والعوامل المؤثرة فيهما.

وتتناول **المحاضرة التاسعة** نماذج من بطاريات القياس، من خلال استعراض أهم اختبارات القدرات العقلية مثل ستانفورد–بينيه، وكسلر، ورافن، وأوتس–لينون، ومقياس الينوي للقدرات السيكولوجية.

وتتجه **المحاضرة العاشرة** إلى مدخل النظرية الحديثة في القياس، من خلال التطور التاريخي لنظرية الاستجابة للمفردة (IRT)، وافتراضاتها الأساسية كأحادية البعد، والاستقلال الموضعي، ومنحنيات الاستجابة للمفردة، وأثر التخمين.

وتعالج **المحاضرة الحاديـة عشرة** أسس ومبادئ النظرية الحديثة في القياس، من خلال مفهوم الموضوعية في نموذج راش، والبعد السيكولوجي له، وتطبيقات IRT في الاختبارات التحصيلية، إضافة إلى الصدق والثبات في إطار النموذج اللوغاريتمي، وبنوك الأسئلـة والاختبـارات التكيفية المحوسبة.

أما **المحاضرة الثانيـــة عشرة** فتتناول نماذج النظريـة الحديثـة في القيـاس، وبالأخص النماذج الاستاتيكية أحادية البعد، مثل النموذج أحـادي المعلم (1PL)، وثنائـي المعلم (2PL)، وثلاثـي المعلم (3PL).

وتختتم المطبوعة **بالمحاضرة الثالثـة عشرة** التي تتناول المسؤوليـة الأخلاقيـة في القيـاس والتقييم النفســي، من خلال التركيز على المعايير الأخلاقيـة، وقضايا العدالــة والمساواة، وسريـة البيانات وحمايتها، إضافة إلى أخلاقيات البحث في القيـاس والتقويـم.

المحـــاضرة الأولى

مدخل للقيـــاس النفســـي

الأهداف:

تهدف هذه المحاضرة إلى تمكين الطلبة من فهم الأسس النظرية والتاريخية للقياس في علم النفس، من خلال التعرف على إسهامات الرواد الأوائل وتجاربهم التي مهدت لظهور القياس النفسي، واستيعاب مراحل تطور المنهج العلمي التي أسهمت في ترسيخ هذا المجال، وصولاً إلى نشوء القياس النفسي كحقل علمي مستقل. كما تسعى المحاضرة إلى توضيح مفهوم القياس وتعريفه بدقة، ثم التمييز بينه وبين المفاهيم ذات الصلة، وذلك عبر عرض مفاهيم التقويم والتقييم والتقدير وشرح خصائص كل منها ووظائفها. وفي الأخير، تهدف إلى تنمية قدرة الطلبة على إدراك الفروق الجوهرية بين القياس والتقويم والتقييم والتقدير، بما يعزز فهمهم لاستخداماتها التطبيقية في المجالات النفسية والتربوية.

محتوى المحاضرة:

أولاً: مفهوم القياس في علم النفس
2- الرواد الأوائل وتجاربهم
2- تطور المنهج العلمي
3- نشوء القياس النفسي
ثانياً- تعريف القياس
ثالثاً: المفاهيم ذات الصلة (التقويم – التقييم - التقدير)
1- التقويم (Evaluation)
2- التقييم (Assessment)
3- التقدير (Rating)
4- الفرق بين القياس والتقويم والتقييم والتقدير

مدة المحاضرة: ساعة ونصف

تمهيد:

يعد القياس (Measurement) ممارسة قديمة قدم الإنسان نفسه. ففي العصور البدائية، كان الإنسان يعتمد على جسده بوصفه أداة للقياس؛ فعند إعداد ملابسه، كان يضع جلد الحيوان على جسده، ويقص الزائد عن طوله. وعندما يحفر كهفًا ليقيم فيه مع أسرته، كان يقارن ارتفاع مدخله بطوله، ويقدّر سعته بعدد أفراد أسرته من خلال الجلوس داخله. وبذلك كان يقيس طول الجلد ومدخل الكهف بطوله الشخصي، وكأن جسده يمثل شريط قياس طبيعيًا.

ولا تزال بعض الأساليب المشابهة لهذه الممارسات مستخدمة حتى اليوم، مثل القياس بالشبر والذراع والقدم والأصابع والعصا والحبل. غير أن هذه الوسائل واجهت مشكلات عديدة نتيجة اختلافها من شخص إلى آخر، مما أفقدها الدقة والموضوعية، إذ إن أحجام الأيدي والأصابع والأذرع والأقدام تختلف بين الأفراد. وقد أدى ذلك إلى الحاجة إلى تطوير وسائل قياس موحّدة، أكثر دقة وموضوعية وثباتًا، فكان ظهور وحدات قياس معيارية كالمتر واللتر والكيلوجرام وغيرها من الوحدات المتعارف عليها عالميًا.

أولاً: مفهوم القيــاس في علم النفس: القياس هو الأسلوب العلمي الذي يحول الأوصاف اللفظية الكيفية لخصائص الأشياء والأشخاص إلى كميات، مما يجعلها قابلة للتناول الموضوعي. وعلم النفس، كغيره من العلوم الإنسانية وغير الإنسانية، يتطور باستمرار. فقد كان تابعاً للفلسفة، يتبع منهجاً في البحث، وخاصة الجدل المنطقي. ولكنه ما لبث أن انفصل عنها في النصف الثاني من القرن التاسع عشر، واتخذ المنهج العلمي وأساليبه في البحث.

1- الرواد الأوائل وتجاربهم:

اعتمد علم النفس في البداية على دراسات الحالة الفردية والخبرة الشعورية. وقد أسهم في هذا الاتجاه كل من:

- جوستاف تيودور فخنر (G. T. Fechner): في تجاربه على العلاقة بين الجسم والنفس سنة 1860.
- هرمان فون هلمهولتز (H. V. Helmholtz): في تجاربه على الإحساسات السمعية والبصرية.
- فلهلم فونت (W. Wundt): في تجاربه على الخبرة الشعورية سنة 1879. وقد توصل إلى أن الخبرة الشعورية تتضمن ثلاثة مكونات هي: المعرفة والوجدان والنزوع.
- هرمان إبنجهاوس (H. Ebbinghaus): في تجاربه على الذاكرة سنة 1885. وقد اتخذ من نفسه موضوعاً لتجاربه واستنبط منحنى النسيان.

2- تطور المنهج العلمي: أجرى هؤلاء الرواد دراساتهم على حالات فردية، ولم يكونوا يستخدمون الإحصاء لاستخلاص المتوسطات الحسابية كما يفعل الباحثون اليوم، لأن الإحصاء لم يلتقِ بعلم النفس في تلك الفترة. بل كانوا يعتمدون على تكرار التجارب على نفس الأفراد وعلى أفراد آخرين لاستنباط قوانين قابلة للتعميم. ومع مطلع القرن العشرين، تطور المنهج من دراسة الحالة الفردية إلى دراسة المجموعات، ومن الوصف الكيفي إلى الوصف الكمي. كما وضع **"بينيه"** أول اختبار مدرج للذكاء.

3- نشوء القياس النفسي: ساهمت ثلاثة تيارات فكرية وعلمية في نشوء القياس النفسي وهي:

1. الاهتمام بالفروق الفردية.
2. حركة الاختبارات النفسية.
3. الاهتمام بإدخال الإحصاء في البحوث النفسية.

كان الدافع القوي لإدخال القياس هو رغبة السيكولوجيين في أن يحذوا حذو العلوم الطبيعية عبر تحديد الخصائص النفسية كمياً. والطريقة الشائعة هي إعطاء الفرد درجة معينة تصف كمياً ما يوجد لديه من خاصية معينة (معمرية، 2012، ص29).

ثانياً- تعريف القياس:

يرجع أصل مصطلح **القياس** في اللغة العربية إلى الفعل قاس، أي قدّر، فيُقال: قاس الشيء بغيره أو على غيره، أي قدّره بمثاله. أما في الاصطلاح العلمي، فقد تعددت تعريفات القياس لدى الباحثين تبعًا لزاوية النظر إليه.

- فقد عرّفه كامبل (Campbell, 1952) بأنه: تمثيل الصفات أو الخصائص بأرقام.

- بينما يرى ستانلي سميث ستيفنس (S. S. Stevens, 1951) أن القياس، في أوسع معانيه، هو: عملية إسناد أرقام إلى الأشياء أو الأحداث وفقًا لقوانين محددة. ويعني ذلك أن قياس أي شيء يستلزم إجراء مقارنات بينه وبين معيار أو مقياس معين، وفق قواعد منظمة.

- أما جوي ب. جلفورد (J. P. Guilford, 1954) فيعرّف القياس بأنه: وصف البيانات بالأرقام.

- في حين يرى ج. ل. كرونباخ (J. Lee Cronbach, 1969) أن القياس هو: طريقة منظمة لمقارنة سلوك شخصين أو أكثر.

- ويذهب بين (K. Bean) إلى أن القياس يتمثل في: مجموعة من المثيرات المرتبة التي تهدف إلى القياس الكمي لبعض العمليات الانفعالية أو العقلية أو النزوعية، وقد تكون هذه المثيرات في صورة أسئلة مكتوبة أو شفهية، أو سلاسل من الأعداد أو الأشكال أو النغمات.

- أما ج. س. نانالي (J. C. Nunnally, 1970) فيعرّف القياس بأنه: مجموعة من القواعد لاستخدام الأعداد للدلالة على الأشياء بطريقة تعكس مقدار خاصية معينة. ويتضمن تعريف "نانالي" ثلاثة مفاهيم أساسية هي: **القواعد، والأعداد، والخاصية**.

- يشير مفهوم **القواعد** إلى الإجراءات المنظمة التي تحكم استخدام الأعداد، والتي ينبغي أن تكون واضحة وصريحة وقابلة للفهم والتواصل. وتُستخدم هذه القواعد في التعامل مع البيانات العددية تبعًا لمستويات القياس المختلفة، كالقياس الاسمي أو الرتبي أو الفتري، وهو ما يترتب عليه اختيار الأساليب الإحصائية الملائمة، سواء كانت بارامترية أو لا بارامترية. وقد تكون هذه القواعد شديدة الدقة والوضوح كما في قياس الأطوال والمسافات، إلا أنها تكون أقل تحديدًا في ميدان العلوم الإنسانية.

- أما مفهوم **العدد** فيرتبط بفكرة **الكم**، أي مقدار ما يتوافر في الشيء أو الفرد من خاصية معينة. وتُستخدم الأعداد للتعبير عن هذا المقدار، ويصل التداخل بين مفهومي القياس والكم إلى حد استخدامهما أحيانًا على نحو مترادف.

- في حين يشير مفهوم **الخاصية** إلى أن القياس لا ينصب على الأشياء أو الأفراد في ذاتهم، وإنما على صفاتهم وخصائصهم. فنحن لا نقيس الشخص ذاته، بل نقيس ذكاءه أو دافعيته أو مستوى قلقه، كما نقيس الطول والوزن وغير ذلك من الخصائص.

وقد ورد في قاموس وبستر (Webster) أن القياس هو: التحقق بالتجربة أو الاختبار من المدى أو الدرجة أو الكمية أو الأبعاد أو السمة باستخدام معيار محدد. ويتضح من هذا التعريف، كما من التعريفات السابقة، أن القياس عملية كمية، وأن نتائجه يُعبَّر عنها بالأرقام.

وبناءً على ما سبق، يمكن تعريف القياس تعريفًا جامعًا بأنه: العملية التي يتم من خلالها تحديد مقدار ما يتوافر من خاصية معينة لدى الفرد أو في الشيء أو الظاهرة.

ويُستخدم مصطلح القياس في الحياة اليومية للدلالة على مجموعة من العمليات التي تمكّن الإنسان من الحصول على قياسات دقيقة للأشياء، مما يسهم في تنظيم التعامل بين الأفراد. وتتحدد قيم الأشياء المتداولة وفق مقاييس متعارف عليها؛ فالأطوال تُقاس لتحديد المسافات والمساحات والأحجام

باستخدام المقاييس المترية كوحدات المتر والسنتيمتر، كما تُقاس الأوزان بوحدات الجرام والكيلوجرام، وتُقاس درجات الحرارة بوحدات السنتيغريد أو الفهرنهايت.

وهكذا، تقاس الخصائص المختلفة، كل بحسب مقياسه ووحداته الثابتة التي اتفق الناس على استخدامها، وذلك بهدف تحديد الكميات أو الفروق الكمية بين الأشياء، بما يحقق قدرًا من الثبات والاتساق في القيم والمفاهيم والمعاني، بحيث لا تختلف من فرد إلى آخر أو من موقف إلى آخر.

وبناءً على ذلك، تستعمل كلمة **"القياس"** في معانٍ متعددة. فقد تستخدم للدلالة على **عملية القياس** ذاتها، أو للإشارة إلى **نتائج القياس**، أو إلى **أدوات القياس**، كما قد يُقصد بها **الوحدات** التي تتضمنها المقاييس.

ويشير إنجلش وإنجلش (English & English, 1958) إلى أن مصطلح القياس يستخدم بعدة دلالات، سواء جاء في صورة فعل أو اسم، ومن أبرز هذه الدلالات ما يأتي:

1- يدل القياس على **النتيجة** المتحصل عليها من عملية القياس، أي القيمة التي يتم التوصل إليها عند قياس شيء ما، بما يتضمنه ذلك من تقدير كمي لوجود الخاصية أو غيابها، أو لدرجة توافرها.

2- ويشير القياس إلى **الوحدة أو المعيار** المستخدم، كأن يقال إن القياس تم بالجرامات أو الأمتار أو الساعات أو الدقائق، أو غيرها من الوحدات المعتمدة في المقاييس المختلفة.

3- كما يستخدم مصطلح القياس للدلالة على **تقديرات إحصائية** لخصائص الأشياء؛ فالمتوسط الحسابي يعد مقياسًا، وكذلك الانحراف المعياري ومعامل الارتباط، إذ يعبّر كل منها عن خاصية أو سمة تميّز الظواهر أو الأشياء المدروسة.

وفي ضوء هذه الدلالات المتعددة، يُقصد بالقياس أنه **عملية مقارنة** شيء ما بوحدات محددة، أو بكمية معيارية، أو بمقدار مقنن من الشيء ذاته أو من خاصيته، وذلك بغرض تحديد عدد الوحدات التي يتضمنها هذا الشيء. فنقيس الطول بوحدات مترية، ونقيس درجة حرارة الجو بوحدات منتظمة ناتجة عن تمدد الزئبق في الترمومتر بفعل الحرارة، كما نقارن سلوك فرد معين بعينة من السلوك الشائع في المجتمع كما هو ممثل في الاختبارات النفسية.

ووفقًا لهذا المفهوم الواسع، يشمل القياس **عملية القياس نفسها، والأداة المستخدمة فيه، ووحدات هذه الأداة أو المقياس**ـــــسواء كانت سنتيمترات أو غرامات أو بنودًا اختباريةـــــإضافة إلى **القيمة العددية** التي تعبّر عن نتيجة استخدام هذه الأداة في قياس شيء ما.

وعليه، يُستخدم القياس لتحديد **كميات خصائص الأشياء والفروق الكمية بينها ومدى هذه الفروق**، سواء في مجالات البيع والشراء أو في دراسة السلوك والقدرات والسمات المختلفة.

وتُعبَّر نتائج القياس باستثناء الحالات التي تتسم بانخفاض الدقة وتعتمد على التقديرات الذاتية في صورة **كميات عددية أو درجات** تُقابَل بخصائص الأشياء وفق قواعد محددة، وهو ما يفترض توافر **ثلاثة عناصر أساسية** في عملية القياس (علي، 2010، ص11).

ثالثاً: المفاهيم ذات الصلة (التقويم – التقييم - التقدير).

1- التقويم (Evaluation)**:**

إن الميدان الذي يستخدم فيه بصورة أساسية مصطلح التقويم Evaluation هو الميدان التربوي، وأول شيء نصطدم به عندما نراجع ما كتب عنه هو تعدد التعاريف المقدمة حوله. وهذا يدل على تنوع الممارسات واختلاف الآراء حول التعليم والتكوين. ففي ميدان تقويم التكوين المهني والأفراد الذين يتابعون التكوين يقدم باربيي (1983) Barbier التعريف التالي: "إنه فعل مقصود ومنظم اجتماعياً يؤدي إلى إنتاج حكم قيمة" إن باربييه يجعل من حكم القيمة Jugement de valeur الخاصية المميزة للتقويم.

وبالنسبة إلى ميدان التربية يوجد العديد من الكتابات إلا أنها في أغلبيتها تشير إلى أن نقطة البداية هي جمع المعلومات ونقطة النهاية هي اختيار قرار فعل بين العديد من البدائل، فمثلاً نجد "أستافل بيم وجماعته" (Stuffle Beam et al, 1980) يعتبر التقويم في التربية هو السيرورة التي من خلالها نحدد ونحصل ونقدم معلومات مفيدة تسمح بالحكم على قرارات محتملة" وفي نفس الاتجاه يذهب بلوم وجماعته (Bloom et al, 1971) إلى أن التقويم في رأيه هو الجمع المنظم للحقائق من أجل تحديد فيما إذا كانت بعض التغيرات تحدث فعلاً عند المتعلمين، وتحديد مقدار وكمية هذه التغيرات.

من خلال هذه التعاريف نلاحظ وجود قواسم مشتركة فيما بينها مثل: **جمع المعلومات والحكم واتخاذ القرار.** إن هذه المميزات تضبط موضوع التقويم في ميدان التربية والتكوين، وتحدد معالمه التي تميزه عن مواضيع البحث الأخرى المتصلة بالتربية والتكوين مثل اقتصاد التربية والتخطيط التربوي وبناء برامج التعليم والتكوين.

إن التركيز على سيرورة عملية التقويم كما هو واضح في التعاريف السابقة وليس على النتائج فقط، تبرز مدى تعقد هذا الموضوع وتعديه للتصور المبسط الذي يقلصه إلى مجرد امتحانات من أجل ترتيب الأفراد وتوزيعهم إلى فئتين (الناجحين والراسبين).

إن عملية التقويم تتضمن في البداية مقارنة بين حالتين (**ما هو كائن وما يجب أن يكون**) وتؤدي إلى اقتراح التعديلات التي يجب القيام بها، بمعنى آخر أنها تقيس الفارق الموجود بين النتائج المحققة والأهداف المسطرة كما أنها تكشف عن أسباب هذا الفارق وبالتالي تسمح لنا بتصور التصحيحات الملائمة (بوسنة، 2007، ص55).

وبعد هذا التحليل يمكننا أن نعرف التقويم على أنه يعني جمع وتقديم معلومات منظمة وموضوعية، كمية وكيفية حول التلميذ أو المتربص، بالاعتماد على العديد من التقنيات مثل الامتحانات والاختبارات والملاحظة والسجلات، وذلك من أجل:

- التعرف على مدى تحقق الأهداف المسطرة للعملية التربوية.
- الكشف عن طبيعة ومقدار التغير الذي حصل في مهارات التلميذ أو المتربص.
- تحديد مواطن القوة والضعف عند التلميذ أو المتربص.

ويكون الهدف الأساسي من التقويم هو العمل على تدعيم النتائج الإيجابية وتدارك النتائج السلبية عن طريق حصص علاجية مناسبة (بوسنه 2007).

2- التقييم (Assessment):

يعتبر مفهوم التقييم المقابل لمفهوم Assessment باللغة الإنجليزية من المفاهيم الواسعة الاستخدام في علم النفس وفي كثير من الأحيان يستخدم بشكل تبادلي وكأنه مرادف لمفهوم التقويم (مع العلم إننا نجد المصطلح Evaluation بالفرنسية يستخدم بصورة تبادلية بالنسبة إلى مفهومي التقييم والتقويم وتكون التفرقة من خلال محتوى النص أو أنه في أغلب الأحيان يرفق بالمصطلح الإنجليزي Assessment للتعبير على أن المقصود هو التقييم وليس التقويم.

استُخدم مصطلح التقييم أثناء الحرب العالمية الثانية من طرف علماء النفس الذين أوكلت إليهم مهمة انتقاء الأفراد الذين يصلحون للقيام بمهام خاصة فيما وراء البحار (مكتب الخدمات الإستراتيجية بالولايات المتحدة)، وقد ظهر هذا المفهوم لأول مرة في كتاب "تقييم الإنسان" الذي أصدره هذا المكتب (صلاح الدين محمود علام 2002). يعرف سندبرج Sundberg (1977) التقييم بأنه مجموعة من العمليات التي تستخدم بواسطة أخصائيين متمكنين، للوصول إلى وضع تصورات وانطباعات حول فرد معين واتخاذ قرارات واختبار فروض تتعلق بنمط خصائص سلوكه أو تفاعله مع بيئته.

إن التقييم يقترن عادة بالتصورات والنماذج النظرية المتصلة بالاختيار والتوظيف في ميدان علم نفس العمل من جهة ومن جهة أخرى بالمجالات النظرية العيادية، حيث يميل المختصون في علم

النفس العيادي إلى استخدام هذا المفهوم في فحوصهم لشخصية المصابين بأمراض نفسية (بوسنة، 2007، ص57).

يستخدم المختصون في عمليات التقييم أدوات قياس متعددة، الكثير منها يعتمد على القياس الكمي، مثل: اختبارات الذكاء والتحصيل والميول والاستعدادات والشخصية... والبعض منها يعتمد على التقديرات الكيفية والتحليلية، مثل: استمارة القبول ورسائل التزكية والمقابلات... وذلك بهدف اختيار الأفراد أو التشخيص العيادي وغير ذلك. والجدير بالملاحظة هو أنه في عمليات التقويم نستخدم نفس الأدوات لكن لتحقيق أهداف في النهاية مختلفة.

3- التقدير (Rating):

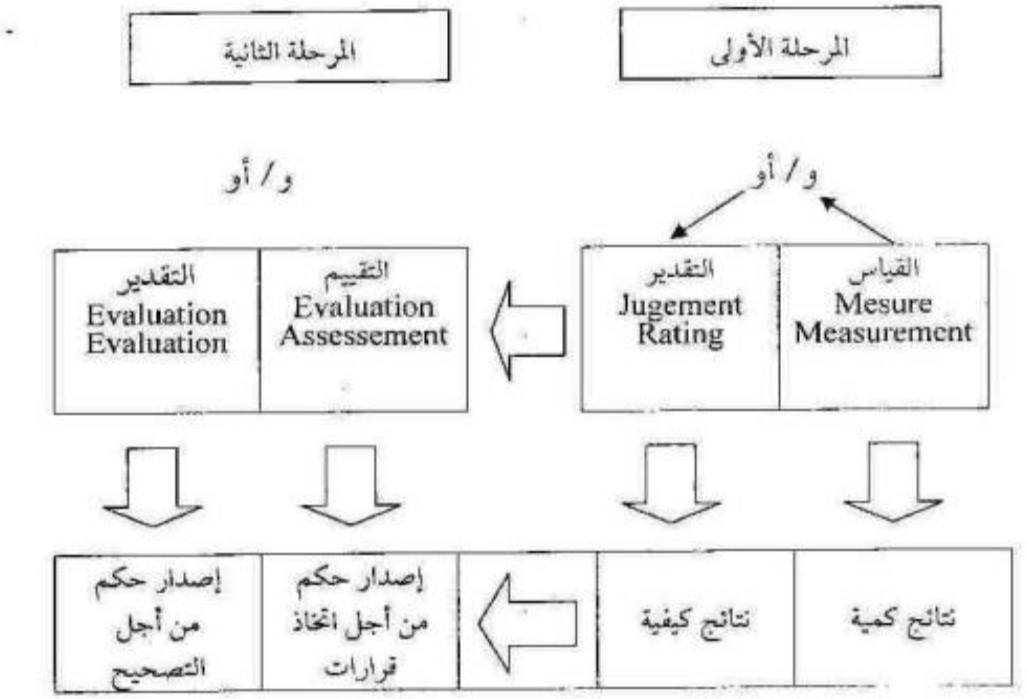
يعتبر مفهوم التقدير الذي يقابل المصطلح الإنجليزي Rating والمصطلح الفرنسي Jugement أقدم من مفهوم القياس من حيث الاستخدام من طرف الإنسان، حيث أن الإنسان بدأ منذ القديم يقوم بإعطاء تقديرات شخصية للأشياء فيما يخص مختلف المواضيع المحيطة به، ثم شرع في القياس فالتقييم وأخيراً التقويم.

إن المقصود بالتقدير هو التخمين الذاتي، فمن الناحية اللغوية نقول قدر الشيء أي بين مقداره. إن الأفراد يقومون بتقديرات متعددة يومياً بناءً على تخمينات شخصية وكيفية، مثلاً: أقل من، أكثر من، أكبر، أصغر، ضخم، هزيل، أكثر دقة، أقل دقة... إن هذه التقديرات التي تستخدم من طرف الأفراد هي تقديرات كيفية لا ترتقي إلى القياس الكمي. وتجدر الإشارة إلى أنه تم تطوير العديد من أدوات التقدير من أجل الاستخدام في مختلف الميادين التربوية والعيادية والمهنية، وهذا في الحالات التي لا يوجد فيها أدوات قياس مناسبة لجمع البيانات المراد الحصول عليها فيما يخص موضوع معين (بوسنة، 2007، ص57).

4- الفرق بين القياس والتقويم والتقييم والتقدير:

بناءً على التوضيحات التي قدمناها أعلاه فيما يخص كل مصطلح، يتبين لنا أنه يوجد من جهة القياس والتقدير ومن جهة أخرى نجد التقويم والتقييم. مع العلم أن نواتج المفهوم الأول **كمية** ونواتج المفهوم الثاني **كيفية**، أما نواتج التقييم والتقويم فإنها تتجاوز التحديد الكمي أو الكيفي للسلوك. وبالإضافة إلى ذلك فإنهما يعتمدان في مختلف إجراءاتهما على القياس و/أو التقدير، بمعنى أن القياس والتقدير يسبقان التقييم والتقويم أنظر الشكل رقم 01.

شكل رقم (01) يوضح العلاقة بين بعض المفاهيم لعلم النفس القياسي.



(بوسنة،2007،ص59)

يوضح الشكل السابق تسلسلًا منطقيًا لعملية القياس والتقويم عبر مرحلتين مترابطتين؛ حيث تبدأ العملية بالمرحلة الأولى التي يتم فيها جمع البيانات من خلال القياس (Measurement) والتقدير (Judgment/Rating)، إذ يعتمد القياس على أدوات موضوعية لإنتاج نتائج كمية في شكل درجات أو أرقام، بينما يعتمد التقدير على الأحكام الوصفية لإنتاج نتائج كيفية. وتمثل هذه النتائج (الكمية والكيفية) المدخلات الأساسية للمرحلة الثانية، التي تتضمن التقويم (Evaluation) والتقييم (Assessment)، حيث يتم في هذه المرحلة تحليل وتفسير البيانات المتحصل عليها. فالتقويم يُستخدم أساسًا لاتخاذ القرارات المناسبة بناءً على النتائج، في حين يركز التقييم على إصدار حكم نهائي يهدف إلى التصحيح والتحسين. وبذلك يتضح أن العملية تبدأ بجمع المعطيات ثم تنتقل إلى تفسيرها واستثمارها في اتخاذ القرارات أو تطوير الأداء.

المحـــاضرة الثانية

التطور التاريخي (حركة القياس النفسي)

الأهداف:

تهدف هذه المحاضرة إلى تعريف الطلبة ببدايات ظهور حركة القياس النفسي من خلال استعراض السياق التاريخي والعلمي الذي أسهم في نشأتها، مع التركيز على مفهوم المعادلة الشخصية ودوره في إبراز الفروق الفردية، إلى جانب التعرف على إسهامات السيكوفيزيقا في الربط بين المثيرات الفيزيائية والاستجابات النفسية. كما تسعى المحاضرة إلى توضيح العلاقة بين القياس النفسي والتأخر العقلي، من خلال إبراز جهود بينيه في قياس الذكاء وتطوير أولى الاختبارات العقلية. وتهدف كذلك إلى تتبع تطور القياس النفسي في إنجلترا، وما تميز به من توجهات، ثم عرض ملامح تطوره في أمريكا، مع بيان الإسهامات التي أسهمت في ترسيخه وتطبيقه في المجالات التربوية والنفسية.

محتوى المحاضرة:

أولاً: بدايات ظهور حركة القياس النفسي

1- مفهوم المعادلة الشخصية

2- السيكوفيزيقا

ثانيًا: القياس والتأخر العقلي

1- بينيه وقياس الذكـــاء

ثالثاً: القياس النفسي في إنجلترا

رابعًا: القياس النفسي في أمريكا

مدة المحاضرة: ساعة ونصف

تمهيد:

يعد القياس النفسي أحد الركائز الأساسية التي قام عليها علم النفس الحديث، إذ مكن الباحثين من الانتقال من الوصف التأملي للظواهر النفسية إلى دراستها دراسة علمية قائمة على الملاحظة والضبط والتكميم. وقد ارتبط تطور القياس النفسي ارتباطًا وثيقًا بتطور علم النفس نفسه، حيث ظهرت الحاجة إلى أدوات موضوعية تُمكّن من فهم الفروق بين الأفراد، وتفسير السلوك الإنساني، وتشخيص القدرات العقلية والاضطرابات النفسية.

ولم يظهر القياس النفسي دفعة واحدة، بل تبلور تدريجيًا عبر مراحل تاريخية متعاقبة، ساهمت فيها مدارس علمية مختلفة، وشخصيات بارزة في أوروبا وأمريكا. وقد ارتبطت بداياته الأولى بمحاولات قياس الإحساس والإدراك، ثم اتسع مجاله ليشمل تشخيص التأخر العقلي، ودراسة الذكاء، والفروق الفردية، إلى أن أصبح علمًا قائمًا بذاته يعتمد على أسس نفسية وإحصائية دقيقة. وانطلاقًا من هذا الإطار، تتناول هذه المحاضرة المحاور الآتية:

أولاً: بدايات ظهور حركة القياس النفسي، والظروف العلمية التي مهّدت لها.

ثانياً: إسهام القياس النفسي في دراسة التأخر العقلي والتمييز بين درجاته.

ثالثاً: تطور القياس النفسي في إنجلترا ودور الإحصاء والتحليل العاملي في دعمه.

رابعاً: ازدهار القياس النفسي في الولايات المتحدة الأمريكية، خاصة في مجال اختبارات الذكاء الفردية والجمعية.

ويهدف هذا العرض إلى إبراز المسار التاريخي لتطور القياس النفسي، وفهم الأسس التي قام عليها، تمهيدًا لاستيعاب تطبيقاته الحديثة في مجالات التشخيص، والإرشاد، والتربية، والبحث العلمي.

أولاً: بدايات ظهور حركة القياس النفسي:

ارتبط نشوء القياس النفسي ارتباطًا وثيقًا بظهور علم النفس ذاته، إذ سارا جنبًا إلى جنب منذ منتصف القرن التاسع عشر، تزامنًا مع أولى المحاولات الجادة لدراسة الظواهر النفسية دراسة علمية، تعتمد على الملاحظة الخارجية المضبوطة، وتبتعد عن التأمل الفلسفي والعقلي المجرد.

ويختلف الباحثون في تحديد النقطة الزمنية التي يمكن اعتبارها بداية فعلية لحركة القياس النفسي؛ فبعضهم يرجعها إلى مفهوم المعادلة الشخصية الذي قدّمه علماء الفلك، بوصفه أول إشارة علمية إلى الفروق الفردية في الاستجابة. بينما يرى آخرون أن بدايتها تعود إلى أعمال فيبر في ميدان

السيكوفيزيقا، ولا سيما ما قدمه عام 1878، إلى جانب صياغة قانون فيبر–فخنر. في حين يربط فريق ثالث نشأة القياس النفسي بأعمال فونت سنة 1879، وما تلاها من جهود تلامذته، وعلى رأسهم كاتل، فضلًا عن الإسهامات البارزة للعالم الإنجليزي جالتون.

1- مفهوم المعادلة الشخصية:

يقوم القياس النفسي في جوهره على مبدأ الفروق الفردية بين الأفراد. غير أن الانتباه الأول لهذه الفروق في الاستجابات للمثيرات لم يكن من قِبل علماء النفس، بل جاء على يد علماء الفلك.

ويعد الفلكي الإنجليزي كينيبروك (Kinneybrook) أول من تنبّه إلى وجود فروق بين الأفراد في ردود أفعالهم، وذلك سنة 1796 أثناء عمله في مرصد غرينتش (Greenwich). فقد لاحظ وجود اختلاف ثابت بين تقديراته الزمنية وتقديرات رئيسه ماسكلين (Maskelyne) فيما يتعلق بزمن مرور بعض الكواكب، إذ كان الفرق بين التقديرين يبلغ في البداية نحو ثُمن ثانية، ثم أخذ في الازدياد ليصل أحيانًا إلى ثانية كاملة .وقد عد هذا الأمر بالغ الخطورة، لأن الملاحظات الفلكية تُستخدم أساسًا في الاكتشافات العلمية وضبط الساعات. والمثير للانتباه أن هذا الفرق الزمني ظل ثابتًا في عدد كبير من التسجيلات التي قام بها الفلكيان. ونتيجة لذلك، شكّ رئيس مرصد غرينتش في أمانة كينيبروك العلمية، ما أدى إلى فصله من عمله.

غير أن هذه الواقعة لفتت انتباه الفلكي الألماني فريدريك فيلهلم بيسل (F. W. Bessel-1784-1846)، الذي كان يعمل في مرصد كونيغسبرغ (Konigsberg). وقد اطّلع بيسل على هذه الحادثة أثناء مراجعته لتاريخ الأبحاث التي نشرها مرصد غرينتش عام 1806، حيث كانت الواقعة مدونة ضمن سجلات سنة 1796. وبتأثير من العالم الرياضي والفلكي الشهير كارل فريدريك غاوس (Carl Friedrich Gauss-1777-1855)، بدأ بيسل في تحليل هذه الظاهرة تحليلاً علميًا رياضيًا.

وتوصل بيسل إلى مفهوم "المعادلة الشخصية"، التي تشير إلى الفروق القائمة بين تقدير فرد وآخر للظاهرة نفسها عند تسجيلها. ويُعزى هذا الفرق، الذي يُقاس بالثواني، إلى اختلاف الأفراد في زمن الرجوع (Reaction Time)، أي المدة الزمنية الفاصلة بين ظهور المثير (الكوكب) وإدراكه من قبل الفرد (الاستجابة) . وقد لاحظ بيسل هذه الظاهرة مرة أخرى أثناء تسجيله للفروق الزمنية بينه وبين مساعده فالبيك (Walbeck) عند رصد مرور أحد الكواكب عبر خط الزوال، فأطلق على هذا الفرق تسمية المعادلة الشخصية (Personal Equation)، بوصفها سمة تميز الفرد عن غيره من

الأفراد. ويعني ذلك أن اختلاف الأفراد في الزمن الذي يستغرقونه للاستجابة لمنبه معين يُعد خاصية فردية مميزة. وقد اكتشف بيسل هذه المعادلة سنة 1820، ونشر تقريرًا علميًا شهيرًا عنها عام 1822.

وبعد ذلك، أصبحت الفروق الفردية في زمن الرجوع مجالًا رئيسيًا للبحث في علم النفس، خاصة في أعمال كل من كاتل (Cattell) وجالتون (Galton)، مما أسهم في ترسيخ الأساس العلمي للقياس النفسي (معمرية، 2022، ص43).

2- السيكوفيزيقا: يعيش الإنسان في بيئة تزخر بمختلف أنواع المنبهات، ولا تقوم أجهزة الاستقبال الحسية لديه بإدراكها جميعًا، وإنما تنتقي بعضها بصورة انتقائية. ويستجيب الفرد لهذه المنبهات باستجابات مناسبة، فيرى ويسمع ويشم ويتذوق وغيرها من صور التفاعل الحسي .وتعد هذه الاستجابات عمليات نفسية تتزامن مع أحداث فيزيقية تقع في البيئة الخارجية. وعند دراسة الإحساس، فإن ما يتم الكشف عنه هو طبيعة العلاقة القائمة بين هذين البعدين: النفسي والفيزيقي.

ومن هذا المنطلق نشأ علم السيكوفيزيقا (Psychophysics) أو علم النفس الفيزيقي، وهو مصطلح يتكون من شقين: نفسي وفيزيقي (أو طبيعي). ويُعد هذا العلم من أقدم فروع علم النفس التجريبي، بل إنه المجال الذي انطلقت منه الدراسات التجريبية في علم النفس منذ نحو عام 1860. ويسعى علم السيكوفيزيقا إلى الإجابة عن السؤال الآتي: ما طبيعة العلاقة الدقيقة بين خصائص المنبه الفيزيقي، كما تقاس موضوعيًا، وبين خصائص الخبرة الحسية كما يحددها المفحوص؟

وقد شكلت دراسة العلاقة بين التغيرات التي تطرأ على التنبيه الفيزيقي والخبرة الحسية الناتجة عنها إحدى أقدم القضايا الأساسية في علم النفس. واستنادًا إلى ذلك، يمكن تعريف السيكوفيزيقا بأنها: دراسة العلاقة بين الخصائص الفيزيقية للمنبه والخصائص الكمية للإحساس، مثل العلاقة بين شدة التنبيه وقوة الإحساس به. كما تعرَف أيضًا بأنها: فرع من علم النفس يعنى بالقياس الذاتي؛ إذ تهتم النظرية السيكوفيزيقية بالأسس المنطقية للقياس الذاتي، في حين تعنى الطرائق السيكوفيزيقية بالأساليب التجريبية المستخدمة في هذا المجال. وقد بدأت الدراسات السيكوفيزيقية في إطار مشروع علمي ألماني قاده ثلاثة من الرواد البارزين، وهم (معمرية، 2022، ص44):

- **هيرمان فون هلمهولتز (H. V. Helmholtz) (1821 - 1894):** درس هلمهولتز بطريقة تجريبية العلاقة المنتظمة بين الجوانب القابلة للقياس في الطاقات الفيزيائية والخصائص القابلة للقياس في الخبرة الحسية لدى الفرد، ما ساهم في فهم أعمق للتفاعل بين العالم الفيزيائي والتجربة النفسية.

- **أرنست هاينريخ فيبر (E. H. Weber) (1795 - 1878):** يعتبر فيبر أول من قاس قدرة المفحوص على إدراك الفرق بين أزواج من المنبهات. بدأ دراساته بقياس أقل الفرق الملحوظ بين المنبهات باستخدام جهاز يُسمى الأستيزيومتر (Esthesiometer)، وهو فرجار دائري يقيس أقل مسافة يستطيع المفحوص عندها تمييز ما إذا تم لمس جلده بطرف واحد أو بطرفين من الفرجار.

قدم فيبر قانونه الشهير المعروف بـ "قانون العتبة الفارقة" أو "قانون الفرق الأدنى المدرك"، والذي ينص على أن: "يزداد أقل فرق يمكن إدراكه بين إثارتين بازدياد شدة المثير الأصلي، ويظل هذا الفرق نسبة ثابتة من شدة المثير".

ويُعبر عن القانون بالمعادلة:

$$\text{ثابت فيبر} = \frac{\text{العتبة الفارقة}}{\text{المنبه الأصلي}}$$

أي أن:

القيمة العتبية الفارقة = ثابت × مقدار المثير المعياري

مثال توضيحي: إذا كنت تحمل وزنًا مقداره 1000 غرام وأضفت 20 غرامًا، تشعر بالفرق. أما إذا كان الوزن 2000 غرام، فلا يمكنك إدراك الفرق إلا بإضافة 40 غرامًا، وعند 3000 غرام يجب إضافة 60 غرامًا لتلاحظ التغيير. ومن هنا نجد أن أقل فرق يمكن إدراكه يمثل دائمًا حوالي 2% من الوزن الأصلي، وهو ما يعرف بـ "ثابت فيبر"، وقد أشار إليه عام 1834.

مثال آخر: عرض على فرد بطاقة تحتوي على خط مستقيم طوله 10 سم، ثم عرضت بطاقات أخرى تحتوي على خطوط أطوالها 9.9 سم، 9.8 سم، 9.7 سم، وهكذا حتى 8 سم. لم يستطع الفرد تمييز الفرق بين الخطوط حتى وصل الطول إلى 8 سم. عند تكرار التجربة مع خط طوله 15 سم، بدأ الفرد يلاحظ الفرق عندما أصبح الفرق 3 سم، أما مع خط طوله 20 سم، فكان أقل فرق يدركه هو 4 سم.

نلاحظ هنا أن النسبة بين الفرق الملحوظ والطول الأصلي ثابتة تقريباً:

$$0.2 = \frac{4}{20} \quad ,0.2 = \frac{3}{15} \quad ,0.2 = \frac{2}{10}$$

وهذا يؤكد أن نسبة التغير في التنبيه إلى مقدار المثير الأصلي ثابتة، وهي 0.2، ولهذا يطلق على هذا القانون أحيانًا اسم "ثابت فيبر".

- **جوستاف تيودور فخنر (G. T. Fechner) (1801 - 1887):** يعتبر فخنر المؤسس الحقيقي لعلم السيكوفيزيقا. وقد انطلقت مشكلته الأساسية من التساؤل التالي: إذا كانت الإحساسات هي جوهر الشعور، وترتبط ارتباطًا مباشرًا بالأحداث الفيزيائية الناتجة عن المنبهات، فكيف يمكن قياس هذه الإحساسات نفسها؟ فبينما يمكننا قياس خصائص المنبهات مثل شدة ضوء الشمس أو طول قضيب أو وزن جسم، يظل قياس ما تثيره هذه المنبهات من إحساسات أمرًا غير مباشر وغير واضح.

انطلق فخنر من هذه المشكلة ليبحث عن طريقة لقياس الظواهر النفسية بشكل كمي، ورغب في ربط علم النفس بالعلوم الطبيعية بشكل أكثر منهجية. وبعد تسع سنوات من البحث والمناقشات، أصدر عام 1860 كتابه الشهير "أسس السيكوفيزيقا"، الذي يعتبره بعض مؤرخي علم النفس الانطلاقة الفعلية لعلم النفس التجريبي. في هذا الكتاب، قدم فخنر قانونًا رياضيًا يربط بين شدة المنبه وشدة الإحساس الناتج عنه، ويُعد هذا أول قانون في علم النفس يحدد العلاقة بين متغيرين نفسيين بشكل رياضي واضح. ويُعرف هذا القانون أحيانًا باسم "قانون فيبر-فخنر"، حيث يمثل تطويرًا وتوسيعًا لقانون فيبر السابق. وينص القانون على أن: شدة الإحساس تتناسب تناسبًا طرديًا مع لوغاريتم شدة المنبه.

أي أن:

$$\text{شدة الإحساس} = k \times \log(\text{شدة المنبه})$$

حيث (k) هو ثابت فيبر.

-أمثلة توضيحية:

- إذا كانت شدة المنبه الضوئي 100 شمعة، تكون شدة الإحساس:

$$9.20 = 4.60 \times 2 = \log 100 \times 2$$

- أما إذا كانت شدة المنبه 200 شمعة:

$$21.20 = 5.30 \times 4 = \log 200 \times 4$$

- وعندما تكون شدة المنبه 300 شمعة:

$$34.20 = 5.70 \times 6 = \log 300 \times 6$$

يتضح من هذه الأمثلة أن زيادة شدة المنبه تؤدي إلى زيادة في شدة الإحساس، ولكن بطريقة لوغاريتمية، حيث تتزايد شدة الإحساس بشكل أبطأ مقارنة بزيادة شدة المنبه.

- فيلهلم فونت (W. Wundt) (1832–1920): كان فونت من أبرز الداعين إلى قياس الظواهر النفسية المختلفة، إلا أنه لم يولِ اهتمامًا كافيًا لمشكلة الفروق الفردية، على الرغم من كونها الأساس الجوهري الذي يقوم عليه القياس النفسي بالمعنى الحديث. ورغم اعترافه بوجود متغيرات دخيلة قد تؤثر في الأداء، وإمكانية تقديرها من خلال الاستبطان وفي ظروف تجريبية محددة، فإن هذا الاعتراف لم يتطور إلى صياغة منهجية شاملة للمشكلة. ذلك أن الاستبطان، بوصفه المنهج الرئيس في التجريب النفسي آنذاك، يُعد في حقيقته الصورة الأولية لما أصبح يُعرف لاحقًا بـ الاختبارات العقلية، أو— بقدر أكبر من الدقة—الأساس الذي تطورت عنه أساليب مقاييس التقدير الذاتي (Self-Rating Scales) المستخدمة في قياس الشخصية. وعند الحديث عن فونت ومعمله، لا يمكن إغفال الإشارة إلى أحد أبرز تلامذته، الذي ينسب إليه وضع اللبنات الأولى للقياس النفسي، وهو:

- جيمس ماكين كاتل (J. Mc. Cattell) (1850–1944): تميزت أبحاث كاتل عن بقية تلامذة فونت؛ إذ حصل على درجة الدكتوراه تحت إشرافه في موضوع قياس الفروق الفردية في زمن الرجوع. ففي حين كان فونت يدرس الخبرة الشعورية ذاتيًا من خلال الاستبطان، اتجه كاتل إلى قياسها قياسًا موضوعيًا، معتمدًا على تحديد الزمن الفاصل بين تعرض المفحوص لمنبه معين وصدور الاستجابة المطلوبة منه. وقدم كاتل مثالًا على ذلك بإجراء تجارب تتطلب إرسال رسالة تلغرافية بأقصى سرعة ممكنة عقب التعرض لتنبيه محدد، مثل ظهور ضوء معين. وبهذه الطريقة وضع الأسس الأولى للدراسة العلمية لـ زمن الرجوع، باعتباره الفترة الزمنية الفاصلة بين التنبيه (ظهور الضوء) والاستجابة (إرسال الرسالة)، معتمدًا في ذلك على قياس موضوعي دقيق.

ومن خلال هذه الجهود، حقق كاتل هدفين أساسيين: أولهما قياس سرعة العمليات الإدراكية عبر مستويات مختلفة من التعقيد، وثانيهما الانتقال من الصياغات العامة للقوانين السلوكية إلى التحديد الكمي للفروق الفردية. وقد واصل كاتل أبحاثه في هذا الاتجاه، ولا سيما دراسة الفروق الفردية في زمن الرجوع، بعد عودته إلى الولايات المتحدة الأمريكية.

- فرنسيس جالتون 1822 - 1911(F. Galton) : لكن ورغم العمل العلمي الموضوعي الذي قام به كاتل، إلا أن الإنجليزي ف. جالتون يعتبر المكتشف الحقيقي لمجال الفروق الفردية، واتجه باهتمامه نحو دراستها من خلال المقاييس التي وضعها، إلا أنه كان مهتماً بالفروق الفردية في المجال البيولوجي وليس في المجال السيكولوجي.

وذكر ذلك في مقدمة كتابه الذي أصدره عام 1883 بعنوان **"دراسة الملكات الإنسانية"**، حيث يشير إلى هدفه العام وهو معرفة الخصائص الوراثية عند البشر، ومعرفة الفروق الأساسية بين السلالات والعائلات المختلفة، وذلك من أجل معرفة مدى إمكان استئصال أو تغيير الاستعدادات المتدنية لبعض الأفراد، وكذلك تحديد ما إذا كانت إجراءات الاستئصال هذه ممكنة ومعقولة، بحيث يمكن تجنيب الأجيال القادمة مثل هذه الحالات من الاستعدادات المتدنية لبعض أفرادها. وهذا الاهتمام من جالتون يندرج ضمن السعي إلى **تحسين النسل.** وحاول جالتون قياس الذكاء، أو هو أول من فكر بجدية في قياس الذكاء عندما أنشأ معملاً صغيراً في متحف لندن لقياس قدرات الإنسان، أسماه **المعمل الأنثروبومتري (Anthropometric)**، كما قضى شهوراً في معرض كنسنجتون (Kensington) قاس خلالها الخصائص البدنية لجمهور المترددين.

وافترض أن القدرات العقلية والقدرات الحسية الإدراكية مرتبطتان ارتباطاً وثيقاً، وأن المتأخرين عقلياً تنقصهم حدة الإحساس. وقام بقياس حدة السمع والإبصار، وإدراك الألوان، وزمن الرجع، والتمييز اللمسي، والتمييز بين الأوزان، والأنشطة الحركية (كالجانب والضغط)، وقوة النفخ. لكن البحوث العلمية دلت بعد ذلك على أن العلاقة بين الجوانب الحسية الإدراكية والذكاء ضعيفة.

ورغم عبقرية جالتون وتأثيره في كل من فندت وكاتل، إلا أن تأثيره في القياس النفسي لم يحدث، وبالأخص في إنجلترا التي عُرف فيها أكثر من أي مكان آخر. إلا بعد ظهور كارل بيرسون صاحب معامل الارتباط لضرب العزوم الشهير الذي طبق أساليب جالتون الإحصائية. لكن دراسات جالتون لفتت الأنظار إلى أهمية القياس النفسي.

- هرمان إبنجهاوس H. Ebbinghaus 1850 - 1909: ولا ننسى أن نشير إلى أن دراسة الألماني إبنجهاوس لقياس الذاكرة سنة 1885، كانت أرقى من المحاولتين السابقتين، لأنها تقيس عملية عقلية عليا موضوعياً ولا تقيس العمليات الحسية.

ثانيًا: القياس والتأخر العقلي:

يعد مجال التأخر العقلي من أبرز الميادين التي ظهرت من خلالها بوضوح إشكالية القياس، ولا سيما فيما يتعلق بتمييز الأفراد بعضهم عن بعض. وقد بدأ الاهتمام العلمي بهذه المشكلة في وقت مبكر، خلال النصف الأول من القرن التاسع عشر. وكان من أهم الإسهامات في هذا المجال ما قدّمه الطبيب الفرنسي جان إسكيرول (J. Esquirol) (1772–1840)، الذي وضع تمييزًا واضحًا بين التأخر العقلي والمرض العقلي، وهما فئتان كانتا مختلطتين وغير محددتي المعالم حتى ذلك الحين.

فقد أوضح إسكيرول أن التأخر العقلي حالة يولد بها الفرد وتستمر معه طوال حياته، ويتميز صاحبها بانخفاض ملحوظ في مستوى القدرات العقلية. أما المرض العقلي، فإنه يصيب الفرد في مرحلة لاحقة من عمره، وقد يصاحبه في بدايته تدهور في الأداء العقلي. واعتمد إسكيرول على اللغة بوصفها مؤشرًا لتحديد المستوى العقلي للفرد، ممتدًا من الحالة السوية إلى أشد درجات التأخر العقلي. ولاحظ أن القدرة على النطق تكون سليمة نسبيًا في حالات التأخر العقلي البسيط، لكنها تضعف تدريجيًا كلما ازدادت شدة الضعف العقلي، إلى أن تختفي المقاطع اللفظية للكلمات تقريبًا في حالات التأخر العقلي الشديد. كما أدخل إسكيرول محكات موضوعية للتمييز بين درجات التأخر العقلي، تمثلت في استخدام بعض المقاييس الفيزيقية، مثل قياس حجم الجمجمة ونسبتها إلى بقية أجزاء الجسم.

إلى جانب ذلك، قدّم الطبيب الفرنسي إدوارد سيغان (1812–1880) (E. Seguin)، تلميذ جان مارك غاسبار إيتارد (1774–1838) (J. M. G. Itard)، إسهامات مهمة في مجال قياس الضعف العقلي. فقد ابتكر لوحة الأشكال المعروفة باسمه (Seguin Form Board)، وهي أداة لا تزال مستخدمة حتى اليوم ضمن بعض مقاييس الذكاء.

1- بينيه وقياس الذكـــاء:

ظل ألفريد بينيه (A. Binet) الذي وُلد في نيس عام 1857 وتوفي في باريس عام 1911 خلال العشرية الأخيرة من القرن التاسع عشر، منشغلاً بمحاولة إيجاد وسيلة مناسبة لقياس الذكاء. حصل على درجة الدكتوراه في العلوم عام 1894 بعنوان: "مساهمة في دراسة الجهاز العصبي تحت المعوي للحشرات". وفي عام 1900 قدّم "الدراسة التجريبية للذكاء". وفي عام 1904 عينه وزير المعارف (التربية) الفرنسي عضواً في لجنة الأقسام الدراسية الخاصة، وكانت مهمة اللجنة أن تجد الوسيلة المناسبة لعزل غير المستفيدين من البرامج التعليمية المدرسية، مع التمييز بين فئتي "المتأخرين عقلياً" و "البلداء".

وأخذ بينيه على عاتقه هذه المهمة بمساعدة تيودور سيمون (Th. Simon 1873 - 1961) الطبيب بمستشفى بيراي - فوكلوز للأمراض العقلية. وكانت المهمة عملية، إذ تتحدد فقط في حل مشكلة تعليمية، لذا لم يكن بيني معنياً بإعطاء تعريف نظري للذكاء، رغم أنه كان منشغلاً به. ونتيجة لإدراكه أن القدرة العقلية تنمو مع الطفل بقدر نضجه، قدم عام 1905 مجموعة الاختبارات التي تقيس الذكاء، متدرجة من الأكثر سهولة إلى الأكثر صعوبة. وتضمنت الاختبارات أعمالاً مثل: تسمية الأشياء؛

مقارنة أطوال؛ تكرار رموز معينة؛ تكملة جمل وفهم أسئلة. نشرها في الحولية النفسية (L' Année psychologique, II, 1905, 191 - 244).

ثالثاً: القياس النفسي في إنجلترا:

في إنجلترا، كان لـ فرانسيس جالتون تأثير بالغ في تلميذه الإحصائي الإنجليزي الشهير كارل بيرسون (K. Pearson) (1857–1936)، الذي أسهم بدور محوري في دفع حركة القياس النفسي قدمًا. فقد قدّم بيرسون إسهامات إحصائية واسعة في مجالي علم النفس والبيولوجيا، من أبرزها تطوير مفهوم الانحراف المعياري بوصفه مقياسًا للتشتت، وصياغة اختبار كاي تربيع (Chi-Square) لقياس حسن المطابقة، إلى جانب ابتكاره معامل الارتباط الخطي المعروف باسمه. ولم يكن هدف بيرسون مقتصرًا على اشتقاق معادلات إحصائية مجردة، بل انصب اهتمامه أساسًا على تطوير أساليب إحصائية عملية تساعد في معالجة البيانات والكشف عن العلاقات الارتباطية، ولا سيما في مجال علم النفس. ورغم أنه لم يسهم بصورة مباشرة في بناء أدوات القياس النفسي، فإن ما قدمه من صيغ ومعادلات إحصائية شكّل أساسًا اعتمد عليه الباحثون في القياس النفسي لاحقًا.

أما تشارلز إدوارد سبيرمان (C. E. Spearman) (1863–1945)، فقد واصل النهج الإحصائي الذي أسسه بيرسون، ويُعد من رواد التحليل العاملي .تلقّى سبيرمان تكوينه النفسي على يد فونت، ثم تنقّل بين عدد من الجامعات الألمانية العريقة، مثل فورزبورغ وبرلين وغوتنغن .وخلال إقامته في ألمانيا، نشر دراسة مهمة حول القياس الموضوعي للذكاء العام، أشار فيها إلى التحليل العاملي بوصفه أسلوبًا إحصائيًا مناسبًا لهذا الغرض. وتتمثل أبرز إسهامات سبيرمان في توظيف التحليل العاملي في قياس الذكاء، وصياغته من خلاله نظرية العاملين في الذكاء، التي تفترض وجود عامل عقلي عام (g)يسهم في جميع الأنشطة العقلية، إلى جانب عوامل نوعية يقتصر تأثير كل منها على مجال معرفي محدد. وقد عرض هذه النظرية في كتابه الشهير »قدرات الإنسان: طبيعتها وقياسها "الصادر عام 1927". وقد واصل هذا الاتجاه في إنجلترا عدد من الباحثين، من أبرزهم ج. هـ. طومبسون (G. H. Thomson) وسيريل لودويك بيرت (C. L. Burt) (1883–1971) وهانز يورغن إيزنك (H. J. Eysenck) (1916– …)، حيث توسع استخدام التحليل العاملي ليشمل، إلى جانب الذكاء، قياس مختلف المتغيرات النفسية. ويُعد توظيف إيزنك للتحليل العاملي في بناء نظرية الأبعاد في الشخصية مثالًا بارزًا على هذا التطور.

رابعًا: القياس النفسي في أمريكا:

عند عودة جيمس ماكين كاتل إلى الولايات المتحدة، اتجه بوضوح إلى دراسة الفروق الفردية، وبدأ منذ عام 1894 بتطبيق مجموعة من الاختبارات النفسية على طلبة جامعة كولومبيا. ويُعد كاتل أول من صاغ مصطلح الاختبارات العقلية (Mental Tests) عام 1890، حيث استخدم بطارية من الاختبارات لقياس عمليات عقلية وحسية–حركية بسيطة، كان من أبرزها زمن الرجع، الذي مثّل محورًا أساسيًا في دراسة الفروق الفردية. وأظهرت نتائج هذه الدراسات الحاجة إلى توظيف الأساليب الإحصائية في تحليل البيانات، كما رسخت فكرة أن الفروق بين الأفراد في القدرات العقلية هي فروق كمية في الدرجة لا نوعية في الكيف. وقد أسهمت جهود كاتل في إعطاء دفعة قوية لحركة القياس النفسي في أمريكا، وهو ما انعكس في اهتمام جمعية علم النفس الأمريكية بدراسة الفروق الفردية منذ عام 1895.

وبدأ انتشار اختبارات الذكاء على نمط بينيه في أمريكا بعد ترجمة هنري جودارد لاختبار بيني عام 1910، ثم تلتها محاولات متعددة كان أبرزها عمل لويس ترمان بجامعة ستانفورد، الذي أعاد تقنين وتعديل اختبار بينيه على عينات أمريكية واسعة، ونشره عام 1916 تحت اسم اختبار ستانفورد–بينيه للذكاء، ليصبح من أشهر الاختبارات الفردية وأكثرها استخدامًا عالميًا، مع خضوعه لاحقًا لعدة مراجعات وتعديلات. ومع تزايد الحاجة إلى قياس ذكاء أعداد كبيرة من الأفراد، خاصة في المدارس والجيش، ظهر اتجاه جديد نحو الاختبارات الجمعية، وتعزز هذا الاتجاه بشكل كبير خلال الحرب العالمية الأولى، حيث أُعدّ اختبارا ألفا وبيتا العسكريان عام 1917 بقيادة روبرت يركس، مما أدى إلى توسع غير مسبوق في القياس النفسي. وفي عام 1939 قدّم ديفيد وكسلر اختبار وكسلر–بلفيو لقياس ذكاء الراشدين، متجاوزًا بعض قيود اختبار ستانفورد–بينيه، ومميزًا بين الاختبارات اللفظية والأدائية، مع استخدامات واسعة في المجال العيادي.

كما شهدت أمريكا تطورًا مهمًا في التحليل العاملي مع أعمال لويس ليون ثرستون، الذي رفض فكرة العامل العام الواحد، واقترح أن الذكاء يتكون من قدرات عقلية أولية متعددة، تجسدت في بطاريته الشهيرة للقدرات العقلية الأولية عام 1941.

وهكذا شكّلت الولايات المتحدة مركزًا رئيسيًا لتطور القياس النفسي، سواء في اختبارات الذكاء الفردية والجمعية أو في النماذج النظرية والإحصائية التي أسهمت في بناء هذا المجال وتوسيعه.

المحـــاضرة الثالثة

مسلمات القياس النفسي وخصائصه

الأهداف:

تهدف هذه المحاضرة إلى تعريف الطلبة بالمسلمات الأساسية التي تقوم عليها نظريات القياس في علم النفس، من خلال توضيح مبدأ قابلية أداء الفرد للقياس وإمكانية تمثيله كمياً، إلى جانب إبراز أن أداء الأفراد يتسم بخصائص مميزة تعكس الفروق الفردية بينهم. كما تسعى إلى توضيح أن العلاقة بين الخصائص النفسية والأداء ليست ثابتة، بل تختلف من فرد لآخر، وهو ما يستدعي الحذر في تفسير النتائج. إضافة إلى ذلك، تهدف المحاضرة إلى ترسيخ فهم أن الخطأ يُعد مكونًا أساسياً في عملية القياس، ولا يمكن فصله عنها، مما يتطلب التعامل معه علمياً عند تفسير الدرجات واتخاذ القرارات.

محتوى المحاضرة:

أولاً- مسلمات تستند إليها نظريات القياس في علم النفس

1- قابلية أداء الفرد للقياس

2- أداء الأفراد له خصائص مميزة

3- اختلاف العلاقة بين الخصائص والأداء من فرد لآخر

4-الخطأ مكون أساسي من مكونات عملية القياس

مدة المحاضرة: ساعة ونصف

تمهيد:

يعد القياس النفسي أحد الركائز الأساسية في ميادين علم النفس والتربية، إذ يمكن الباحثين والممارسين من تقدير الفروق الفردية بين الأشخاص في السمات والقدرات والاتجاهات والدوافع وغيرها من الظواهر النفسية غير المرئية بشكل مباشر. إلا أن القياس في هذا المجال يختلف جذرياً عن القياس في العلوم الطبيعية، لأنه يتناول خصائص إنسانية معقدة ومتداخلة يصعب فصلها أو عزلها عن بعضها البعض. ولذلك ظهرت نظريات القياس النفسي لتضع أسساً ومبادئ علمية تضمن دقة النتائج وصدقها، وتوضح كيفية بناء الأدوات التي تقيس السمات النفسية بطريقة موضوعية وقابلة للتفسير. وتستند هذه النظريات إلى مجموعة من المسلمات الأساسية التي تمثل الإطار الفلسفي والمنهجي لعملية القياس النفسي، وهي بمثابة القواعد التي تحدد ما يمكن قياسه وكيفية تفسير الدرجات المتحصل عليها. وفي هذه المحاضرة سنتناول بالتفصيل أبرز المسلمات التي يقوم عليها القياس النفسي، والمتمثلة فيما يلي:

1- قابلية أداء الفرد للقياس أي إمكانية قياس السلوك الإنساني كمياً من خلال مؤشرات يمكن ملاحظتها.

2- أداء الأفراد له خصائص مميزة أي أن سلوك الفرد يعكس خصائصه النفسية المميزة له عن غيره.

3- اختلاف العلاقة بين الخصائص والأداء من فرد لآخر مما يشير إلى وجود الفروق الفردية كأساس لكل عملية قياس نفسي.

4- الخطأ مكون أساسي من مكونات عملية القياس باعتبار أن أي نتيجة كمية تتضمن قدراً من الخطأ لا يمكن إلغاؤه تماماً وإنما يمكن التقليل من أثره.

من خلال دراسة هذه المسلمات سيتضح لنا المنطق الذي تستند إليه أدوات القياس النفسي وكيف تؤثر هذه المسلمات في بناء الاختبارات والمقاييس وتفسير نتائجها، بما يضمن فهماً أعمق لطبيعة السلوك الإنساني وأساليب تقديره بدقة وموضوعية.

أولاً- مسلمات تستند إليها نظريات القياس في علم النفس:

تهدف النظرية العلمية بوجه عام إلى تفسير مجموعة من السمات أو السلوكيات ضمن إطار منظم، استنادًا إلى مجموعة من الفروض أو المسلّمات التي تنطلق منها لتقديم تفسير علمي لتلك السمات. ويُشترط في أي نظرية أن تمتلك قدرة تفسيرية وتعليلية تؤهلها لأن تكون صالحة للتطبيق الميداني ومحققة للفائدة العلمية المرجوة منها. وبالمثل فإن نظريات القياس في علم النفس شأنها شأن

نظريات العلوم الأخرى تعتمد في دراستها للظواهر النفسيـة على مجموعة من المسلمات الأساسية التي ينبغي الإلمام بها قبل الشروع في أي عملية قيـاس.

1- قابليـة أداء الفرد للقيـاس:

من المعلوم أن لكل فرد مجموعة من الأنماط السلوكية التي تختلف بدرجات معينة عن أنماط سلوك الأفراد الآخرين. ويُطلق على هذه الأنماط في ميدان القياس النفسي مصطلح "الأداء" كما أشار إلى ذلك مصطفى كامل (1999). وتجسد هذه الفكرة المضمون الأساسي للمسلمة الأولى في نظريات القياس النفسي، والتي تنص على أن السلوك الإنساني المميز لكل فرد قابل للقياس والتقدير بطريقة كمية .أي أن أداء الفرد يمكن تحويله من وصف نوعي أو كيفي إلى تعبير كمي رقمي باستخدام أدوات وأساليب قياس محددة، ووفق قواعد ومعايير قياسية متفق عليها تضمن موضوعية التقدير ودقته.

تشير المسلمة الأولى من مسلمات نظريات القياس في علم النفس إلى أن أداء الفرد قابل للقياس والتقدير الكمي، وهو ما يُعدّ الأساس الذي تُبنى عليه مختلف العمليات اللاحقة في مجال القياس النفسي. فلو لم نُقر منذ البداية بإمكانية التعبير الكمي عن أداء الأفراد لما أمكننا الحديث عن أي من العمليات القياسية الأخرى. وعندما يقاس أداء الفرد أو يُقدّر في مرحلة معينة يصبح من الممكن لاحقًا التنبؤ بمراحل أخرى من أدائه أو بردود أفعاله المستقبلية (مصري عبد الحميد حنورة، 1998).

إن القول بقابلية أداء الفرد للقياس والتقدير الكمي يعني ضمنيًا إمكانية إخضاع هذا الأداء لتأثير العوامل والظروف الخارجية، لا سيما تلك التي تحدث في المواقف التجريبية والتي قد تؤثر بدرجات متفاوتة في عملية القياس والتقدير. ومع ذلك يعد من الصعب في معظم الحالات عزل الأداء المراد قياسه عن باقي مكونات شخصية الفرد، إذ تمثل الشخصية كلاً متكاملاً ويُعدّ فصل أحد أجزائها لدراسته بمعزل عن السمات الأخرى عملية معقدة ومحدودة الدقة. فإذا كان الهدف من القياس هو تقدير أداء الفرد في مواقف تتطلب حلّ مشكلات رياضية أو تركيب أشكال هندسية، فقد يصعب في كثير من الأحيان عزل هذا الأداء عن جوانب أخرى من أدائه مثل قدرته على التعبير اللغوي أو استخدام الرموز أو معالجة الأشكال الهندسية. وبالتالي فإن النتائج المتحصل عليها قد تتأثر جزئيًا بهذه العوامل المتداخلة. ومن هنا يتضح أن قياس سلوك الأفراد ووضعهم في مواقف تجريبية يتطلب مراعاة طبيعة التداخل والتفاعل بين مكونات الأداء الإنساني المختلفة، إذ ترتبط هذه الجوانب بعلاقات متبادلة من التأثير والتأثر .وبناءً على ذلك يجب أن تُصمَّم أدوات القياس أو التقدير – كالاختبارات

في أغلب الحالات – بطريقة تراعي هذا التفاعل بحيث تكون البنود الموضوعة قادرة على تمثيل الأداء المستهدف بدقة مع تقليل تأثير العوامل الخارجية أو غير المقصودة قدر الإمكان.

تختلف القياسات النفسية عن القياسات الطبيعية من حيث طبيعة الظاهرة المقاسة وطريقة التعامل معها. ففي القياس الطبيعي مثل قياس الأطوال أو الأوزان لا يتأثر القياس بعوامل جانبية؛ فقياس طول شخصٍ ما لا يتأثر بوزنه أو بلونه، وقياس وزن قطعة من الحديد يبقى ثابتًا سواء كانت على شكل كرة أو مكعب، إذ يمكن عزل كل خاصية عن الأخرى وقياسها بدقة دون تدخل من العوامل المحيطة مما يضمن موضوعية عملية القياس واستقرار نتائجها. أما في القياس النفسي فالأمر أكثر تعقيدًا، إذ تشير المسلمة الأولى من مسلمات نظريات القياس النفسي إلى أن أداء الإنسان قابل للقياس والتقدير الكمي، غير أن تحقيق ذلك يتطلب استخدام أدوات قياس من نوع خاص تراعي خصوصية الأداء الإنساني وتعقيداته. لذلك تختلف أدوات القياس النفسي فيما بينها باختلاف السمة أو الظاهرة المقاسة؛ فأداة قياس الذكاء، على سبيل المثال، لا تصلح لقياس الدافعية، كما أنها تختلف جذريًا عن الأدوات المستخدمة في القياسات الطبيعية الدقيقة. وتتميز أدوات القياس النفسي بامتلاكها منطقًا رياضيًا وإحصائيًا خاصًا، إضافةً إلى مفاهيم وإجراءات منهجية تهدف إلى تقدير الأداء الإنساني بدقة قدر الإمكان رغم الطبيعة المركّبة والمتغيرة للسلوك البشري.

2- أداء الأفراد له خصائص مميزة:

تشير المسلمة الثانية من مسلمات نظرية القياس النفسي إلى أن أداء الفرد يُعد دالة لخصائصه الشخصية، أي أن كل سلوك أو أداء يصدر عن الفرد يعكس خاصية معينة أو مجموعة من الخصائص التي تميّزه عن غيره من الأفراد. فمثلاً تمثل خاصية كالذكاء مصدرًا لعدة أنماط من الأداء؛ فالفرد ذو الذكاء العالي قد يُظهر سرعة في حل المسائل الرياضية أو قدرة على إيجاد العلاقات بين المفاهيم والأشياء بطريقة مبتكرة أو مهارة في التفكير الإبداعي — وجميعها صور مختلفة تعبّر عن خاصية واحدة هي الذكاء. وفي المقابل فإن الأداء الواحد قد يكون ناتجًا عن عدة خصائص مختلفة .فعلى سبيل المثال عدم مشاركة التلميذ في الأنشطة الصفية قد يُعزى إلى أسباب متنوعة؛ فقد يكون نتيجة مرض عضوي يخشى التلميذ أن يكتشفه زملاؤه، أو بسبب الرهاب الاجتماعي (الفوبيا) الذي يعاني منه بعض الأفراد أو نتيجة تجربة سلبية سابقة في العمل الجماعي. وهكذا فإن الأداء الإنساني لا يُفهم بمعزل عن خصائص الفرد النفسية والجسمية والاجتماعية، إذ تتداخل هذه الخصائص لتنتج مؤشرات سلوكية متعددة تعبّر عن تكوين الشخصية ككل.

يتضح من ذلك أن العلاقة بين الخصائص والأداء الإنساني علاقة معقدة ومتشابكة، وهو ما ينعكس بطبيعة الحال على أدوات القياس المستخدمة من حيث مراحل بنائها وطرق اختبار المؤشرات الدالة على السمة المستهدفة دون غيرها من الظواهر. كما يؤثر هذا التعقيد في دلالة النتائج وتفسيرها خصوصًا فيما يتعلق بدرجة صدقها في تمثيل السمة المقاسة؛ أي كيف يمكن للنتيجة الكمية المتحصل عليها أن تعبّر بدقة وصدق عن السمة الحقيقية التي يفترض قياسها. فعلى سبيل المثال عند إعداد أداة لقياس التحصيل الدراسي يجب إدراك أن الأداء المقاس لا يعكس خاصية التحصيل وحدها بل هو نتاج تفاعل مجموعة من الخصائص مثل الذكاء والثروة اللغوية والقدرة على التركيز، وغيرها من العوامل المؤثرة. لذا من الضروري أخذ هذه العوامل بعين الاعتبار عند تحليل دلالة بنود الاختبار وعند تفسير النتائج لضمان أن يكون التقدير الكمي ناتجًا عن السمة المستهدفة فعلاً وليس عن خصائص أو متغيرات أخرى متداخلة. إن الخصوصية المميزة للعلاقة بين الخصائص والأداء جعلت أدوات القياس النفسي والتربوي تتأثر بطبيعة هذه العلاقة من حيث بناء الأداة وتكوينها .وقد أشار عبد الرحمن سليمان الطريري (1998) إلى ضرورة إضافة بعد آخر إلى دراسة هذه العلاقة يتمثل في مدى شدة الارتباط بين الخاصية والأداء.

فعلى سبيل المثال إذا كانت الخاصية المقاسة هي العدوانية وكان الأداء الملاحظ هو الاعتداء المادي على الأقران في المدرسة فإن أداة القياس ينبغي أن تكون حساسة بدرجة كافية لشدة العلاقة بين الخاصية والأداء، بحيث يمكنها قياس هذا السلوك وإرجاعه بدقة إلى السمة المقصودة، سواء كانت خاصية واحدة أو مجموعة من الخصائص. وبمعنى آخر يجب أن تسمح الأداة بتقدير العلاقة بين المتغيرين (الخاصية والأداء) دون أن تتأثر بوجود متغيرات دخيلة أو طارئة. ففي المثال ذاته إذا كانت الأداة قادرة على قياس العلاقة بين العدوانية والسلوك الاعتدائي بدقة، فإنها لن تتأثر بتدخل عوامل أخرى كسوء المعاملة أو الشعور بالاحتكار أو ضعف المشاركة في الأنشطة، وهي عوامل قد تتداخل أحيانًا مع مؤشرات العدوانية الحقيقية.

وبناءً على ذلك تؤكد المسلمة الثانية من مسلمات القياس النفسي التي تفيد بأن أداء الإنسان يعكس خصائص معينة ويدل عليها على بُعدين أساسيين في عملية القياس، هما:

1- علاقة الخاصية بالأداء من حيث النوع والكم، أي مدى صدق الأداة في قياس السمة التي وُضعت لقياسها ودقة هذا القياس.

2- مدى تأثر أداة القياس بطبيعة العلاقة بين الخاصية والأداء، أي قدرتها على تفسير سلوك الفرد وإرجاعه فعلاً إلى السمة المقصودة دون انحرافات ناجمة عن عوامل خارجية.

وهكذا فإن الاختبارات والمقاييس النفسية لا تقتصر على قياس أداء الفرد فقط بل تمتد لتقدير شدة العلاقة بين الخصائص والأداء، مما يمنح نتائج القياس عمقًا تفسيرياً وصدقًا أعلى في التمثيل.

3- اختلاف العلاقة بين الخصائص والأداء من فرد لآخر:

تشير المسلمة الثالثة من مسلمات نظرية القياس النفسي إلى جوهر عملية القياس في ميادين علم النفس والتربية والمتمثل في مفهوم الفروق الفردية .فالعلاقة بين الخصائص والأداء ليست ثابتة بل تختلف من فردٍ إلى آخر. وقد تناول فرانسيس غالتون (F. Galton) هذا المبدأ في دراساته الميدانية، حيث ركز على الفروق بين الأفراد أكثر من أوجه التشابه بينهم معتبرًا أن التنوع الإنساني هو الأساس الذي يُبنى عليه القياس النفسي. فعندما تقوم الاختبارات النفسية بقياس أداء فردٍ معين، فإنها في الواقع لا تقيس كمية الأداء كما نقيس الطول أو الوزن في القياسات الطبيعية، وكذلك عند قياس خاصية أو قدرة معينة – مثل الذكاء – فإن الأداة لا تقيس كمية الذكاء بمفهومها الكمي المطلق. وكما أشار حسين بشير محمود(2001) ، فإن قياس العلاقة بين الخاصية والأداء لا يتم بوحدات مطلقة، بل يتم ضمن إطار نسبي قائم على الاختلافات والتباينات بين الأفراد في خصائصهم وأدائهم الفعلي.

ومن ثم فإن ما يقاس فعليًا في الاختبارات النفسية هو الاختلاف بين الأفراد في سماتهم العقلية والشخصية والقدرات المختلفة، وليس القيم المطلقة لتلك السمات. فعملية القياس هنا نسبية بطبيعتها، لا تسعى إلى إيجاد قانون عام موحّد كما هو الحال في القياسات الطبيعية، بل تهدف إلى الكشف عن الفروق الفردية وتحديد دلالتها الإحصائية. وعليه فإن الإقرار بوجود الفروق الفردية يُعدّ أساسًا جوهريًا في القياس النفسي وينعكس على طريقة التعامل مع النتائج العددية الناتجة عن عملية القياس، حيث تُفسَّر هذه الأرقام على أنها مؤشرات للعلاقة بين الأداء والخاصية، لا كقيم مطلقة لكمية السمة نفسها. فبينما تسعى القياسات الطبيعية إلى الوصول إلى قوانين عامة، يركّز القياس النفسي على تحليل التباين بين الأفراد وفهم دلالاته العلمية.

وعليه فإن الاختلاف بين القياس في الميادين النفسية والتربوية والقياس في الميادين الطبيعية يظهر بوضوح في الهدف والأسلوب معًا. فمفهوم التباين أو الفروق الفردية يلعب دورًا حاسمًا في تصميم أدوات القياس واختيار وحداتها، وفي التأكد من مدى فعاليتها في تقديم قياس موضوعي للسمات أو الظواهر النفسية والتربوية. إذ تختلف الأداة المصممة لقياس الفروق الفردية عن تلك التي

تهدف إلى قياس الكميات المطلقة، كما أشار إلى ذلك صلاح الدين علام (2000). وبمعنى آخر، فإن الأداة المبنية على أساس القياس النسبي تختلف جوهريًا عن الأداة المبنية على أساس القياس المطلق، لما لذلك من انعكاس مباشر على تفسير النتائج. فعلى سبيل المثال عند قياس التحصيل الدراسي لا يمكن تفسير الدرجات المحصلة بمعزل عن إطار مرجعي يُنسب إليه القياس، والذي قد يتمثل في جداول المعايير أو الدرجات التائية أو نسبة الأهداف المحققة، تبعًا للغرض من القياس. ومن ثمّ يُعدّ مفهوم الفروق الفردية حجر الأساس في عملية القياس النفسي، إذ يؤثر في طرق بناء الأدوات وأساليب تحليل البيانات وتفسير النتائج بشكل عام.

4-الخطأ مكون أساسي من مكونات عملية القياس:

يُعدّ الخطأ مكونًا أساسيًا في عملية القياس كما توضّح المسلّمة الرابعة من مسلّمات نظرية القياس، والتي تتناول جانبًا حاولت العديد من العلوم الأخرى معالجته أو الحدّ من تأثيره وهو الخطأ في القياس. فكل نتيجة نحصل عليها من أي أداة قياس تتكوّن في جوهرها من درجتين: الأولى تمثل الدرجة الحقيقية التي تعبّر عن السمة أو الخاصية المراد قياسها والثانية تمثل الخطأ في القياس الناتج عن عوامل متعددة. وبذلك يمكن القول إن الدرجة الكلية = الدرجة الحقيقية + درجة الخطأ، وهو ما يعكس وجود الخطأ كعنصر ملازم لأي عملية قياس، خاصة في الميادين النفسية والتربوية حيث يصعب عزل العوامل المؤثرة بشكل كامل. وقد جرى تصنيف أخطاء القياس إلى ثلاثة أنماط رئيسية تختلف في مصادرها وطبيعتها وتأثيرها على دقة النتائج.

أ- يعد الخطأ الثابت (المنتظم) النمط الأول من أخطاء القياس، وهو نوع من الخطأ ينشأ من الأداة نفسها ويتكرر بطريقة منتظمة ليؤثر بنفس المقدار على جميع درجات الأفراد الذين يُطبق عليهم الاختبار. فمثلاً إذا وُجد خلل في تدريج ميزان لقياس الأوزان يؤدي إلى زيادة مقدارها نصف كيلوغرام، يمكن بسهولة تصحيح الوزن الحقيقي بطرح هذا المقدار الثابت من النتيجة الظاهرة مما يجعل هذا النوع من الخطأ غير مسبب لمشكلة حقيقية في عملية القياس طالما أمكن تحديد قيمته بدقة.

غير أن الأدوات النفسية والتربوية لا تتيح عادة معرفة كمية هذا الخطأ المنتظم أو ما يُعرف بخطأ الأداء، ومع ذلك وبما أنه يؤثر بشكل متماثل على جميع الدرجات فإن المعنى الأساسي للقياس القائم على الفروق الفردية لا يتغير. أي أن وجود مقدار ثابت من الخطأ المرتبط بالأداة لا يُضعف من فعالية القياس في الميدان النفسي لأن جوهر القياس فيه يقوم على مقارنة الفروق بين الأفراد وليس على القيم المطلقة للدرجات.

ب- النمط الثاني هو خطأ القياس، ويقصد به الخطأ الناتج عن استخدام الدرجة الظاهرية بدلاً من الدرجة الحقيقية في عملية القياس. ويُعد هذا النوع من الأخطاء قابلاً للتحكم والمعالجة من قِبل الباحث، وذلك من خلال تطبيق إجراءات إحصائية مناسبة كتصحيح الدرجات من أثر التخمين أو تحويل الدرجات الخام إلى درجات معيارية أو تائية، بهدف تقليل تأثير هذا الخطأ وتحسين دقة النتائج المستخلصة من عملية القياس.

ج- أما النمط الثالث فهو الخطأ العشوائي أو ما يعرف بـ خطأ الصدفة، وهو نوع من الأخطاء يكون مصدره غير محدد ويصعب ضبطه أو التحكم فيه .تمتاز الأخطاء العشوائية بأنها تلغي بعضها البعض، خصوصاً عندما يكون حجم العينة كبيراً. ولتوضيح العلاقة بين هذه الأخطاء والدرجات المختلفة يمكن القول إن:

الدرجة الكلية = الدرجة الحقيقية + درجة الخطأ العشوائي

$$\boldsymbol{X = T + E}$$

ويعني ذلك أن الدرجة الكلية تمثل المجموع الجبري لكل من الدرجة الحقيقية ودرجة الخطأ العشوائي، إذ قد يكون هذا الأخير موجباً أو سالباً، مما يؤدي إلى زيادة أو نقصان في قيمة الدرجة الحقيقية. ويُفترض أن متوسط الأخطاء العشوائية يساوي صفراً $\sum E = 0$ عندما يكون حجم العينة كبيراً، وبالتالي يكون معامل الارتباط بين الدرجات الحقيقية ودرجات الخطأ العشوائي مساوياً للصفر $r_{T_i E_i} = 0$.

ويُعزى ذلك إلى أن الأخطاء العشوائية الموجبة قد تظهر لدى أصحاب الدرجات العالية، في حين قد ترتبط الأخطاء السالبة بالدرجات المنخفضة أو بالعكس، مما يؤدي في النهاية إلى عدم وجود أي علاقة منتظمة بين الدرجات الحقيقية وأخطاء القياس العشوائية (العنكبي، 2009، ص 21).

كما أن درجات الخطأ العشوائي الناتجة عند تطبيق اختبار معين على مجموعة من الأفراد لا ترتبط بدرجات الخطأ العشوائي الناتجة عن تطبيق اختبار آخر على نفس المجموعة أو تطبيق الاختبار نفسه على مجموعة مختلفة، أي أن معامل الارتباط بينهما يساوي صفر $r_{E1E2} = 0$.

وبوجه عام يعد الخطأ أحد المكونات الأساسية في عملية القياس في ميادين التربية وعلم النفس، إذ لا توجد أي عملية قياس ـــ مهما بلغت دقتها أو درجة التحكم فيها من قبل الباحث ـــ تخلو تماماً من درجة من الخطأ.

فكل درجة يتم الحصول عليها تتكون من جزء يمثل الخطأ بمصادره المختلفة، وجزء يعكس الدرجة الحقيقية أو السمة الفعلية للفرد. وكلما تمكن الباحث من تقليل أثر هذا الخطأ، ازدادت موضوعية النتائج ودقة تفسيرها، مما يرفع من مصداقية القرارات المستندة إليها. إن الإقرار بأن الدرجة الظاهرية ليست درجة حقيقية خالصة، بل مزيج من الدرجة الحقيقية ودرجة الخطأ، يجعل من الضروري التعامل معها بمنطق نسبي وتقريبي عند التحليل الاحصائي والتفسير (بوسالم، 2014، ص 30–38).

الدكتور زيـاد رشيـد

المحـــاضرة الرابعة

مجـــالات القيـــاس النفســـي
(القياس العقلي)

الأهداف:

تهدف هذه المحاضرة إلى تزويد الطلبة بفهم شامل لمجال القياس العقلي من خلال التعرف على مفهوم القدرات العقلية وأنواعها وتصنيفاتها، واستيعاب مفهوم الذكاء والإشكاليات المرتبطة بتعريفه. كما تسعى إلى توضيح العوامل المؤثرة في الذكاء، مع إبراز طبيعة التفاعل النسبي بين الوراثة والبيئة في تشكيله. وتهدف أيضًا إلى عرض أبرز نماذج ونظريات الذكاء، والتعرف على أدوات قياس الذكاء العام وأساليب استخدامها. إضافة إلى ذلك، تركز المحاضرة على التطبيقات العملية للقياس العقلي في مجالات التعليم والإرشاد والعيادات النفسية، مع مناقشة خصوصية القياس العقلي في البيئة العربية والتحديات الثقافية المرتبطة به وعمليات التقنين. كما تهدف إلى توضيح مفهوم الاستعداد وعلاقته بالقدرة، والتعريف بمفهوم الابتكارية وعلاقته بالتفكير التباعدي، بما يسهم في تنمية فهم متكامل للقدرات العقلية وأساليب قياسها وتوظيفها في الواقع التطبيقي.

محتوى المحاضرة:

أولاً: القياس العقلي (قياس القدرات العقلية)
1- القدرات العقلية (Mental abilities / Cognitive abilities)
2- تصنيف القدرات العقلية
ثانياً: مفهوم الذكـــاء
ثالثاً: مشكلة تعريف الذكـــاء
رابعاً: العوامل المؤثرة في الذكـــاء
خامساً: التفاعل النسبي بين الوراثة والبيئة
سادساً: نماذج ونظريات الذكاء
سابعاً- أدوات قياس الذكاء العام
ثامناً: تطبيقات القياس العقلي في التعليم والإرشاد والعيادات النفسية
تاسعاً: القياس العقلي في البيئة العربية (التحديات الثقافية والتقنين)
عاشراً: مفهوم الاستعداد وعلاقته بمفهوم القدرة
الحادي عشر: الإبتكاريـــة ومفهوم التفكير التباعـدي

مدة المحاضرة: أربع ساعات ونصف

تمهيد:

يعد القيــاس العقلـي أحد الركائز الأساسيـة في علم النفس، حيث يهدف إلى فهم القدرات العقلية للأفراد وتقدير الفروق الفردية بينهم بطريقة علمية موضوعية. وقد حظي هذا المجال باهتمام كبير منذ بدايات علم النفس الحديث، نظرًا لأهميته في تفسير السلوك الإنساني وتوجيه القرارات التربوية والإرشادية والعيادية. ويعد الذكاء من أبرز المفاهيم التي شغلت الباحثين، ليس فقط من حيث قياسه، بل أيضًا من حيث تحديد طبيعته والعوامل المؤثرة فيه، وما يرتبط به من مفاهيم كـالقدرة والاستعداد والابتكارية. كما أن تطور نماذج ونظريات الذكاء، وتعدد أدوات قياسه، يعكس تعقيد هذا البناء النفسي وتداخله مع العوامل الوراثية والبيئية. وفي ظل التحديات الثقافية، خاصة في البيئة العربية، تبرز أهمية التقنين والتكييف الثقافي لأدوات القياس لضمان دقتها وصدقها. ومن هذا المنطلق، تأتي هذه المحاضرة لتسليط الضوء على أسس القياس العقلي، وأهم مفاهيمه ونظرياته، وتطبيقاته في مختلف المجالات.

أولاً: القيــاس العقلــي (قياس القدرات العقلية)

1- القدرات العقليــة (Mental abilities / Cognitive abilities):

تشير القدرات العقلية إلى مجموعة من العمليات المعرفية التي تمكن الفرد من التفكير والفهم والتعلم وحل المشكلات واتخاذ القرار والتكيّف مع البيئة. وتعد القدرات العقلية مفهوماً أوسع من الذكاء العام، إذ تتضمن قدرات محددة كالانتباه، والذاكرة، والقدرة اللغوية، والقدرة العددية، والقدرة المكانية، وسرعة المعالجة، والتفكير المجرد، والقدرة الإبداعية. ويؤكد الباحثون أنّ القدرات العقلية تتوزع عبر مستويات من القدرة العامة (g) إلى قدرات واسعة ثم قدرات ضيقة كما في نموذج كارول–هورن–كاتل (CHC)، الذي يعد من أبرز النماذج الحديثة في فهم بنية القدرات العقلية. وينظر إلى هذه القدرات باعتبارها أساساً للتعلم والأداء الأكاديمي والمهني، حيث تشكل المعرفة والمهارات العقلية أدوات أساسية لتحقيق التكيف والنجاح في الحياة اليومية والسياقات التعليمية والمهنية (McGrew, 2009; Sternberg, 2020).

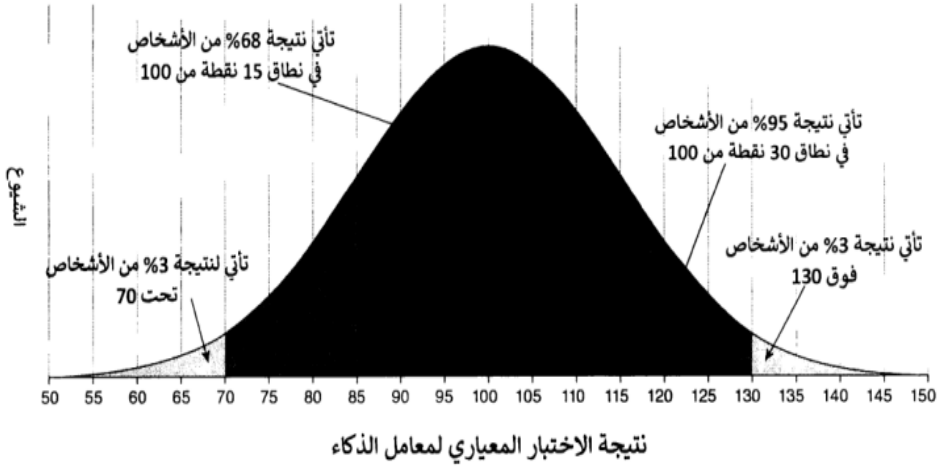
2- تصنيف القدرات العقلية:

يوضح الشكل التالي التوزيع الطبيعي لدرجات معامل الذكاء، وبناءً عليه يمكن تصنيف القدرات كما يلي: تعد الدرجة المتوسطة لمعامل الذكاء (100) هي المستوى العادي، حيث يقع أغلب الأفراد (حوالي 68%) ضمن المجال الممتد بين (85–115)، ويُصنفون ضمن القدرات المتوسطة. أما

الأفراد الذين تقع درجاتهم بين (70–85) فيمثلون فئة أقل من المتوسط، في حين أن من تقل درجاتهم عن (70) (حوالي 3%) يُصنفون ضمن ضعف القدرة أو تأخر عقلي. في المقابل، الأفراد الذين تتراوح درجاتهم بين (115–130) يُصنفون ضمن فوق المتوسط، بينما الذين تزيد درجاتهم عن (130) (حوالي 3%) يُعدّون ضمن فئة القدرات العالية أو الموهبة. كما يشير الشكل إلى أن حوالي 95% من الأفراد يقعون بين (70–130)، وهو ما يعكس الانتشار الطبيعي للقدرات في المجتمع.

شكل رقم 02 يوضح المنحنى الطبيعي أو الجرس للذكاء البشري

(فالك، 2024، ص50)



ثانياً: مفهوم الذكـــاء:

حاول علماء النفس المبكرون إلى التعرف على مفهوم الذكاء وبالفعل ظهرت تعريفات عديدة في هذا المجال، ولكن سرعان ما تبينوا أن ما بينهم من اختلاف أكثر مما بينهم من اتفاق، وقد يرجع ذلك إلى ارتباط الذكاء بالعديد والعديد من الأنشطة المختلفة للسلوك الإنساني. وقد يرجع هذا الاختلاف أيضاً إلى أن الذكاء ليس شيئاً مادياً محسوساً وإنما هو شيء معنوي، وبالتالي لا يمكن قياسه مباشراً وإنما يقاس بطريقة غير مباشرة. وفيما يلي نحاول أن نستعرض أهم التعريفات التي ظهرت منذ نشأة مفهوم الذكاء والتي حاولت أن تقدم تفسيراً لطبيعته (محمود وعفاف، 2006، ص185-191).

1- المفهوم اللغوي للذكـــاء (Linguistic Intelligence): ترجم علماء النفس العرب مصطلح Intelligence بكلمة ذكاء، وهي مشتقة من الفعل الثلاثي ذكا ويوضح المعجم الوسيط أصل هذه الكلمة لغوياً، فـ ذكاء النار يعني اشتد لهيبها واشتعلت، ويقال ذكت الشمس أي اشتدت حرارتها، وذكاء فلان ذكاء أي سرع فهمه. ويعني المفهوم اللغوي للذكاء القدرة الشفهية على استخدام الكلمات بكفاءة كما في الخطابة وإلقاء الشعر والصحافة والتأليف وكذلك استخدام اللغة في توصيل معلومات معينة. أو

محاولة تذكرها والاعتماد عليها في عمليات التعليم والتعلم ومحاولة فهم قواعد اللغة النحوية وفهم معاني الكلمات.

2- المفهوم الفلسفي للذكاء: درج أرسطو أن نشاط الكائن الحي ومنها بالطبع النشاط العقلي موجود في جسمه على هيئة استعداد كامن وهو المسئول عن حركات جسم الكائن الحي والتي تقوم بوظائف الدافعية قوة من داخل جسم الكائن الحي سواء النبات أو الحيوان أو الإنسان ، إلا أن الإنسان يتميز بأن لديه أرقى صور النفس وهو العقل الذي يتحكم في نشاطه وتفكيره ، وهكذا اعتبر الذهن أو العقل الخاصية التي تميز الإنسان عن الحيوان. وقد نقلت آراء أرسطو إلى أوروبا عن طريق فلاسفة من العرب والمسلمين مثل الفارابي وابن سينا والغزالي وابن رشد وغيرهم من الفلاسفة. وقد اهتم ابن رشد بفكرة الفروق الفردية في النواحي العقلية في كتابه " الدعاوي القلبية".

كما اهتم أيضاً بفكرة وحدة العقول والتي بدت عند بعض فلاسفة العصور الوسطى. ومن هاتين الفكرتين نجد تصور ابن رشد أقرب إلى التصور الحديث للنشاط العقلي. وقد تتابعت المذاهب الفلسفية في محاولتها تفسير النشاط العقلي ومن أهمها:

أ- نظرية الملكات: وترى نظرية الملكات أن العقل يتكون من عدد من القوى كالإرادة والانتباه التي تؤدي إلى الأنشطة العقلية المختلفة. وقد استطاع كل من " جال وسبيرزهايم " من جمع كثير من البيانات التي تقيم الدليل على نظرية الملكات من زيارتهما للمؤسسات التربوية والمستشفيات و ضعاف العقول والسجون.

ب- الفلسفة الارتباطية: تمتد الفلسفة الارتباطية Associationism بجذورها إلى الفيلسوف اليوناني أرسطو الذي صاغ قوانين الارتباط الثلاثة وهي الاقتران والتشابه والتضاد، إلا أن الارتباط بمعناها الحديث تعود إلى الفيلسوف الإنجليزي توماس الذي اختصر قوانين أرسطو في قانون واحد هو الاقتران، كما حاول التمييز بين التداعي الحر والتداعي المقيد. وقد دعم الفيلسوف جون لوك هذا الاتجاه بنظريته الهامة في أن الخبرة الحسية هي مصدر المعرفة الوحيد، ويدعونه إلى اختزال العمليات المركبة وردها إلى مكوناتها الارتباطية البسيطة، وهذه تعتبر نقطة ضعف عندما اختزلت العمليات العقلية العليا إلى مجرد ارتباطات حسية بسيطة مما يؤدي إلى تبسيط العقل البشري تبسيطاً مخلاً، ومع ذلك فإن الاتجاه المعاصر الذي يتمثل في تجهيز المعلومات يعتمد بأصوله إلى الفلسفة الارتباطية.

3- المفهوم البيولوجي للذكاء: عرف الفيلسوف الإنجليزي هربرت سبنسر مفهوم الذكاء على أنه قدرة موحدة، كما اعترف بوجود جانبين للحياة العقلية هما الجانب المعرفي والجانب الوجداني. وبذلك يعتبر سبنسر أول من استخدم مصطلح الذكاء مؤكداً أهميته في النواحي البيولوجية، كما عرف سبنسر الحياة بأنها عملية تكيف ومستمر ومتصل من جانب العلاقات الداخلية للعلاقات الخارجية، بمعنى التكيف الذاتي للفرد مع بيئته الخارجية التي يعيشها. ويرى سبنسر أن هذا التكيف يمكن الوصول إليه عن طريق الغريزة عند الكائنات الحية الدنيا، أما الإنسان فيتكيف عن طريق ما يتعلم. وقد سادت نظريات سبنسر في الذكاء عند علماء البيولوجيا البريطانيين، ولعب مفهوم التوافق باعتباره جوهر الذكاء دوراً هاماً عند بينيه وأثر في المدرسة الوظيفية الأمريكية لعلم النفس.

ويفسر "بياجيه" النشاط العقلي لمفهوم الذكاء في صور بني معرفية يسمح للكائنات الحية كل حسب نوعه بالتوافق سواء من خلال التمثل (أي تغيير في جوانب البيئة) أو عن طريق المواءمة (أي التغيير الذي يحدث للكائن العضوي نفسه)[1].

4- المفهوم الفسيولوجي للذكاء: أكدت الدراسات التجريبية للمخ أنه يعمل ككل ولا يوجد جزء من المخ يعمل منفصلاً عن باقي الأجزاء، وأن النشاط العصبي يتميز بخاصية التنميط[2] وهي التي تتضمن بدورها خاصية التكامل وهذا ما أكده " شرنجتون " ، ويدعم هذا نتائج "لاشلي" عن النشاط الكتلي للمخ ، وقد اعتبر كل منهما أن هذا النشاط الكتلي مطابقاً لمفهوم العامل العام للذكاء.

5- المفهوم الاجتماعي للذكاء: نظراً لأن الإنسان اجتماعي بطبعه وطبيعته، ولا يعيش في فراغ وإنما يعيش في مجتمع يتأثر به ويؤثر فيه، من هنا ربط أصحاب هذا الاتجاه بين الذكاء والعوامل الاجتماعية السائدة في المجتمع من عادات وتقاليد وقيم وأنماط تفكير وأساليب سلوك إلى غير ذلك من العوامل التي تعتبر نتاجاً للتفاعل الاجتماعي. ويتضمن هذا الذكاء معرفة باللغة والثقافة والمفاهيم

---

[1] - **التمثّل (الاستيعاب) (Assimilation):** هو أن يستخدم الطفل المعارف أو المخططات العقلية الموجودة لديه لفهم خبرة جديدة بدون تغييرها.
مثال: طفل صغير يعرف كلمة "كلب" ويعتقد أن كل الحيوانات ذات الأربع أرجل هي "كلب." عندما يرى بقرة لأول مرة يقول" :انظر! كلب"! هنا الطفل استعمل نفس المخطط السابق (حيوان بأربع أرجل = كلب) لاستيعاب المعلومة الجديدة.
-المواءمة (التكيف) (Accommodation): هي عندما يُعدّل الطفل مخططاته العقلية أو ينشئ مخططًا جديدًا لفهم شيء جديد بشكل صحيح.
مثال: بعد أن يشرح له أحد الكبار أن هذا الحيوان هو "بقرة" وليس "كلب"، يبدأ الطفل بتعديل مخططه: الكلب: حيوان بأربع أرجل، صغير، ينبح البقرة: حيوان بأربع أرجل، كبير، يعيش في المزرعة، يصدر صوتًا آخر. هنا الطفل غيّر المخطط وأضاف فئة جديدة.

[2] - يُشير التنميط العصبي إلى قدرة الدماغ على تشكيل أنماط منتظمة من النشاط العصبي عند معالجة المعلومات، مما يُعدّ أساسًا مهمًا لفهم الذكاء. فكلما كانت هذه الأنماط أكثر كفاءة وتنظيمًا، زادت قدرة الفرد على التعلم، وحل المشكلات، واتخاذ القرارات. وفي القياس النفسي، يُفسَّر الذكاء بوصفه نتاجًا لتفاعل شبكات عصبية مترابطة، وليس مجرد محصلة لقدرات منفصلة أو درجة ثابتة في اختبار موحد، إذ تتطور هذه الأنماط عبر الخبرة والتدريب.
مثال :عندما يتعلم الطالب حل مسائل رياضية، يُنشئ الدماغ نمطًا عصبيًا متخصصًا. ومع التكرار، يصبح حل النوع نفسه من المسائل أسرع وأكثر دقة، ما يعكس تطورًا في الكفاءة العقلية وفعالية المعالجة المعرفية.

الاجتماعية المختلفة التي تتعلق بالزمان والمكان والأعراف والقيم والتقاليد الاجتماعية، ومن ثم فإن ذكاء الفرد الاجتماعي يتمثل في قدرته على التفاعل الإيجابي مع المجتمع الذي يعيشه ، وبالتالي فإن ذكاء الفرد الذي يعيش في مجتمعات متحضرة لا شك أنه أكبر من ذلك الفرد الذي يعيش في مجتمعات بدائية أولية.

ولذلك يعرف ثورنديك Thorndike الذكاء الاجتماعي بأنه قدرة الفرد على التفاعل بإيجابية مع الآخرين، وهو يرى أن الذكاء الاجتماعي للفرد يعتمد على سنه وجنسه وثقافته ومكانته الاجتماعية. ويؤكد بعض علماء هذا الاتجاه على دور الذكاء في النجاح الاجتماعي لذلك وجب انتقاء الأذكياء من الأفراد لتولي المناصب الاجتماعية القيادية في المجتمع (رمزية الغريب، 1985).

ثالثاً: مشكلة تعريف الذكـــاء:

يتضح لنا من عرضنا السابق لتطور مفاهيم الذكاء أن هناك عدة اتجاهات مختلفة تفسر مفهوم الذكاء، وبالتالي فإنه لا يوجد اتفاق تام على تعريفه. ونظراً لتعدد مظاهر النشاط العقلي تعددت تبعاً له تعريفات الذكاء والتي يمكن عرضها من خلال الأقسام التالية:

1 - القدرة على التعلم: فبعض علماء النفس من المهتمين بالعملية التعليمية عرفوا الذكاء على أنه قدرة الفرد على التعلم ومن أمثلة ذلك:

-تعريف كلفن Colvin للذكاء بأنه القدرة على التعلم.

-تعريف وودرو Woodrow بأنه القدرة على اكتساب القدرة.

-تعريف جودانف Coodenough بأنه القدرة على الإفادة من الخبرة.

وطبقاً لهذه يتحدد مدى ذكاء الفرد بمدى قدرته على التعلم والاستفادة من خبراته السابقة.

2 - القدرة على التوافق: وبعض علماء النفس يؤكد على أن الذكاء يعتمد على مدى تكيف الفرد وتوافقه مع المواقف الجديدة التي يمر بها، ومن أمثلتها:

- تعريف شترن Stern بأنه القدرة على التصرف السليم في المواقف الجديدة.

- تعريف بنتر Pintnre بأنه قدرة الفرد على التوافق بنجاح للعلاقات الجديدة في الحياة.

وطبقاً لهذا يعتبر الذكاء قدرة عقلية تساعد الفرد على مواجهة المواقف الجديدة ومحاولته حل المشكلات التي تواجهه.

3 - القدرة على التفكير: ومن أمثلة هذه التعريفات:

- تعريف ترمان Terman بأنه القدرة على التفكير المجرد.

-تعريف بينيه Bient بأنه القدرة على الحكم السليم.

-تعريف كوهلر Kohler بأنه القدرة على إدراك العلاقات وهذا يعني أن الذكاء قدرة الفرد على استعمال المفاهيم المجردة والرموز المختلفة في مواجهة المواقف التي يتطلب حلها استعمال الرموز اللغوية أو العددية مثل حل مسابقات الكلمات المتقاطعة أو الألغاز الرياضية.

4 - القدرة الفطرية (الوراثية): ومن أمثلتها:

- تعريف بيرت Burt بأنه القدرة العقلية المعرفية الفطرية العامة.

- تعريف سبيرمان Sperman بأنه قدرة فطرية عامة.

- تعريف ثورنديك بأنه عبارة عن إمكانيات كامنة في التكوين الجسمي للكائن الحي، موروثة وليست مكتسبة.

5 - التعريف الإجرائي: وقد اتجه بعض علماء النفس إلى التعريف الإجرائي والعملي للذكاء ومن أمثلة ذلك:

- تعريف جاريت Garrett بأنه القدرة على النجاح في المدرسة أو الكلية.

- تعريف بورنج Boring بأنه القدرة على الأداء الجيد على اختبارات الذكاء، ويعد هذا التعريف من أكثر تعاريف الذكاء شيوعاً.

وقد أمكن اختصار تعريف بورنج (Boring) للذكاء هو أحد أشهر التعريفات في تاريخ علم النفس، وقدَّمه عالم النفس الأمريكي إدوين بورنج عام 1923، وجاء بصيغة نقدية ساخرة تسلط الضوء على محدودية تعريفات الذكاء آنذاك. في القول المشهور: "أن الذكاء هو ما تقيسه اختبارات الذكاء"، وبالطبع لابد وأن تتوافر في هذه الاختبارات الشروط الأساسية التجريبية وفي مقدمتها الصدق والثبات. ومعنى ذلك أن تعريف بورنج يوجه الأنظار إلى الاتجاه الصحيح نحو دراسة الاختبارات المستخدمة في القياس دراسة دقيقة لتحديد ما تقيسه، وبالتالي فإننا نحصل على تعريفات عديدة للذكاء بقدر ما لدينا من اختبارات تقيس سمات مختلفة للذكاء.

ورغم أن مفهوم الذكاء لا يزال موضوع اختلاف بين علماء النفس كل حسب اهتماماته ودراساته، إلا أنهم يتفقون والغالبية من الناس على الصفات التي تميز الشخص الذكي من غيره، فالفرد الذكي يتميز بأنه:

أ- أشد يقظة وأسرع في الفهم عن غيره.

ب- أقدر وأسرع على التعلم.

ج- أقدر على تطبيق ما تعلمه لحل ما يتعرض له من مشكلات.

د- أقدر على إدراك ما بين الأشياء والأعداد والألفاظ من علاقات.

ه- أقدر على التبصر في عواقب الأمور.

و- غالباً ما يكون أنجح في الدراسة أو العمل أو في الحياة بصفة عامة.

رابعاً: العوامل المؤثرة في الذكـــاء:

تصنف العوامل المؤثرة في النشاط العقلي الإنساني إلى عاملين رئيسيين هما: أولهما العوامل الوراثية، وثانيهما العوامل البيئية، إلا أن هذا التصنيف الثنائي هو فقط من أجل التبسيط والدراسة فقط، لأنه من الصعب الفصل التام بين هذين العاملين، كما سوف نوضح ذلك تفصيلاً فيما بعد.

1- العوامل الوراثية:

تبدأ حياة الفرد عند اتحاد الأمشاج الذكرية بالأنثوية وإخصاب الحي المنوي الذكري للبويضة الأنثوية حيث يتكون الزيجوت Zygote وهي الخلية الحية المخصبة والتي تحتوي على (23) زوج من الكروموزومات أو الصبغيات لأنه لا يمكن رؤيتها إلا تحت المجهر إذا لونت، ويوجد على هذه الكروموزومات آلاف الجينات وهي المسئولة عن انتقال الصفات والخصائص الوراثية سواء الجسمية أو العقلية أو الخلقية من الأجداد الأول إلى الآباء ثم الأبناء فالأحفاد وهكذا. فكل طفل يرث نصف ما لأبوية من موروثات (جينات)، فمن طريق الجينات يرث الأبناء عن والديهم الخصائص الجسمية والعقلية وأيضاً الخلقية والمزاجية،

1-1- الذكاء والوراثة: هناك رأي سائد بين علماء النفس أن الذكاء عبارة عن قدرة كامنة تتحدد أصلاً نتيجة التكوين الوراثي الأصلي للفرد. وقد عرف الذكاء كما ذكرنا سابقاً على أنه قدرة عقلية فطرية عامة، فهو عبارة عن الإرث البيولوجي الذي يرثه الأبناء عن الآباء والأجداد. ويحدد ثورنديك Thorndike الذكاء بأنه إمكانيات كامنة في التكوين الجسمي للكائن الحي هي موروثة وليست مكتسبة. كما يتضح لنا ذلك أيضاً من تعريف بيرت Burt للذكاء بأنه القدرة العقلية المعرفية النظرية العامة. وتعريف خليفة بركات بأنه " قدرة عضوية لها أساس في التكوين الجسماني. فأن الذكاء يعتبر استعداد كامن في التكوين العضوي للكائن الحي عن طريق الوراثة، فأن العوامل البيئية هي التي تعمل على قدحه ونموه وازدهاره.

2- العوامل البيئية:

إذا كانت الوراثة تمثل كافة العوامل الداخلية التي تؤثر في نمو الجنين منذ اللحظة التي يخصب فيها الحي المنوي للبويضة الأنثوية حيث يتم منذ لحظة الإخصاب وضع البصمة الوراثية للجنين، ويعد هذه اللحظة مباشرة يبدأ تأثير العوامل البيئية، ولذلك فأن رحم الأم ما هو إلا بيئة مهيئة لنمو الجنين من درجة حرارة ورطوبة وتغذية إلى غير ذلك من العوامل البيئية المناسبة لنموه نمواً سوياً سليماً. ويمكن تعريف البيئة من الناحية النفسية بأنها مجموع الاستثارة التي يتلقاها الفرد من لحظة إخصاب البويضة في رحم الأم حتى وفاته، وعليه فإن البيئة يتسع مجالها ليشمل بيئة ما قبل الولادة في المرحلة الجنينية، وكذلك بيئة ما بعد الولادة، وهكذا يتسع مجال البيئة ليشمل كل صور الاستثارة التي تمتد طول حياة الفرد، منذ لحظة الإخصاب حتى وفاته كما ذكرنا. وأن جميع صور الاستثارة التي يمر بها الفرد في حياته تؤثر من قريب أو حتى من بعيد على نمو ذكائه من حيث أنها تتيح إمكانيات عديدة لهذه القدرة الكامنة لأن تنمو وتزدهر وتتطور وتعبر عن نفسها بالطريقة التي تقيسه اختبارات الذكاء ، وفيما يلي نتناول بعض العوامل البيئية التي تؤثر في الذكاء من جهة وتتأثر بها وهي:

2-1- الذكاء والتعليم: أثبت البحوث والدراسات في هذا الصدد أن هناك علاقة إرتباطية موجبة ومرتفعة بين ذكاء التلميذ وتحصيله الدراسي، لدرجة أننا قد نستعمل مصطلح الذكاء ليدلل عن المستوى التحصيلي المرتفع للتلميذ، وبالتالي فإن الذكاء مرتبط بالنجاح في الدراسة، صحيح أن هناك عوامل بيئية أخرى تساهم في نجاح التلميذ دراسياً من حالته النفسية والصحية والاقتصادية إلى غير ذلك من العوامل إلا أن الذكاء يعتبر العامل الرئيسي الأول من بين تلك العوامل الذي يؤثر في نجاحه الدراسي، و لعل أدل على ذلك من البحوث التجريبية التي أجريت على مدى النجاح الدراسي لمجموعتين من الأطفال أحداهما التحقت بمدارس ما قبل التعليم الابتدائي والأخرى ظلت في منازلهم حتى التحقوا بالمدرسة الابتدائية مباشرة، وقد دلت النتائج على تفوق التلاميذ الذين التحقوا بمدارس ما قبل المدرسة الابتدائية تحصيلاً على المجموعة الأخرى التي لم تلتحق بمثل هذه المدارس، كما وجد أن درجات ذكائهم أعلى من تلك المجموعة تتلقى فرصة الالتحاق بمدارس ما قبل المدرسة الابتدائية. فالذكاء والتعليم هما وجهان لعملة واحدة، فإذا كان حدوث عملية التعلم يتوقف على قدر مناسب من الذكاء، فإن التفوق في الذكاء يؤدي بالضرورة إلى التفوق الدراسي.

2-2- الذكاء والبيئة الثقافية: لاشك أن ذكاء الفرد يتأثر بالبيئة الثقافية التي يعيشها وذلك من خلال القراءة والاطلاع والبحث وكتابة الشعر والقصص والخطابة والرسم إلى غير ذلك من العوامل الثقافية التي تؤثر في ذكاء الفرد عن طريق الخبرات والمعارف والمهارات التي يكتسبها، كما يتأثر بالذكاء من جهة أخرى.

2-3- الذكاء والبيئة الاقتصادية: إن البيئة الاقتصادية الغنية تساهم مساهمة فاعلة في تنمية ذكاء الفرد من خلال توفير الوسائل والأدوات التي من شأنها أن تساهم في تنمية الذكاء وتطويره.

2-4- الذكاء والأسرة: وتعد الأسرة من أهم عوامل تطور ونمو ذكاء أبنائها من خلال استخدام أساليب التربية السوية السليمة التي تساهم في نموه انفعالياً وخلقياً وثقافياً وعلمياً، وعن طريق توفيرها المناخ النفسي المناسب لنمو قدراتهم واستعداداتهم وتوفير الأدوات والوسائل والإمكانيات التي تساهم في قدح استعداداتهم العقلية واحترام فرديته.

2-5- الذكاء والبيئة الاجتماعية: فالإنسان كما نعلم اجتماعي بطبعه وطبيعته، ولا يستطيع أن يعيش منعزلاً عن الآخرين وبالتالي لابد له من أن يتشبع بعادات وتقاليد وأعراف المجتمع الذي يعيشه مما يمكنه التعامل بذكاء مع الآخرين في ضوء عادات وتقاليد وقيم الجماعة.

2-6- الذكاء والبيئة الصحية: إن العقل السليم في الجسم السليم، وبالتالي فإن الاهتمام والعناية والرعاية الصحية لأفراد المجتمع لابد أن يكون لها مردود إيجابي على حالتهم الجسمية، ومن ثم الصحة وأيضاً النفسية وبالتالي العقلية.

خامساً: التفاعل النسبي بين الوراثة والبيئة:

احتدم الجدل بين كل من أصحاب الوراثة والبيئة، فقد أرجع أنصار الوراثة قدرات الإنسان واستعداداته وجميع سمات شخصيته إلى الجينات الوراثية، واعتمدوا في ذلك قديماً على نظرية الغرائز التي تخلى عنها حالياً علماء النفس، ويؤكدون أهمية ودور العوامل الوراثية من خلال دراسة التطابق التام للقوائم المتصلة والتي هي عبارة عن بويضة مخصبة ثم تنشطر إلى قسمين متماثلين تمام التماثل، فنلاحظ تشابه تام بين هذا التوأم والذي يرجع إلى تطابق العوامل الوراثية له.

أما أنصار البيئة فقد عارضوا دعوى أنصار الوراثة وتطرفوا برفضهم كل أثر للوراثة، وقد تزعم هذا الاتجاه واطسن Watson عندما قال " أعطوني اثني عشر طفلاً أصحاء وأعطوني عالمي الخاص الذي فيه أتمكن من تنشئتهم، وأنا أضمن لكم أني اخترت أي واحد منهم عشوائياً ودربته على أن يتخصص في أي ناحية اختارها، فإني أستطيع أن أجعله منه طبيباً أو محامياً أو فناناً أو تاجراً

ناجحاً، بل وشحاذاً أو لصاً بصرف النظر عن قدراته واستعداداته، أو حرفة أجداده، والجنس الذي ينتمي إليه " ويبدو في قوله هذا مدى التطرف الذي وصل إليه. وفي حقيقة الأمر أنه يصعب فصل أثر الوراثة عن البيئة في تأثيرها على النمو الإنساني، هذا التفاعل يحدد صفات الفرد ومستويات نضجه ومدى تكيفه.

سادساً: نماذج ونظريات الذكاء:

نظراً لما تتمتع به القدرات العقلية اهتماماً كبيراً من قبل علماء النفس المهتمين بقضايا القياس العقلي، وقد تجاوز هذا الاهتمام إلى المسؤولين عن اتخاذ القرار في مختلف المجالات التطبيقية التربوية والنفسية والاجتماعية. وقد اتجه القياس العقلي في أواخر القرن العشرين إلى استعمال الوسائل الإحصائية في الكشف عن التكوين العقلي، وإلى استعمال طريقة التحليل العاملي بصفة خاصة، و تفسير العوامل التي تنتج عن هذا التحليل. ونتيجة لذلك ظهرت عدة نظريات لتفسير التكوين العقلي، وكيف تنظم القدرات العقلية، ونستعرض فيما يلي أهم هذه النظريات.

1- نظرية العاملين(Two-Factor Theory) لـ(Charles Spearman):

تعد نظرية العاملين في الذكاء لعالم النفس "تشارلز سبيرمان" في بداية القرن العشرين، من أبرز المحاولات المبكرة التي وظّفت أسلوب التحليل العاملي (Factor Analysis) للكشف عن البنية النفسية للذكاء وفهم كيفية تنظيم القدرات العقلية، ومن أوائل النظريات العلمية التي فسّرت طبيعة الذكاء الإنساني. تفترض هذه النظرية أن الذكاء يتكوّن من عاملين رئيسيين: العامل العام (g) وهو القدرة العقلية العامة المشتركة التي تؤثر في أداء الفرد في جميع المهام الذهنية مثل الفهم والاستدلال وحل المشكلات؛ والعامل النوعي (s) وهو قدرة خاصة مرتبطة بكل مجال معرفي محدد مثل القدرة اللغوية أو الحسابية أو الذاكرة. وبناءً على هذه الفكرة، يرى سبيرمان أن نجاح الفرد في أي مهمة معرفية يعتمد على الذكاء العام إضافةً إلى القدرة الخاصة المناسبة لذلك النشاط. ورغم أن هذه النظرية أسهمت في تأسيس مقاييس الذكاء الحديثة وترسيخ مفهوم الذكاء العام، فقد تعرّضت للنقد لأنّها تبسِّط الذكاء الإنساني وتقلل من أهمية اختلاف القدرات المتخصصة بين الأفراد.

2- نظرية القدرات العقلية الأولية لثرستون Thurstone:

قام "ثرستون" بتطبيق عدد كبير من اختبارات الذكاء المتنوعة وذلك على عينة كبيرة من الأفراد، واستخدم طرق أفضل في التحليل الإحصائي، وقد توصل إلى نتائج مختلفة تماماً عما توصل إليه ثرستون عن طبيعة الذكاء. ونتيجة لاستخدامه للتحليل العاملي المناسب لمصفوفة معاملات

الارتباط للاختبارات المتعددة التي شملت جميع الأنماط السلوكية الممكنة التي يجمع علماء النفس على أنها تتعلق بقياس الذكاء، أمكنه استخراج عدد من العوامل أو المتغيرات التي يرجع إليها الأداء العقلي، ثم يحاول القائم بالتحليل العاملي بتجميع الاختبارات التي ترتبط مع بعضها ارتباطاً عالياً وعندئذ يقوم بتسمية هذا المتغير بالاسم الذي يصف الوظيفة العقلية المقيسة، وهكذا يمكن جمع أكثر من عامل أو متغير مع ملاحظة أن العامل الأول يتميز بأكبر نسبة من التباين ، ثم يقل هذا التباين إلى أن يصبح التباين الباقي ضئيلاً بحيث يتعذر معه استخراج عوامل أو متغيرات تستحق الذكر.

ولكي يضمن ثرستون عدم تداخل العوامل التي يخرج بها، وأن تكون بسيطة وفي نفس الوقت قاطعة الدلالة ، راعى أن تكون الاختبارات التي استخدمها في دراساته بسيطة جداً بحيث لا تتضمن عمليات عقلية معقدة حتى يسهل تسمية وتفسير العوامل التي يحصل عليها. وقد تمكن من تحديد عدد من العوامل الطائفية والتي أطلق عليها اسم **القدرات الأولية**، أهمها:

أ- القدرة اللغوية (الفهم اللغوي): وتشمل الاختبارات التي تتطلب معرفة معنى الكلمات وتصنيفها والعلاقات اللفظية وغيرها.

ب - القدرة على الطلاقة اللفظية: وتوجد في اختبارات مثل إيجاد الكلمات وتكوينها وذكر كلمات تبدأ أو تنتهي بحروف معينة .. إلى غير ذلك.

ج - القدرة العددية: وتظهر في الاختبارات التي تتعلق بإجراء العمليات الحسابية وتقيس سرعة إجراء هذه العمليات بدقة.

د- القدرة المكانية: وتكشف عنها الاختبارات التي تتطلب إدراك العلاقات المكانية والهندسية، وكذلك تتطلب تصور أبعاد الأشياء بعد تغيير وضعها.

ه - القدرة على السرعة الإدراكية: وتتمثل في الاختبارات التي تهدف إلى التعرف السريع الدقيق للتفاصيل وخاصة في الأشياء المرئية وأوجه الاختلاف والتشابه بينها.

و - القدرة على التذكر: وتكشف عنها الاختبارات التي تتطلب التعرف على الأشكال أو الربط بين أسماء سبق تعلمها معاً، أو أسماء وأرقام سبق تعلمها .... الخ.

ز - القدرة على الاستدلال: وتتمثل في الاختبارات التي تتطلب اكتشاف قاعدة كما في تكملة سلاسل الأعداد أو الحروف .... الخ.

ويرى "ثرستون" أن تلك المجموعة من القدرات هي التي يتوقف عليها النشاط العقلي الإنساني، وأن أي عملية عقلية يقوم بها الإنسان تدخل فيها هذه القدرات بأوزان مختلفة ، فعلى سبيل المثال فإن

حل تمرين هندسي يعتمد على قدرات عقلية مثل القدرة على الاستدلال والقدرة المكانية والقدرة العددية، بينما قد تعتمد كتابة موضوع تعبيري على القدرة اللغوية والقدرة اللفظية والقدرة على الطلاقة اللفظية أكثر من غيرها من القدرات ... وهكذا.

وبناءً على هذه القدرات العقلية الأولية أنكر ثرستون العامل العام واكتفى فقط بالعوامل الطائفية وحدها، بسبب عدم عثوره على ارتباط عام بين الاختبارات التي طبقها في دراساته، ولكن بعد ذلك ثبت له وجود ارتباطات عالية بين العوامل الطائفية ، والتي فسرت علي أنها صفة شائعة بين تلك العوامل هو **عامل العوامل**، وهذا هو المعنى الذي أعطاه ثرستون للعامل العام.

تعقيب: من نتائج البحوث والدراسات التي أجريت لتفسير التكوين العقلي الإنساني، أمكن التوصل إلى العوامل العقلية الثلاثة التالية:

أ- هو العامل العام: الذي يشمل جميع الاختبارات التي تقيس النشاط العقلي المعرفي، وهو الذي يمكن أن نوجد بينه وبين الذكاء العام.

ب- هو العامل الطائفي: يمثل هذا العامل الصفة التي تشترك فيها مجموعة من الاختبارات التي تقيس بعض جوانب النشاط العقلي، ولا تشترك فيها بقية الاختبارات، بمعنى آخر أنها خاصة بطائفة معينة من النشاط العقلي.

ج- هو العامل الخاص: هذا العامل يمثل الصفة التي يتفرد بها اختبار معين ولا يوجد في الاختبارات الأخرى.

3-التنظيم الهرمي للقدرات العقلية :

من خلال تلك العوامل العقلية الثلاثة يمكن أن نتصورها على شكل تنظيم هرمي حيث تأتي في قمة الهرم القدرة العقلية العامة، وهي التي تشترك في جميع العمليات العقلية المعرفية، ثم تأتي بعدها القدرات الطائفية أما القدرات الخاصة فهي تقع عند قاعدة هذا التنظيم الهرمي.

3-1- النموذج الهرمي للقدرات العقلية:

من أكثر النماذج الهرمية شيوعاً هو نموذج سيرت بيرت C. Burt و فيليب فرنون Vernon P.E وكذلك نموذج ريموند كاتل R. W. Cattel ومن النماذج التي كانت لها صدى في المنطقة العربية هو نموذج فؤاد البهي السيد. ونظراً لوجود تشابه بين نموذج كل من بيرت وفيرنون فإننا سوف نكتفي بأن نستعرض أحدهما بالشرح وليكن نموذج فيرنون .

أ – نموذج فيرنون Vernon Vernon's Hierarchical Theory of Intelligence

نجد أن في قمة النموذج الهرمي للتنظيم العقلي لفيرنون، العامل g يليه مجموعتان من العوامل الطائفية هما القدرة التحصيلية (أو الاستعداد اللفظي التعليمي) والقدرة المهنية (أو الاستعداد العملي الميكانيكي).

وتتقسم هذه العوامل الطائفية الكبرى إلى العوامل الفرعية الخاصة اللفظية والعددية والمكانية وغيرها من القدرات. وكذلك ينقسم العامل العملي الميكانيكي إلى عدة عوامل فرعية منها المعلومات الميكانيكية وعامل السرعة اليدوية والعامل المكاني. وهكذا يمكن أن نحلل هذه العوامل الطائفية الصغرى إلى عوامل أبسط منها. ويوجد في أدنى هذا النموذج الهرمي العوامل الخاصة. ويمكن تمثيل النموذج الهرمي لفيرنون وكأنه عبارة عن شجرة مقلوبة.

ب- التحقيق التجريبي لنموذج فيرنون الهرمي: أجرى فيرنون أكثر من (60) تحليلاً عاملياً باستخدام أسلوب التحليل العاملي وتدوير المحاور لحذف التشبعات العاملية السالبة. وقد تأكد فيرنون في ثمانية بحوث عاملية أن العامل العام يشمل أكثر من ضعف التباين الذي تشمله العوامل الطائفية الأخرى مجتمعة، وكان من أبرز نتائج تحليله لعدد (13) اختبار طبقت على (1000) مجند في الجيش البريطاني تأكد النموذج الهرمي الذي أشرنا إليه في الشكل السابق.

4 – نظرية جيلفورد Guilford (نظرية التصنيف الثلاثي)

عرض العالم الأمريكي جيلفورد (1955) تصنيفاً ثلاثياً للقدرات العقلية تمت اسم بنية العقل Structure of Intellect. وتعرّض تصنيفه الثلاثي إلى عدة تطورات حتى أصبح في صورته الحالية منذ آخر تطور له عام 1967. ويصنف جيلفورد العوامل العقلية لأسس ثلاثة هي:

4-1- نوع العمليات (الإجراءات) Operations: ويزودنا هذا النوع من العمليات بخمس مجموعات أساسية للقدرات العقلية هي :

أ- عوامل المعرفة Cognition Factors: وتتعلق بقدرة الفرد على الكشف عن شئ ما أو التعرف عليه.

ب- عوامل التذكر Memeory Factors: وتعني قدرة الفرد على الاحتفاظ بما عرفه.

ج- عوامل التفكير التقاربي Convergent Thinking Factors: بمقتضى هذا النوع من التفكير تؤدي المعلومات التي لدينا إلى إجابة واحدة، تعرف بأنها أفضل الإجابات.

د- عوامل التفكير التباعدي Divergent Thinking Factors: ويتضمن هذا النوع من التفكير في اتجاهات مختلفة، ومن ثم تصل إلى إجابات متنوعة وليست إجابة واحدة كما في التفكير التقاربي.

ه- التقويم: وتشير إلى قدرة الفرد على إصدار أحكام على المعلومات أو الأفكار أو الحلول في ضوء معايير محددة، من خلال التمييز بين الصحيح والخاطئ أو المناسب وغير المناسب، وتقدير قيمة البدائل المختلفة واختيار أفضلها. ويتطلب هذا النوع من العمليات استخدام معايير منطقية أو خبرات سابقة للحكم على دقة المعلومات، واتساقها، وملاءمتها للموقف، مما يجعله عملية عقلية عليا ترتبط بالنقد والتحليل واتخاذ القرار.

4-2- بعد المحتوى أو المضمون Content: وفي هذا البعد تصنف العوامل العقلية حسب نوع المحتوى أو المادة المتضمنة في العمليات العقلية، ويشتمل هذا البعد على أربعة أنواع وهي :

أ- محتوى الأشكال Figural: وتشتمل على المعلومات التي لها خصائص عيانية محسوسة، قد تكون مدركات حسية بصرية (أي لها أشكال أو أوزان أو أحجام أو ألوان... الخ) وقد تكون مدركات حسية سمعية أي أصوات أو كلام أو إيقاعات .... الخ)، أو مدركات حسية لمسية، أو تتعلق بالإحساسات الحركية.

ب- محتوى الرمز Symbolic: ويشتمل هذا البعد على المعلومات التي ليست لها خصائص حسية كما في محتوى الأشكال، وإنما يتميز بخصائص مجردة، ولا يلعب فيها عنصر المعنى دوراً كبيراً، ومن أمثلتها الأرقام والحروف.

ج- محتوى المعاني (السيانتي) Semantic: وهي نوع من المعلومات يتمثل فيه الأفكار والمعنى، وتتشكل عادة في صورة لغوية.

د- المحتوى السلوكي Behavioral: ويعني الاستدلال من ظواهر السلوك على مشاعر الفرد وذكائه الاجتماعي.

4-3- بعد النواتج Products: وهو البعد الثالث في تصنيف جيلفورد، ويقصد به الطريقة التي يتم بها التعامل مع بعض المحتويات (سواء كانت أشكالاً، أو رموزاً أو معاني أو مواقف سلوكية) وهي عبارة عن ستة أنواع من النواتج وهي:

أ- الوحدات Units : وهي تمثل أبسط ما يمكن أن تحلل إليه معلومات المحتوى، كما تتميز بالاستقلال النسبي. وعلى سبيل المثال فإن وحدة المعاني اللغوية هي الكلمة ووحدة محتوى الأشكال هو الشكل الواحد على أرضية ومحتوى الرمزى ممثل الأصوات الكلامية أو الأرقام العددية.

ب- الفئات Classes (Seats) : ويقصد بها مجموعة من الوحدات تجمع بينها خصائص مشتركة.

ج- العلاقات Relations : وهي ما يربط الوحدات ببعضها ببعض كعلاقات التشابه أو الاختلاف.

د- المنظومات Systems: وتدل المنظومة على مجموعة من العلاقات المنظمة المتداخلة والتي تربط بين أجزاء بحيث يتكون منها مركب معقد، ومن أمثلة ذلك المسألة الحسابية والفقرة اللغوية.

ه- التحويلات Transformation: ويقصد بها التعديلات أو التغييرات التي تطرأ على معلومات الاختبار سواء من حيث الشكل أو البنية أو التركيب أو المعنى.

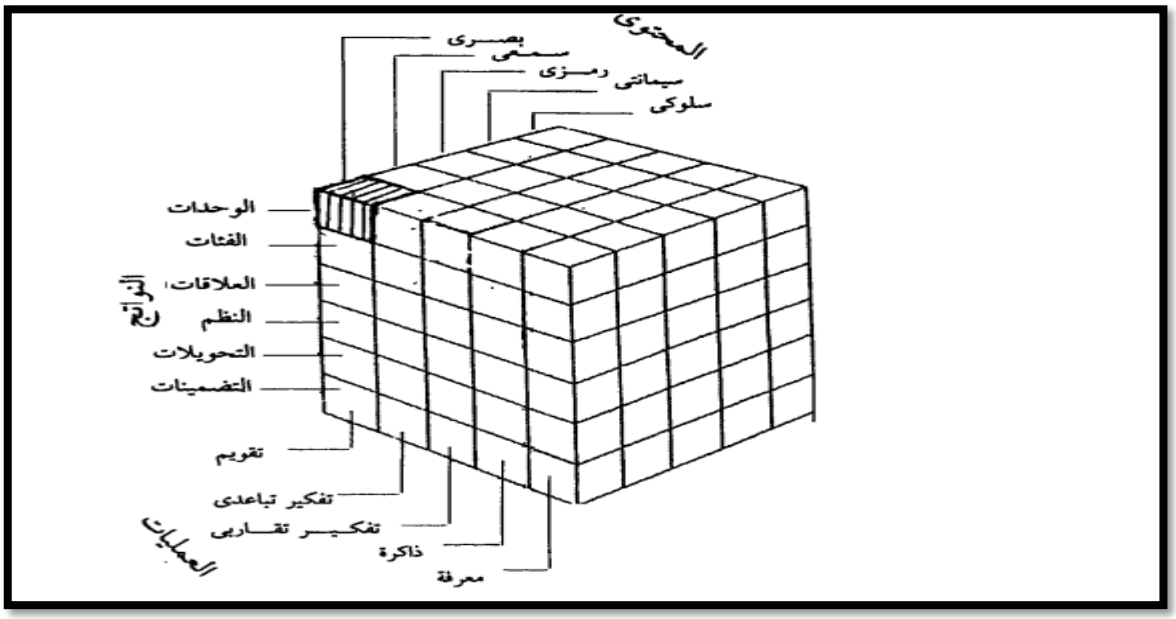
و- التضمينات (اللوازم) Implication: وهي ما يتوقعه المفحوص أو يستدل عليه من المعلومات المتاحة في الاختبار، ويعني ذلك أن هذه التضمينات تتطلب أن أي وحدة من وحدات المعلومات تؤدي إلى أخرى بطريقة طبيعية.

ويصور لنا جيلفورد الوضع النهائي لنموذجه في تكوين العقل في شكل ثلاثي الأبعاد، حيث تمثل العمليات العقلية الخمسة بعده الأول، والمحتويات الأربع بعده الثاني والنواتج الست بعده الثالث. وفي هذا النموذج العقلي يدل تفاعل كل عملية معينة مع محتوى معين مع ناتج معين على قدرة عقلية معينة، وعلى ذلك يمكن حساب عدد العوامل العقلية كالتالي:

عدد العوامل العقلية = (5) عمليات x (4) محتويات x (6) نواتج = 120 عاملاً

ويمثل الشكل التالي نموذج مكعب جيلفورد.

شكل 03 يوضح نموذج جيلفورد ثلاثي الأبعاد أو الأوجه



(علام، 2000، ص360)

5- نموذج كاتل-هورن-كارول (CHC):

يعد نموذج كاتل-هورن-كارول (Cattell–Horn–Carroll Model: CHC) من أكثر النماذج شمولًا وتأثيرًا في مجال قياس الذكاء وفهم البنية الهرمية للقدرات العقلية. يجمع هذا النموذج بين أعمال ريموند كاتل وجون هورن حول الذكاء السائل (Gf) والمتبلور(Gc) ، وبين نظرية جون كارول في القدرات المعرفية ذات الثلاثة طبقات. وفقًا لهذا النموذج، ينظم الذكاء في ثلاثة مستويات هرمية: في القمة يوجد العامل العام للذكاء(g) ، يليه مستوى القدرات الواسعة كالذكاء السائل، المعرفة المتبلورة، الذاكرة العاملة، المعالجة البصرية، المعالجة السمعية، سرعة المعالجة، والقدرات الكمية، ثم مستوى القدرات الضيقة التي تمثل مهارات أكثر تخصصًا داخل كل قدرة واسعة. وقد ساهم نموذج CHC بشكل كبير في تطوير وتفسير اختبارات الذكاء الحديثة كاختبار وكسلر واختبار وودكوك-جونسون، حيث يوفر إطارًا نظريًا دقيقًا لتحليل الأداء المعرفي وتفسير نقاط الاختبارات الفردية. وبفضل شموليته وقابليته للتطبيق, أصبح نموذج CHC المرجع الأكاديمي الرئيس في فهم وتقييم القدرات العقلية لدى الأفراد في مختلف الأعمار والسياقات التربوية والنفسية.

6- نظرية الذكاءات المتعددة (جاردنر):

تعد نظرية الذكاءات المتعددة التي قدّمها عالم النفس الأمريكي هوارد جاردنر في عام 1983 من النظريات الحديثة التي أحدثت تحولًا جوهريًا في فهم طبيعة الذكاء الإنساني. رفض جاردنر التصور التقليدي للذكاء باعتباره قدرة عقلية واحدة قابلة للقياس باختبارات معيارية، واقترح بدلاً من ذلك أن الذكاء يتكون من مجموعة من القدرات المتمايزة التي تعمل بشكل نسبي مستقل عن بعضها. وقد حدد جاردنر في نسخته الأولى سبعة ذكاءات، ثم أضاف لاحقًا الذكاء الطبيعي، لتصبح الأنواع الأساسية ثمانية ذكاءات: اللغوي، المنطقي–الرياضي، البصري–المكاني، الجسمي–الحركي، الموسيقي، الاجتماعي (التفاعلي)، الشخصي (الذاتي)، والطبيعي. تؤكد النظرية على أن كل فرد يمتلك مزيجًا فريدًا من هذه الذكاءات بنسب متفاوتة، مما يعزز أهمية التعليم المتنوع الذي يراعي الفروق الفردية ويحفّز أنماط التعلّم المختلفة. وبفضل تأثيرها الواسع، أصبحت هذه النظرية أحد الأسس المعاصرة في تطوير البرامج التربوية والمناهج التعليمية التي تركز على تمكين المتعلمين من توظيف نقاط قوتهم الذهنية في التعلم والحياة العملية.

سابعاً- أدوات قيــاس الذكــاء العــام:

يمكن تصنيف اختبارات الذكاء العام إلى نوعين وفقاً لطريقة التطبيق .فالاختبارات التي تطبق على فرد واحد في وقت واحد تسمى الاختبارات الفردية (Individual Intelligence Tests). أما الاختبارات التي تطبق على عدد كبير من الأفراد في وقت واحد فتسمى الاختبارات الجماعية (Group Intelligence Tests)، ومعظم الاختبارات الفردية كمقياس ستانفورد بينيه (Stanford - Binet)، ومقاييس ويكسلر (Wechsler)، تحتاج إلى شخص مدرب على تطبيقها وتفسير نتائجها حيث يهتم بكيفية استجابة الفرد أو المفحوص. لذلك تطبق عادة الاختبارات الفردية بواسطة الأخصائيين النفسيين بالمدارس والجامعات والمؤسسات حيث يُستفاد من نتائجها في التشخيص الكلينيكي، واتخاذ قرارات مهمة كتشخيص حالات الضعف العقلي. أما الاختبارات الجماعية فإنها تُعد أفضل وسائل تقييم القدرات العقلية لعدد كبير من المختبرين في هذه المؤسسات بغرض التوجيه التعليمي والمهني.

ونظراً لأنه يصعب بالطبع تناول جميع اختبارات الذكاء العام المستخدمة في وقتنا الحاضر، فإننا سوف نوضح بإيجاز أمثلة لبعض الاختبارات الفردية والجماعية الشائعة الاستخدام.

1- اختبارات الذكاء الفردية:

1-1- مقياس ستانفورد بينيه للذكاء Stanford - Binet Intelligence Scale (SBIS):

ويعد مقياس ستانفورد - بينيه كما ذكرنا امتداداً للمقياس بينيه (Binet) الأصلي. وقد نشرت الطبعة الأولى لهذا المقياس عام 1916 في الولايات المتحدة الأمريكية وأجريت عليه عدة تعديلات في الأعوام 1937، 1960، 1972، 1984. وسوف نلقي الضوء فيما يلي على هذا المقياس:

أ- مقياس ستانفورد - بينيه (طبعة عام 1916):

قام تيرمان (Terman) بجامعة ستانفورد الأمريكية بترجمة مقياس بينيه - سيمون (Binet - Semon) وأجرى تطبيقه على عينة تشتمل على 1000 طفل و 400 فرد راشد. وكما في مقياس بينيه الأصلي اعتمد تيرمان على ميزان العمر الزمني (Age Scale)، حيث قام بتجميع الاختبارات التي يشتمل عليها المقياس في مستويات عمرية. واشتمل المقياس على مفردات منتقاة من مقياس بينيه وإضافة مفردات جديدة تُمثّل عينة واسعة من المهام التي تتطلب ذكاء، ولكن هذه المهام لا تعتمد على خبرات مدرسية معينة، وأحد محكات انتقاء مفردة في اختبار فرعي هو تزايد النسبة المئوية من الأطفال في المستويات العمرية المتتالية الذين يستطيعون أن يجيبوا عنها إجابة صحيحة أي التمايز العمري (Age

Differentiation)، وتوضع المفردة عندئذ في المستوى العمري للاختبار الفرعي الأكثر ملاءمة له (علام، 2000، ص370).

ومن المحكات الأخرى لانتقاء المفردات الارتباط بين درجات كل مفردة والدرجات الكلية في المقياس .وهذا يعني أن كل مفردة تقيس ما يتعلق بالمقياس ككل، وهو ما يشار إليه باسم تجانس المفردات (Item Homogeneity). ويعتمد العمر العقلي ((Mental Age (MA)، ونسبة الذكاء (Intelligent Quotient (IQ))، في القياس على عدد الاختبارات الفرعية التي يجتازها المفحوص في المستويات العمرية المختلفة. وتحسب نسبة الذكاء باستخدام الصيغة:

$$\text{نسبة الذكاء } (\boldsymbol{IQ}) = \frac{\text{العمر العقلي}}{\text{العمر الزمني}} \times \mathbf{100}$$

وذلك بعد تحويل كل من العمر العقلي والعمر الزمني إلى شهور (علام، 2000، ص371).

ب- النسخة الأحدث من اختبار ستانفورد–بينيه للذكاء Stanford-Binet Intelligence Scales, Fifth Edition (SB-5):

النسخة الأحدث لمقاييس الذكاء ستانفورد–بينيه، الإصدار الخامس (SB-5)، التي طوّرها غيل هـ. رويد، تُعد من أكثر أدوات قياس القدرات العقلية موثوقية عبر مراحل العمر المختلفة، حيث تعتمد على عينة معيارية واسعة تضم 4,800 فرد، وتمت مواءمتها مع اختبار بندر–جشطالت البصري-الحركي (الطبعة الثانية) وربطها باختبارات وودكوك-جونسون للإنجاز. كما تتوفر نسخة مخصّصة للأطفال في مرحلة ما قبل المدرسة (Early SB-5) مبنية على عينة معيارية من 1,800 طفل، وبكلفة أقل. يقدم الاختبار درجات معيارية للاختبارات الفرعية، بالإضافة إلى مؤشرات تركيبية تشمل: الذكاء اللفظي، الذكاء غير اللفظي، الذكاء المختصر، معامل الذكاء الكلي، وخمسة عوامل معرفية رئيسية هي: الاستدلال السائل، المعرفة، القدرة العددية، المعالجة البصرية-المكانية، والذاكرة العاملة، مع إمكانية الحصول على درجات حساسة للتغيرات المرجعية عند الحاجة. يتألف الاختبار من عشرة اختبارات فرعية تُطبق فرديًا، ويستغرق كل منها نحو خمس دقائق، ويُطبّق على الفئات العمرية من 2 إلى أكثر من 85 عامًا في النسخة الأساسية، ومن 2 إلى 5 سنوات و11 شهرًا في النسخة الكاملة الخاصة بالأطفال، ومن 6 سنوات إلى 7 سنوات و3 أشهر في النسخة المختصرة. ويتطلب تطبيقه مستوى تأهيليًا مهنيًا متقدمًا (مستوى **C**) (Roid, n.d).

2-1- مقاييس ويكسلر للذكاء (Wechsler Intelligence Scales):

طوَّر ويكسلر مجموعة من مقاييس الذكاء الفردية انطلاقًا من رؤية تختلف عن تلك التي اعتمد عليها مقياس ستانفورد–بينيه، إذ سعى إلى معالجة أوجه القصور التي شابت هذا المقياس، ولاسيما ما يتعلق بضعف ملاءمته لتقييم ذكاء الراشدين من حيث طبيعة بنوده ومعاييره، واعتماده على درجة كلية واحدة تُعد مؤشرًا عامًا للذكاء، بالإضافة إلى اعتماده في طبعاته الأولى على مفهومي العمر العقلي ونسبة الذكاء؛ وهما مفهومان تعرَّضا لنقد واسع خاصة في سياق قياس ذكاء البالغين. كما أُخذ على المقياس تركيزه على عامل السرعة في الإجابة عن بنود محددة بزمن، وهو ما قد لا يتناسب مع قدرات الراشدين وخصائصهم النفسية والمعرفية. استند ويكسلر في بناء مقياسه إلى تصور للذكاء باعتباره منظومة مترابطة من القدرات وليس قدرة عامة واحدة، وأدرج مكونات لفظية وأخرى أدائية للتقليل من تأثير اللغة والثقافة والخبرة التعليمية، بهدف إتاحة تقييم أكثر عدالة للأفراد ذوي الصعوبات اللغوية أو محدودية الخبرة اللفظية (Wechsler, 1939). ومن خلال هذا النهج أصبح بالإمكان الحصول على ثلاث درجات رئيسية تشمل القدرة اللفظية، والقدرة الأدائية، والذكاء العام، بحيث قد يحصل الفرد على الدرجة نفسها بطرق مختلفة تبعًا لطبيعة المفردات التي نجح أو أخفق فيها.

أتاح مقياس ويكسلر إمكانية تشخيص أسلوب تفكير الفرد بدرجة أدق من مقياس ستانفورد–بينيه، كما أتاح ملاحظة السلوك أثناء المهام العملية، الأمر الذي يوفّر بيانات نوعية مهمة في السياق الإكلينيكي. فبينما تمكّن المهام اللفظية الأخصائي من متابعة استجابات المفحوص اللغوية، تتيح المهام العملية – بما تتطلبه من تركيز وانتباه واستغراق زمني أطول – فهمًا أفضل لأساليب حل المشكلات لدى المفحوص، مما يجعل مقياس ويكسلر أداة فعّالة في التشخيص السريري.

ظهر أول إصدار لمقياس ويكسلر للراشدين عام 1939 تحت اسم "مقياس ويكسلر–بيلفو للذكاء"، نسبة إلى مستشفى بيلفو للأمراض النفسية. غير أن العينة المعيارية المستخدمة آنذاك لم تكن ممثلة للمجتمع الأمريكي، ما دفع ويكسلر إلى إعادة تقنينه عام 1955 وإطلاق اسم "مقياس ويكسلر لذكاء الراشدين (WAIS)"، ثم جرى تقنينه مجددًا ووضع معايير جديدة له عام 1981 تحت مسمى (WAIS-R). كما قدم ويكسلر عام 1949 مقياسًا لذكاء الأطفال ضمن الفئة العمرية 5–15 سنة (WISC)، ثم تبعه في عام 1967 بإصدار مقياس لمرحلة ما قبل المدرسة والمرحلة الابتدائية (WPPSI) للفئة العمرية 4–6 سنوات (علام، 2000، ص375).

أ- النسخة الأحدث لمقياس وكسلر Wechsler Adult Intelligence Scale – Fifth Edition (WAIS-) (5) (2024):

يعد مقياس وكسلر لذكاء الراشدين – النسخة الخامسة (WAIS-V) أحدث إصدار من سلسلة مقاييس وكسلر المخصّصة لقياس القدرات العقلية لدى الأفراد البالغين. وقد طوّر هذا المقياس استنادًا إلى أحدث النظريات النفسية والأبحاث في مجال علم النفس العصبي، بهدف توفير تقييم شامل ودقيق لمستوى الأداء العقلي عبر مجموعة من المجالات المعرفية الأساسية. ويقيس المقياس خمس قدرات رئيسية تتمثل في :الفهم اللفظي، والاستدلال البصري–المكاني، والاستدلال السائل، والذاكرة العاملة، وسرعة المعالجة، كما يوفر درجة عامة للذكاء (FSIQ) تعد مؤشراً شاملاً لمستوى القدرة العقلية الكلية لدى الفرد. يمتاز (WAIS-V) بتحديث المعايير، وتحسين بنود الاختبار، وإضافة أدوات رقمية لتسهيل التطبيق والتصحيح، مما يجعله مناسبًا للاستخدام في السياقات الإكلينيكية والتعليمية والبحثية. ويُستخدم هذا المقياس للكشف عن جوانب القوة والضعف المعرفية، وتشخيص الاضطرابات النمائية والمعرفية مثل صعوبات التعلم والخرف وإصابات الدماغ، بالإضافة إلى الاستعانة به في التقييمات المهنية والإرشادية. ورغم دقته العالية وموثوقيته، إلا أنه يُعد جزءًا من منظومة تشخيصية متكاملة يجب أن تشمل مصادر معلومات أخرى للوصول إلى تقييم شامل للوظائف العقلية للفرد (Jacobsen & George, 2024).

1-3- مقارنة بين مقاييس ويكسلر ومقياس ستانفورد-بينيه:

في مجال القياس العقلي يُعد كلٌّ من مقياس ستانفورد–بينيه ومقياس وكسلر من أبرز أدوات قياس الذكاء عبر التاريخ، إلا أن لكل منهما فلسفة وبنية تختلف عن الآخر. فمقياس ستانفورد–بينيه ظهر أولًا وركز في بداياته على مفهوم الذكاء العام من خلال الاعتماد على العمر العقلي ونسبة الذكاء مع تركيز أكبر على القدرات اللفظية، وقد اعتُبر مناسبًا بشكل خاص للأطفال وإن كانت له حدود في قياس ذكاء الراشدين وصعوبات في التعامل مع الفروق الثقافية واللغوية. أما مقياس وكسلر فقد جاء لاحقًا استجابة لهذه المحدوديات، إذ قدم تصورًا للذكاء بوصفه مجموعة من القدرات المتداخلة وقسم الأداء إلى مجالات لفظية وأدائية (غير لفظية)، مما سمح بقياس أوسع وأكثر واقعية للوظائف العقلية لدى الراشدين والأطفال على حد سواء. كما يوفر مقياس وكسلر درجات متعددة تعكس جوانب معرفية متباينة إلى جانب ملاحظة السلوك أثناء الأداء الأمر الذي يمنح الأخصائي النفسي معلومات

تشخيصية أعمق. وبذلك ينظر إلى مقياس ستانفورد–بينيه على أنه أساس تقليدي مهم في القياس العقلي بينما يمثل مقياس وكسلر تطورًا نوعيًا يعكس فهمًا أكثر شمولًا للقدرات المعرفية.

2- اختبارات الذكـاء الجماعية:

اختبارات الذكاء الجماعية ظهرت تاريخيًا لتلبية الحاجة إلى تصنيف وانتقاء الجنود، ثم توسع استخدامها لتشمل ميادين مدنية متعددة. وتمتاز هذه الاختبارات بإمكانية تطبيقها على مجموعات كبيرة في وقت واحد مما يقلل من الزمن والتكلفة والموارد اللازمة مقارنة باختبارات الذكاء الفردية. ومن أبرز النماذج الأولية لهذه الاختبارات: اختبار الجيش ألفا، واختبار الجيش بيتا، واختبار التصنيف العام للقوات المسلحة (AGCT)، والاختبار البديل له (AFQT). تقدم هذه الاختبارات غالبًا في صورة ورقية وتعتمد على مفردات من نوع الاختيار من متعدد ما يسهم في سهولة وموضوعية التصحيح سواء يدويًا أو باستخدام الحاسوب ودون الحاجة إلى فاحصين ذوي خبرة كبيرة أو أدوات مكلفة. ومع ذلك فإن الطابع اللفظي لهذه المفردات يجعلها غير مناسبة للأطفال والأميين والأفراد ذوي الإعاقات السمعية أو البصرية الأمر الذي استدعى تطوير نسخ غير لفظية من هذه المقاييس.
وتُستخدم الاختبارات الجماعية عادة في المؤسسات التعليمية والقطاعات المهنية والمنشآت الصناعية والعسكرية، إضافة إلى استخدامها الواسع في الأبحاث النفسية والتربوية. وتساعد نتائجها في عمليات اتخاذ قرارات انتقاء العاملين (Personnel Selection)، وأغراض التصفية (Screening)، والتوجيه المهني (Vocational Guidance)، وتحديد القدرات العقلية والاستعدادات الخاصة. غير أن هذه الاختبارات تفتقر إلى التفاعل المباشر بين الفاحص والمفحوص، فلا تسمح بتحفيز الأفراد أو ملاحظة سلوكهم كما هو الحال في الاختبارات الفردية، ما يفترض مستوى عالٍ من الدافعية الذاتية والتعاون لدى المفحوصين أثناء تأديتها (علام، 2000، ص389).

2-1- أمثلة لبعض اختبارات الذكاء الجماعية:

نظراً لتعدد أنواع اختبارات الذكاء الجماعية وتنوعها حيث يوجد عدد كبير من هذه الاختبارات تناسب جميع المستويات العمرية من الحضانة إلى سن الرشد، فإنه يصعب حصر هذه الاختبارات في هذه المحاضرة، وإنما يمكن الرجوع إلى الكتاب السنوي للقياس النفسي (Buros, 1998) للاطلاع على أحدث هذه الاختبارات بالتفصيل، وكذلك أدلة الاختبارات في البيئة العربية.

غير أننا سوف نقدم في الجزء التالي أمثلة لبعض أنواع الاختبارات الأكثر استخداماً في قياس الذكاء، وعلى الرغم من أنها مقننة في البيئة الأمريكية إلا أن بعضاً منها تم تعريبه وتقنينه في البيئة المصرية وفي بيئات عربية أخرى.

2-2- اختبــارات الذكــاء الجماعية متعددة المستويات:

أ- اختبار كاليفورنيا للنضج العقلي (California Test of Mental Maturity (CTMM)-1963): يضم هذا الاختبار سلسلة مكوّنة من 12 اختبارًا فرعيًا تقيس مجموعة من العمليات العقلية، كالذاكرة والعلاقات المكانية والاستدلال المنطقي والعددي والمفاهيم اللفظية. ويتوافر الاختبار في عدة صيغ متتابعة عبر ستة مستويات عمرية تبدأ من مرحلة الحضانة وصولًا إلى مرحلة الرشد، بحيث يُخصص لكل مستوى كتيب مستقل. يسمح الاختبار بالحصول على ثلاث درجات كلية تشمل الدرجة اللفظية وغير اللفظية والدرجة الكلية، ويستغرق تطبيقه ما بين 48 و81 دقيقة بحسب المستوى الأمر الذي أدى لاحقًا إلى تطوير نسخة مختصرة منه لتقليل زمن التطبيق.

ب- اختبار القدرات المعرفية (Cognitive Abilities Test (CAT)-1978): يعد هذا الاختبار كذلك متعدد المستويات، حيث يتضمن بطاريتين: الأولى مخصصة للصفوف من الحضانة حتى الصف الثالث الابتدائي والثانية تشمل الصفوف من الثالث إلى نهاية المرحلة الثانوية. تحتوي بطارية المرحلة الأولى على صيغتين لكل مستوى وكل منهما تتضمن أربعة اختبارات فرعية تقيس المفاهيم العلائقية والعمليات العقلية المتعددة والمفاهيم الكمية والكلمات الخفية ويستغرق كل اختبار فرعي حوالي 15 دقيقة تُطبق في أربع جلسات منفصلة. تفسر الدرجات وفق معايير نسبة الذكاء الانحرافية والمئينيات. أما البطارية الثانية فتنقسم إلى ثلاث مجموعات محتوى: لفظي وكمي وغير لفظي ويشمل كل منها اختبارات فرعية مثل إكمال الجمل، التصنيف اللفظي، متشابهات لفظية، علاقات كمية، تكوين معادلات، معالجة أعداد ورموز كمية، إضافة إلى أشكال هندسية ومكانية في المحتوى غير اللفظي. وتُفسر نتائج البطارية وفق العمر المكافئ، الرتب المئينية، والتساعيات المعيارية. وتجدر الإشارة إلى أن هذا الاختبار يمثل تطويرًا لاختبار لورج وثورندايك (1963).

ج-اختبارات أوتيس–لينون للقدرة العقلية العامة (Otis - Lenon Mental Ability Tests-1967): تتكون هذه الاختبارات من صيغة واحدة تشمل مجموعة متنوعة من المفردات تقدم في شكل حلزوني شامل لقياس القدرة العقلية العامة. يضم الاختبار ستة مستويات تمتد من الصف الأول الابتدائي حتى نهاية المرحلة الثانوية، ويستغرق تطبيقه ما بين 30 و45 دقيقة حسب المستوى. يعطي الاختبار

درجة عامة واحدة ويتم تفسير النتائج باستخدام نسب الذكاء الانحرافية والمئينيات والتساعيات المعيارية للعمر والصف كما يوفر تقديرًا للعمر العقلي في المستويات الأولى.

د- اختبار الاستعداد للدراسة الجامعية ((Scholastic Assessment Test (SAT)): يعد اختبار (SAT) أحد أشهر الاختبارات الدولية المستخدمة لقياس جاهزية الطلاب للالتحاق بالتعليم الجامعي، ويعتمد عليه بشكل واسع في إجراءات القبول بالجامعات والكليات داخل الولايات المتحدة وخارجها. يقيس الاختبار مجموعة من القدرات الأكاديمية الأساسية التي يكتسب معظمها خلال المرحلة الثانوية، وخاصة مهارات القراءة والفهم، والكتابة واللغة، والرياضيات، مما يساعد على تقييم قدرة الطالب على تحليل المعلومات وتطبيق المعارف في سياقات عملية. يتقدم معظم الطلاب لاختبار (SAT) خلال السنة قبل الأخيرة أو الأخيرة من المرحلة الثانوية، كما أن العديد من المؤسسات التعليمية حول العالم، بما فيها المدارس الدولية وبرامج التعليم الأمريكي بالخارج، تعتمد نتائجه لقياس الاستعداد الجامعي والمهني. يقدم الاختبار في مواعيد محددة ويُتاح للطلاب التسجيل ذاتياً للاختبار في عطلات نهاية الأسبوع، بينما توفر بعض المدارس جلسات اختبار خلال اليوم الدراسي لإتاحة فرصة أوسع للطلاب. ويعد (SAT) اختباراً صادراً عن College Board ويهدف إلى توفير مؤشر معياري يساعد الجامعات في اتخاذ قرارات القبول، وتقديم بيانات موضوعية عن مدى استعداد الطلبة لمتابعة التعليم العالي (U.S. Department of Defense Education Activity, n.d., para. 1).

ج- اختبار القبول في الدراسات العليا ((Graduate Record Examination (GRE)): اختبار (GRE) أو امتحان سجل الدراسات العليا هو أحد أهم الاختبارات المعيارية الدولية المستخدمة للقبول في برامج الدراسات العليا، خصوصًا في الولايات المتحدة وكندا وأستراليا. يُشرف على الاختبار مركز الاختبارات التعليمية ETS، وهو الجهة نفسها التي تدير اختبار (TOEFL). يقيس الاختبار القدرات العقلية والأكاديمية للطلبة في ثلاثة مجالات رئيسية :التحليل الكتابي، القدرة اللفظية، والقدرة الكمية. يُقدَّم الاختبار بصيغتين: عبر الكمبيوتر وهي الأكثر شيوعًا أو بصيغة ورقية في أماكن لا تتوفر فيها النسخة الإلكترونية ويستغرق حوالي 3ساعات و45 دقيقة. يعتمد نظام الدرجات على مدى يتراوح بين 130–170 لكل من القسمين اللفظي والكمي، بينما يقيم القسم التحليلي من 0 إلى 6 درجات بنصف نقطة. يوجد أيضًا اختبارات GRE تخصصية في مجالات كالكيمياء والرياضيات والفيزياء وعلم النفس، تقدم على شكل ورقي ثلاث مرات سنويًا. تتراوح تكلفة اختبار GRE العام بين 205 و255 دولارًا، بينما تبلغ تكلفة اختبار التخصص حوالي 150دولارًا. يُعد GRE خيارًا أساسيًا

للراغبين في متابعة دراسات عليا في تخصصات متنوعة، بدءًا من العلوم والهندسة وحتى العلوم الإنسانية وإدارة الأعمال، خاصة وأن العديد من الجامعات تشترطه أو تعتبره عنصرًا داعمًا في ملف الترشح الأكاديمي (Tippett, 2025).

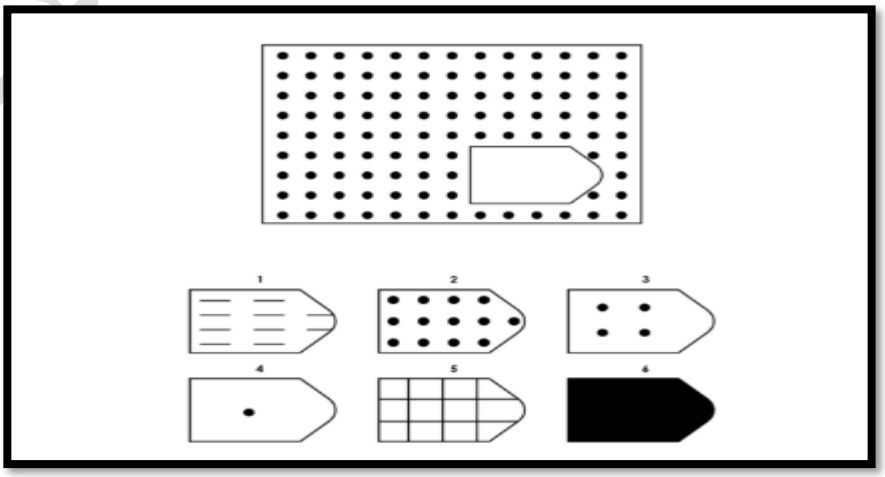
3- اختبارات الذكاء الجماعية غير اللفظية:

تستخدم اختبارات الذكاء الجماعية غير اللفظية (Non-Verbal Intelligence Tests) لقياس القدرات العقلية لدى الأفراد الذين لا يمتلكون مهارات قراءة وكتابة كافية، أو أولئك الذين يعانون من صعوبات لغوية وثقافية أو إعاقات جسدية وكذلك الأطفال في سن صغيرة. وتمتاز هذه الاختبارات بأنها لا تعتمد على اللغة، ويمكن تنفيذها باستخدام الورقة والقلم أو من خلال أداء مهام عملية، وفي بعض الحالات يمكن تطبيقها دون استخدام أي تعليمات لفظية، ومن أمثلة على هذه الاختبارات:

3-1- اختبار المصفوفات المتتابعة لرافن 1947 – (Raven Progressive Matrices):

طوَّر عالم النفس جون رافن (J. Raven) بالتعاون مع عالم الوراثة ليونيل بنروز (L. Penrose) هذا الاختبار في إنجلترا ويُعد أحد أشهر أدوات القياس غير اللفظية المستخدمة لتقدير الذكاء العام خصوصًا في البيئات التعليمية والعسكرية، حيث صُمم لقياس القدرة العقلية دون تأثير الخلفية التعليمية والثقافية. يتكون الاختبار من 60 مصفوفة هندسية يُحذف أحد أجزائها وعلى الفرد اختيار الجزء الناقص من بين بدائل عددها ستة أو ثمانية. وتنظم المواد في خمس مجموعات متدرجة الصعوبة تحتوي كل منها على 12 مفردة وتتطلب الإجابة إدراك العلاقات المنطقية والتماثل والتتابع الهندسي ويوضح شكل التالي مثالاً لهذه المصفوفات (Jensen, 1980):

شكل 04 يوضح أحد مفردات اختبار رافن الملون



يتوافر الاختبار بثلاث صيغ:

- صيغة قياسية للراشدين وتلاميذ المدارس.
- صيغة ملونة موجهة للأطفال قبل سن المدرسة والأفراد ذوي الإعاقات الذهنية.
- وصيغة متقدمة تناسب المستويات الأكاديمية الأعلى.

يصنف الاختبار ضمن اختبارات القوة الذهنية؛ إذ لا يفرض زمنًا محددًا، لكنه عادة يستغرق بين 15 و30 دقيقة، ويمكن تطبيقه فرديًا أو جماعيًا مع تعليمات بسيطة. وتستند معاييره إلى عينات من التلاميذ والبالغين في بريطانيا. وعلى الرغم من أن اختبار رافن يُعد من أدوات القياس الاختبارات غير المتحيزة للثقافة (Culture-Fair Tests) التي تهدف إلى تقليل تأثير الخلفية الثقافية في الأداء، فإن الدراسات تشير إلى صعوبة الوصول لاختبار خالٍ تمامًا من التأثير الثقافي. كما تُظهر نتائج البحث أن الاختبار أقل قدرة على التنبؤ بالتحصيل الدراسي والأداء المهني مقارنة بالاختبارات اللفظية التقليدية، خاصة في مجالات الانتقاء والتوجيه المهني.

ثامناً: تطبيقات القياس العقلي في التعليم والإرشاد والعيادات النفسية:

تُعدّ تطبيقات القياس العقلي من الركائز الأساسية في مجالات التعليم والإرشاد والعيادات النفسية، نظرًا لدورها في فهم القدرات العقلية وتشخيص الصعوبات واتخاذ القرارات الملائمة للتدخل. ففي المجال التعليمي تُستخدم اختبارات الذكاء لتحديد المتعلمين ذوي الاحتياجات الخاصة والمتفوقين وتوجيه البرامج التعليمية المناسبة لهم، بالإضافة إلى تشخيص صعوبات التعلم وتقديم الدعم الأكاديمي الملائم (Sattler, 2018). أما في الإرشاد النفسي فتسهم مقاييس الذكاء في فهم الفروق الفردية وتحديد القدرات المعرفية والميول المهنية مما يساعد المرشدين في توجيه الأفراد نحو الاختيارات الأكاديمية والمهنية المناسبة لهم وتعزيز مهاراتهم المعرفية (Anastasi & Urbina, 1997). وفي السياقات الإكلينيكية تُعد هذه الاختبارات أداة محورية في تشخيص الاضطرابات النمائية والمعرفية كالإعاقات الذهنية واضطرابات الذاكرة والخرف، إضافة إلى تقييم التدهور المعرفي الناتج عن الإصابات الدماغية أو الأمراض النفسية (Groth-Marnat & Wright, 2016).

وبذلك يظل القياس العقلي عنصرًا جوهريًا في دعم القرارات التربوية والإرشادية والإكلينيكية القائمة على الأدلة العلمية.

تاسعاً: القياس العقلي في البيئة العربية (التحديات الثقافية والتقنين):

في البيئة العربية واجه القياس العقلي العديد من التحديات المرتبطة بالثقافة واللغة والسياقات الاجتماعية مما استدعى جهودًا كبيرة في مجال التقنين والتعريب لضمان الصدق والثبات وملاءمة المقاييس المستوردة للواقع العربي. وتتمثل أبرز التحديات في الفروق الثقافية والمعيارية بين البيئات الغربية التي طُوِّرت فيها اختبارات الذكاء وبين البيئة العربية؛ فالعادات والقيم وأنماط التفكير السائدة قد تؤثر في طريقة استجابة الأفراد لبنود الاختبار، كما أن الفروق اللغوية والدلالية تتطلب ترجمة دقيقة وتكييفًا ثقافيًا عميقًا يتجاوز الترجمة الحرفية. إضافة إلى ذلك يبرز نقص المقاييس العربية الأصيلة وقلة الدراسات الموسعة للتقنين على عينات ممثلة لمختلف الدول والفئات العمرية، فضلاً عن محدودية التدريب المتخصص في تطبيق وتفسير الاختبارات النفسية في بعض البلدان العربية. وقد ساهمت محاولات تقنين اختبارات مثل وكسلر وستانفورد–بينيه على عينات عربية في سد جزء من هذه الفجوة، إلا أن الحاجة ما زالت قائمة لتطوير مقاييس عربية حديثة تراعي الخصوصيات الثقافية والمعرفية للمجتمع العربي، مع توظيف الأساليب الإحصائية المتقدمة كنظرية الاستجابة للمفردة الاختبارية (IRT) لتعزيز جودة القياس.

عاشراً: مفهوم الاستعداد وعلاقته بمفهوم القدرة:

الفرق بين الاستعداد والقدرة من المفاهيم الجوهرية في القياس النفسي والتربوي، ويؤثر الفهم الدقيق لهذين المصطلحين على تصميم الاختبارات، تفسير النتائج، ووضع السياسات التعليمية. فيما يلي مدخل واضح ومركّز يوضح الفروق الأساسية، أمثلة، وتبعات قياسية وتربوية.

1- الاستعداد (Aptitude):

- **تعريفه:** الاستعداد هو قابلية كامنة لدى الفرد، تمثل إمكانية أو قابلية للتعلم أو اكتساب مهارة في المستقبل إذا أُتيح له التدريب المناسب.

- **خصائصه:**

- قدرة محتملة لم تتحقق بعد بشكل كامل.

- يتنبأ بما يمكن أن يفعله الفرد لاحقًا.

- يتأثر بعوامل وراثية وبيئية، لكنه أكثر ثباتًا من المهارات المكتسبة.

- يقيسه غالبًا: اختبارات الاستعداد، اختبارات القدرات التحصيلية المستقبلية، مثل اختبارات الذكاء أو الاستعداد المهني.

- **مثال:** طفل لديه استعداد رياضي، أي لديه قابلية للتعلم والنجاح في الرياضة إذا تلقى تدريبًا .لكنه قد لا يتقن اللعب اليوم بشكل جيد.

2- جدول يوضح أهم اختبارات الاستعدادات (Aptitude Tests)

الاختبار بالعربية	الاختبار بالإنجليزية
بطارية اختبارات الاستعدادات الفارقة	Differential Aptitude Test Battery (DAT)
بطارية اختبارات الاستعدادات العامة	General Aptitude Test Battery (GATB)
اختبار الاستعداد الأكاديمي	Scholastic Aptitude Test (SAT)
بطارية اختبارات سجل الدراسات العليا	Graduate Record Examination Aptitude Test (GRE)
بطارية الاستعدادات المهنية العسكرية	Armed Services Vocational Aptitude Battery (ASVAB)
اختبار "بينيت" للفهم الميكانيكي	Bennett Mechanical Comprehension Test (BMCT)
اختبار القدرة المدرسية أوتيس-لينون	Otis–Lennon School Ability Test (OLSAT)
اختبار القبول في كلية الحقوق	Law School Admission Test (LSAT)
اختبار القبول في كلية الطب	Medical College Admission Test (MCAT)
اختبار مينيسوتا للاستعداد الكتابي	Minnesota Clerical Test (MCT)

2- القدرة (Ability):

- **تعريفها:** القدرة هي مستوى الأداء الحالي للفرد في مهارة أو مجال معين، أي ما يستطيع فعله الآن بالفعل.

- **خصائصها:**

- قدرة فعلية حالية وليست كامنة.
- تمثل المهارات التي اكتسبها الفرد من خلال التعلم أو الخبرة.
- قد تتغير بسرعة أكبر لأنها مرتبطة بالتعلم والممارسة.
- تقاس غالبًا باختبارات الأداء، التحصيل أو اختبارات المهارات.

مثال: طالب يستطيع حل مسائل الرياضيات الآن بشكل جيد، فهذا يعكس قدرته الحالية في الرياضيات.

2-1- الفرق بين الاستعداد والقدرة:

الجانب	الاستعداد	القدرة
طبيعتها	كامنة / محتملة	فعلية / حالية
الزمن	يصف ما يمكن أن يحدث مستقبلًا.	يصف ما يحدث الآن
العلاقة بالتعلم	يتوقع النجاح في التعلم.	نتيجة للتعلم السابق
الاستخدام	التوجيه والإرشاد، التنبؤ بالأداء.	وصف مستوى المهارة الحالية
أمثلة	الاستعداد اللغوي، الاستعداد الميكانيكي، الذكاء.	القدرة الحسابية، القدرة القرائية، القدرة على حل المشكلات

2-2- جدول يوضح أهم اختبارات القدرات (Ability Tests)

الاختبار بالعربية	الاختبار بالإنجليزية
مقياس وكسلر لذكاء البالغين / الأطفال	Wechsler Intelligence Scales (WAIS, WISC)
مقياس ستانفورد–بينيه للذكاء	Stanford–Binet Intelligence Scale
اختبار وودكوك–جونسون للقدرات المعرفية	Woodcock–Johnson Tests of Cognitive Abilities
اختبار كاتل للذكاء الخالي من الثقافة	Cattell Culture Fair Intelligence Test
اختبار كوفمان لتقييم الأطفال	Kaufman Assessment Battery for Children (KABC)
اختبار الذكاء غير اللفظي (توني)	Test of Nonverbal Intelligence (TONI)
مقياس وكسلر للتحصيل الفردي	Wechsler Individual Achievement Test (WIAT)
اختبار بيبودي للمفردات المصورة	Peabody Picture Vocabulary Test (PPVT)
مصفوفات رفن المتتابعة كمؤشر للقدرة العامة	Raven’s Progressive Matrices
مقياس القدرات المتباينة	Differential Ability Scales (DAS)

الحادي عشر: الإبتكاريـــة ومفهوم التفكير التباعـدي:

تعد الابتكارية من المفاهيم النفسية التي حظيت باهتمام متزايد من الباحثين في علم النفس والتربية خلال السنوات الأخيرة، إلا أنّ تعريفها ما يزال محلّ تباين بسبب تعدّد الاتجاهات النظرية التي تناولتها. فبعض العلماء ينظرون إلى الابتكارية بوصفها عملية ممتدة زمنياً تتسم بـ الأصالة والقدرة على التكيّف وتحقيق الأفكار، ما يعني ضرورة دراسة الأفراد الذين يمتلكون مستويات مرتفعة من الأداء الابتكاري. ويرى آخرون أنّ الابتكارية تمثل قدرة على اكتشاف الجديد، في حين يعتبر فريق ثالث أنها ليست قدرة قائمة بذاتها، بل مجموعة عمليات نفسية تفضي إلى إنتاج يتسم بالجدة والقيمة. كما يركّز بعض الباحثين على الإنتاج الابتكاري ذاته أكثر من تركيزهم على العملية.

وتتراوح هذه التصورات بين النظر إلى الابتكارية كعملية حلّ مشكلات ذات طابع جديد، وبين اعتبارها تحقيقًا كاملاً لطاقات الفرد المتميزة. ويؤكد ماكينون (Mackinnon) أنه ليس من الضروري البحث عن تعريف واحد أمثل للابتكارية، لأنها مفهوم واسع يجمع كل المعاني السابقة وربما يتجاوزها. ولذلك يُفضّل التعامل معها بوصفها ظاهرة متعددة الأبعاد لا كمتغير نظري أحادي يتطلب تعريفًا صارمًا، مما يعني أنها تضم عدداً من الجوانب المترابطة.

ويُعدّ هذا التوجّه مفيدًا خاصة عند دراسة الابتكارية في إطار شامل، إذ يسمح بتقسيم هذه الظاهرة المعقدة إلى جوانب أو مكوّنات يمكن تحليلها ومعالجتها بفعالية أكبر مقارنة بالتعامل مع مفهوم الابتكارية على نحو عام ومجرّد (علام، 2000، ص453).

1- قياس قدرات التفكير الابتكاري:

يعد كل من جيلفورد (Guilford) وتورانس (Torrance) من أبرز الرواد في مجال قياس قدرات التفكير الابتكاري، وقد طوّرا مجموعة من الاختبارات تهدف إلى قياس قدرات ابتكارية محددة تشمل: الطلاقة، المرونة، الأصالة، وإدراك التفاصيل. وتعد هذه القدرات المكوّن الأساسي لما يعرف بالتفكير التباعدي، وفيما يلي توضيح موجز لهذه القدرات:

أ- الطلاقـــة (Fluency): تشير الطلاقة إلى قدرة الفرد على إنتاج عدد كبير من الأفكار أو الكلمات أو الأساليب للتعبير عن الأشياء. وتتضمن عدة صور، مثل: طلاقة الأفكار (Ideational Fluency)، طلاقة التداعي (Associational Fluency)، والطلاقة التعبيرية (Expressional Fluency). يُطلب من الفرد مثلاً التفكير في أكبر عدد ممكن من الاستخدامات غير المألوفة لشيء معين أو تقديم مرادفات

لكلمة محددة أو تكوين جمل بناءً على شروط خاصة. وتُمنح نقطة عن كل فكرة جديدة ويُعدّ عدد الأفكار المنتَجة مؤشرًا لدرجة الطلاقة لديه.

ب- المرونـــة (Flexibility): المرونة هي قدرة الفرد على تغيير مساره الفكري والانتقال من نمط تفكير إلى آخر بما يتوافق مع متطلبات المهمة. وتعكس هذه القدرة عدم الجمود الفكري وتنوّع أساليب معالجة المعلومات. وتظهر المرونة في صورتين رئيسيتين:

- المرونة التلقائية (Spontaneous Flexibility): عندما تتعلق المهمة بمعانٍ لغوية.

- المرونة التواؤمية (Adaptive Flexibility): عندما تتعلق المهمة بالأشكال البصرية أو الهندسية. تحتسب نقطة لكل انتقال جديد في التفكير، ويعكس مجموع النقاط مستوى المرونة لدى الفرد.

ج- الأصالـــة (Originality): الأصالة تعني قدرة الفرد على إنتاج أفكار نادرة وغير مألوفة من منظور إحصائي. ويُستخدم معيار شائع يقوم على منح نقاط للأفكار التي لا ينتجها سوى عدد قليل من الأفراد، مثلاً (5 من بين 100). وتقييم الأفكار عادة باستخدام مقياس تدريجي من خمس نقاط، حيث تُعطى نقطة واحدة للأفكار الشائعة أو المكررة، بينما تُمنح خمس نقاط للأفكار عالية الجدة والتميز. وتُعطى درجة صفرية للأفكار التي يقدّمها أكثر من خمسة أفراد من بين مائة. ويعكس مجموع النقاط مستوى الأصالة.

د- إدراك التفاصل (Elaboration): يقصد بإدراك التفاصيل قدرة الفرد على توسيع الفكرة الأساسية بإضافة عناصر جديدة تزيد من عمقها أو وضوحها. قد يطلب من الفرد مثلاً كالرسم باستخدام شكل بسيط كالمربع ثم تطوير هذا الشكل ليصبح صورة أكثر تعقيداً كتحويل المربع إلى باب وإضافة زخارف أو فتحة للرؤية أو أي تفاصيل إضافية تمنح الفكرة ثراءً بصرياً. ويعد عدد التفاصيل المضافة مؤشراً على درجة هذه القدرة (علام، 2000، ص455).

2- الفرق بين التفكير التباعدي والتقاربي في إطار قدرات جيلفورد وتورانس:

يمثل التفكير التباعدي (Divergent Thinking) البنية الأساسية التي اعتمد عليها جيلفورد وتورانس في قياس القدرات الابتكارية، إذ يركز هذا النوع من التفكير على إنتاج أكبر عدد ممكن من الأفكار المتنوعة والمبتكرة استجابةً لمثير واحد، دون الالتزام بإجابة واحدة صحيحة. ويظهر التفكير التباعدي في قدرات مثل الطلاقة والمرونة والأصالة وإدراك التفاصيل؛ فهو يسعى إلى توسيع نطاق الاحتمالات والابتعاد عن الأنماط المألوفة واستكشاف حلول جديدة وغير تقليدية للمشكلة. وبالتالي فإن الجودة الأساسية للتفكير التباعدي هي تنوع الأفكار وجدّتها.

أما التفكير التقاربي (Convergent Thinking) فيمثل النقيض الوظيفي للتفكير التباعدي، ويُقصد به البحث المنهجي عن الإجابة الصحيحة أو الأكثر دقة من بين مجموعة من البدائل بالاعتماد على المنطق والتحليل والمعرفة المكتسبة. ويرتبط التفكير التقاربي باختبارات الذكاء وحل المشكلات ذات الإجابة الواحدة، حيث يقوم الفرد بتضييق الخيارات وصولاً إلى حل محدد. وعلى عكس التفكير التباعدي الذي يركز على الكمّ والتنوع يتميز التفكير التقاربي بالتركيز على التحقق والدقة والإجابة المثلى.

وفي سياق نظرية جيلفورد وتطبيقات تورانس، ينظر إلى التفكير التباعدي باعتباره المكون الأساسي للابتكارية؛ بينما يُعد التفكير التقاربي مهماً للعمليات التحليلية لكنه لا يُمثل الابتكارية بحد ذاتها. ومن خلال الفصل بين هذين النمطين يصبح من الممكن قياس القدرات الابتكارية بشكل أدق وتمييز الإنتاج الإبداعي عن الأداء المعرفي التقليدي.

الدكتور زيـاد رشيـد

المحـــاضرة الخامسة

مجـــالات القيـــاس النفســـي
(قياس الشخصية)

الأهداف:

تهدف هذه المحاضرة إلى تمكين الطلبة من فهم مفهوم الشخصية في علم النفس من خلال التعرف على أبرز نظرياتها وتصنيفات مقاييسها ومناهج قياسها، إضافة إلى تتبع التطور التاريخي لقياس الميول واستيعاب مفهومها وأساليب قياسها، مع الوقوف على أهم المشكلات المرتبطة بالاستبيانات المقننة للميول. كما تسعى المحاضرة إلى توضيح مفهوم الاتجاهات من خلال استعراض تطور قياسها وتصنيف مقاييسها وأساليب بناء موازينها، إلى جانب تنمية وعي الطلبة بأبرز التحديات والصعوبات التي تواجه قياس الجوانب غير المعرفية في المجال النفسي والتربوي.

محتوى المحاضرة:

أولاً: الشخصيـــة (Personality)
1- مفهوم الشخصية
2- نظريات الشخصية
3- تصنيف مقاييس الشخصية
4- مناهج قياس الشخصيـــة
ثانيـــاً: قياس الميول من منظور تاريخي
1- تعريف الميول (Interest)
2- قياس الميول (Interest Measurement)
3- بعض المشكلات المتعلقة بالإستبيانات المقننة للميول
ثالثاً: الاتجـاهـــات (Attitudes)
1- قياس الاتجاهات من منظور تاريخي
2- تصنيف مقاييس الاتجاهات
3- أساليب بنـاء موازين الاتجـاهـــات
رابعاً: بعض المشكلات المتعلقة بقياس الجوانب غير المعرفية

مدة المحاضرة: أربع ساعات ونصف

تمهيد:

عند الحديث عن القياس النفسي كثيراً ما يركز الاهتمام على الجوانب المعرفية كقدرات الفرد واستعداداته العقلية غير أن السلوك الإنساني لا يختزل في هذه الجوانب وحدها. فهناك أبعاد أخرى لا تقل أهمية والتي تتعلق بالجوانب الوجدانية أو غير المعرفية (Non-Cognitive)، كالانفعالات والدوافع والاتجاهات والميول والسمات الشخصية. هذه الجوانب تُسهم إلى جانب القدرات العقلية (Cognitive) في تشكيل السلوك الإنساني المتكامل، حيث يتفاعل البعدان المعرفي وغير المعرفي في التأثير على أداء الفرد واستجاباته في المواقف المختلفة. فقياس القدرات العقلية وحدها لا يمكن أن تفسر الأداء الإنساني تفسيراً كاملاً، كما أن الاعتماد على قياس الجوانب الوجدانية فقط لا يقدم صورة دقيقة عن سلوك الفرد. إنّ كلاً من هذين الجانبين يمثل وجهًا من وجوه الشخصية الإنسانية المتعددة التي تتفاعل فيما بينها لتنتج سلوكاً متكاملاً يعكس خصائص الفرد واتجاهاته وقدراته.

وتركز أدوات قياس الجوانب غير المعرفية على تقدير الأداء النمطي للفرد (Typical Performance)، أي ما يعبّر عن أسلوبه المعتاد في التفكير والتفاعل والانفعال. ورغم أن هذه الأدوات تعد أقل دقة في التنبؤ بالنجاح الأكاديمي أو المهني مقارنة بأدوات قياس القدرات العقلية، إلا أنها تعد أكثر فائدة في مجالات الإرشاد والتوجيه النفسي، لأنها تقدم معلومات نوعية غنية تساعد على فهم الشخصية بصورة أكثر شمولية. وتشمل الجوانب غير المعرفية طيفاً واسعاً من الخصائص النفسية كالعواطف (الانفعالات) والمشاعر، والاتجاهات، والميول، والمعتقدات، والقيم، والدوافع، والسمات المزاجية، والحس الاجتماعي، وغيرها من مكونات الشخصية.

وفي هذه المحاضرة سنركز على ثلاثة من أهم هذه الجوانب:الشخصية، الميول، والاتجاهات. وعلى الرغم من أن مفهوم الشخصية في معناه الواسع قد يضم هذه الجوانب جميعها، فإننا سنميز بينها لاعتبارات علمية ومنهجية مع تناول الأدوات الخاصة بكل منها واستخداماتها وأبرز المشكلات المرتبطة بها.

أولاً: الشخصيــة (Personality):

شهد مجال دراسة الشخصية تطورًا كبيرًا مع تطور الأساليب السيكومترية الحديثة، حيث استفاد علماء النفس من تقنيات التحليل العاملي وأساليب الاختيار القسري في بناء أدوات دقيقة لقياس الشخصية. ويُعَدّ كل من كاتل وثيورستون وأيزنك وجيلفورد من أبرز الرواد الذين أسهموا في إرساء الأساس العلمي لقياس السمات الشخصية اعتمادًا على التحليل العاملي. وقد مثّلت جهود جيلفورد في

أوائل الأربعينيات، وجهود كاتل في نهاية العقد نفسه، نقطة تحول مهمة في هذا الميدان، إذ قام كاتل على وجه الخصوص ببناء استبيان العوامل الستة عشر للشخصية (16PF) اعتمادًا على تطبيقات متقدمة للتحليل العاملي. ومع تطور البحث في الشخصية، انتقلت الاهتمامات من التركيز على أدوات القياس ذاتها إلى بناء نظريات أكثر عمقًا حول بنية الشخصية وتكوينها، مما أسهم في صياغة نماذج فرضية تستند إلى أسس نظرية واضحة. لهذا نجد أن استخدام مقاييس الشخصية اليوم يتم بقدر كبير من الحذر من قبل المختصين، نظرًا لحساسية هذا المجال وما يتطلبه من دقة في التفسير والتطبيق.

1- مفهوم الشخصيــة:

يعد قياس الشخصية أحد المجالات الرئيسة في القياس النفسي والتربوي، غير أن فهم مقاييس الشخصية يتطلب أولاً تحديد المقصود بهذا المفهوم. ويواجه الباحثون صعوبة في وضع تعريف موحّد للشخصية بسبب تعدد المنظورات النفسية، إذ يشير ألبورت في دراسته إلى وجود أكثر من خمسين تعريفًا لهذا المفهوم، ما يعكس اتساعه وتعقيده.

تتباين تعريفات الشخصية بين الباحثين؛ فجيلفورد يرى أنها تشمل جميع الخصائص المستقرة نسبيًا لدى الفرد، بما فيها المظهر الجسمي والقدرات والاتجاهات والقيم. بينما يركز باحثون آخرون على الجوانب غير المعرفية فقط مثل الانفعالات والدوافع والخلق، معتبرين أن هذه الجوانب تمثل جوهر الشخصية. وقد دعت الاتجاهات الحديثة إلى ضرورة دمج الجانبين المعرفي والانفعالي معًا لفهم الشخصية بصورة أعمق، لأن السلوك ناتج عن تفاعل العقل مع الموقف. وبسبب اختلاف الاتجاهات، لا يمكن الجزم بوجود تعريف واحد صحيح للشخصية، بل يعتمد التعريف على الهدف من استخدامه. فمثلًا عرف ألبورت الشخصية بأنها تنظيم دينامي داخل الفرد يوجه سلوكه، بينما رأى ساندبرج أنها عمليات يستخدمها الفرد لتكوين انطباعات واتخاذ قرارات حول الآخرين. أما كراون فيعتبر الشخصية نظامًا من الإمكانات السلوكية التي تظهر في تكامل السلوك عبر الزمن والمواقف المختلفة، مؤكدًا ضرورة التمييز بين تحليل الشخصية ككل وبين تحليل السلوك من خلال متغيرات محددة. وبالرغم من التباين، يتفق معظم الباحثين على مجموعة خصائص عامة للشخصية: فهي تجمع عدة سمات وقدرات، وتتسم بالتنظيم والتكامل، وتتميز بالتفرد بين الأفراد، وتؤثر في تفاعل الفرد مع بيئته، كما أنها مستقرة نسبيًا عبر الزمن. ويمكن اعتماد هذه العناصر إطارًا عامًا لتوجيه دراسة الشخصية.

وقد حاولت نظريات الشخصية تقديم نماذج منظمة لهذا المفهوم؛ فبعضها ركز على بنية الشخصية من خلال تحديد الأبعاد أو السمات الأساسية، مثل الانطواء–الانبساط وغيرها. ويعبر عن هذه السمات من خلال الصفحة النفسيـــة (Psychogram) أو البروفيل النفسي (Profile) الذي يوضح موقع الفرد على عدة أبعاد. أما الاتجاه الآخر المعتمد على نظريات التعلم والتحليل النفسي فيركز على عمليات نمو الشخصية وديناميـاتها، مع الاهتمام بالدافعية والتعلم والصراعات وتكوين العادات. كما ظهرت مداخل أخرى تعتمد على نماذج معالجة المعلومات والجوانب المعرفية (علام، 2000، ص578). وستعرض هذه المداخل بمزيد من التفصيل عند مناقشة أساليب وطرق قياس الشخصية في الفقرات اللاحقة.

2- نظريات الشخصية:

تنوعت وجهات النظر في دراسة سيكولوجيا الشخصية نتيجة لتعدد المدارس والاتجاهات النفسية التي ينتمي إليها الباحثون والمفكرون. وسنستعرض مفهوم الشخصية من منظورات علماء النفس من مختلف هذه المدارس، حيث سنتناول موضوع الشخصية من خلال نظريات السمات، والنظريات التحليلية النفسية، والنظريات الإنسانية، بالإضافة إلى النظريات المعرفية الاجتماعية.

2-1- نظريات السمـات:

تعد نظرية السمات من النظريات البسيطة والمفهومة لدى العامة بسبب اتصالها المباشر بحياة الناس اليومية من خلال الأوصاف التي يطلقونها على بعضهم البعض من البخل، الشجاعة، الثرثرة، الذكاء ...إلخ، إذ تقول أن لكل فرد مميزات (سمات) شخصية ثابتة فيه ملاحظة تميزه عن الآخرين لاختلافه عنهم بهذه السمات. ويجمع هؤلاء على أن السمة هي الوحدة الرئيسية للشخصية. وهم يعرفونها بأنها استعداد مسبق أو ميل محدد للاستجابة. ومن هذا المنطلق راحوا يبحثون عن شبكة السمات التي تتكون منها الشخصية. فقد وضع ألبورت وقد أودبرت H. ODBERT عام 1936 قائمة تتضمن 4541 كلمة تستعمل في اللغة الإنجليزية لوصف الشخصية. ورصد بيرت السمات الجسمية والنفسية والاجتماعية. ووجد أن السمات الجسمية تصف حالة الجسم ونواحي القدرة أو العجز فيه وأن السمات النفسية تدل على الجوانب المعرفية والمزاجية الموروثة والمكتسبة.

فالسمات المعرفية الموروثة تشمل الذكاء والاستعدادات الخاصة، بينما تمثل السمات المعرفية المكتسبة المهارات الخاصة التي يكتسبها الفرد. أما السمات المزاجية الموروثة فتعبّر عن الانفعالات العامة والنماذج المزاجية الخاصة، في حين تجسد السمات المزاجية المكتسبة الاتجاهات والميول

والعادات التي تتطور مع الزمن. أما السمات الاجتماعية فتركز على العوامل البيئية كالوضع الاقتصادي للأسرة والظروف المنزلية والعلاقات الأسرية والاجتماعية ونوع العمل وأساليب قضاء أوقات الفراغ.

2-2- نظرية العوامل الستة عشر للشخصية (Cattell, 1943):

قدم ريموند كاتل (Cattell, 1943) نظرية في بنية الشخصية سعى من خلالها إلى تبسيط مكونات الشخصية الإنسانية إلى ستة عشر عاملاً أساسياً ثنائية القطب. وشملت هذه العوامل: الانطلاق، الذكاء، قوة الأنا، السيطرة، التفاؤل، قوة الأنا الأعلى، حب المغامرة، الطراوة، التوحد، الاستقلال، الدهاء، الميل إلى الشعور بالذنب، التحرر، الاكتفاء الذاتي، ضبط العواطف، وضبط الدوافع.

وقد عمل كاتل على تقليص عدد السمات إلى الحد الأدنى الممكن من أجل تسهيل تطبيق التحليل العاملي في دراستها، فبذل جهودًا كبيرة أثمرت عن تطوير استبيان العوامل الستة عشر للشخصية (16PF)، الذي ضمّ عددًا كبيرًا من البنود المصممة لقياس هذه السمات التي اعتبرها كافية لوصف الشخصية بشكل شامل (غباري وأبو شعيرة، 2009، ص119).

2-3- نظرية العوامل الخمس الكبرى (Big Five):

اقترح فيسكي (Fiske, 1949) أن العوامل الأساسية التي تقيس خصائص الشخصية هي خمسة عوامل فقط، وليس ستة عشر عاملاً كما توصل إليها كاتل. ويستند هذا الاقتراح إلى الأساس النظري لنموذج العوامل الخمسة الكبرى (Big Five)، الذي تطور من خلال استخدام أسلوب إحصائي خاص يُعرف بـ التحليل العاملي (Factor Analysis)، والذي أصبح أداة رئيسية في دراسات ونظريات الشخصية. وعلى الرغم من اختلاف النظريات الفردية، إلا أنها تتفق جميعًا على الانتماء إلى عائلة واحدة من النظريات (عبد الخالق والأنصاري، 1996، ص6-19).

وباستخدام التحليل العاملي أصبح من الممكن النظر إلى هذه العوامل الخمسة كمجموعة متكاملة من النظريات. فقد طور كل من تويس وكريستال (1961)، نورمان (1963)، أيزنك (1967)، وكوستا وماكاري (1992) أساسًا قويًا لنموذج العوامل الخمسة الكبرى، وأكدت العديد من الدراسات ثبات هذه العوامل عبر مجموعات سكانية متنوعة.

ومن خلال التقدير الذاتي وتقييم الخبراء والتحليل العاملي، تمكن جولدبرج من عزل حوالي ثلاثة عشر عاملاً، كان ترتيب العوامل الخمسة الأولى متطابقًا مع تلك التي حددها نورمان. وبالتالي تجمع هذه الأبعاد الخمسة الأساسية بين الاتجاهات النفسية التي ينتمي إليها معظم المنظرين.

أما تعاريف العوامل الخمسة الكبرى، فهي محاولة لوصف القاسم المشترك بين مجموعة من الخصال أو العوامل الثانوية، وقد طور كوستا وماكاري قائمة بالخصال الأكثر شيوعًا لهذه العوامل. ويرى كوستا وماكاري أن نموذج العوامل الخمسة يعد تمثيلاً صحيحًا لبنية الخصال الشخصية، وأن صحة هذه النظرية تمثل نقطة تحول مهمة في علم نفس الشخصية. والعوامل الخمسة الكبرى، هي كما يلي:

أ- الانبساطية (Extraversion): تشمل الدفء، الاجتماعية، البحث عن المتعة، العواطف الإيجابية.

ب- العصابية (Neuroticism): تشمل القلق، العدوان، الاكتئاب، الاندفاعية، والحساسية للنقد.

ج- المقبولية (Agreeableness): تشمل الثقة، الخضوع، الحشمة، الإيثار، الاستقامة.

د- الضميرية أو الاهتمامادية (Conscientiousness): تشمل التنافس، النظام، الكفاح من أجل التفوق، القصدية، انضباط الذات، والشعور بالواجب.

هـ- الانفتاح على الخبرة (Openness to Experience): تشمل المغامرة، الرياضة، الفضول الفكري، المشاعر، الأفعال، الأفكار، القيم، والاستقلالية (غباري وأبو شعيرة، 2009، ص131).

2-4- نظريات التحليل النفسي (Psychoanalytic Theory):

يعتبر سيغموند فرويد (1856–1939) Sigmund Freud واحدًا من عمالقة علم نفس الشخصية، وعلى الرغم من مرور ما يقارب القرن على طرح نظريته، إلا أنها لا تزال مثيرة للجدل وتحظى باهتمام واسع. يعود الفضل لفرويد في كونه من الرواد الذين أسهموا بشكل كبير في تطوير فهمنا للسلوك البشري، وذلك لعدة أسباب مهمة:

- أدرك فرويد الحاجة إلى مراقبة السلوك بدقة متناهية، رغم أن معظم العلماء اختلفوا مع العديد من جوانب نظريته واعتبروها مثيرة للجدل أو حتى خاطئة.

- حاول فرويد تبسيط تفسير السلوك الإنساني عبر عدد محدود من العمليات النفسية الأساسية، ما جعل نظريته نموذجًا مركزًا لفهم السلوك.

- طور فرويد نموذجًا عقليًا وظيفيًا للعلاج النفسي يُعرف بـ التحليل النفسي (Psychoanalysis)، والذي كان ثوريًا في طريقة التعامل مع الاضطرابات النفسية.

- سعَى فرويد إلى تفسير جوانب متعددة من حياة الإنسان شملت الدين والفن والأدب ونمو الطفولة والأحلام والانحرافات السلوكية وزلات اللسان والدوافع والقلق والسلوك الجنسي، بالإضافة إلى الاكتئاب والأمراض العصابية المختلفة والعدوان.

- أثرت نظرية فرويد بشكل كبير لا سيما في أوروبا وتجاوزت حدود علم النفس لتؤثر في مجالات معرفية أخرى مثل النقد الأدبي والثقافة بشكل عام.

وأخيرًا، كانت نظرية فرويد الأولى التي تناولت مفهوم اللاشعور (Unconscious)، الذي يشبه سلة مهملات تضم الخبرات والمشاعر التي يحاول الفرد نسيانها أو حجبها، مما أضاف بعدًا جديدًا لفهم النفس البشرية.

بهذا الشكل ظل تأثير فرويد حاضرًا في الدراسات النفسية رغم كل الجدل، حيث لا يمكن إنكار دوره الكبير في تأسيس علم نفس الشخصية (غباري وأبو شعيرة، 2009، ص136).

3- تصنيف مقاييس الشخصية:

يتضح من المراجعة السابقة أن علماء النفس لم يتفقوا على تعريف موحد لمفهوم الشخصية (Personality)، مما أدى إلى تنوع وتعدد النظريات التي تسلط الضوء على هذا المفهوم من وجهات نظر مختلفة. ونتيجة لذلك، تعددت وتنوعت طرق وأساليب قياس الشخصية (Personality Assessment) والعديد من المتغيرات المرتبطة بها. يُعتبر مجال قياس الشخصية من أكثر مجالات القياس النفسي والتربوي تنوعًا في الأساليب والطرق، ويعود ذلك كما سبق أن أشرنا إلى شمول مفهوم الشخصية لمجموعة واسعة من المتغيرات والمفاهيم غير المتجانسة.

على الرغم من اختلاف طرق تصنيف مقاييس الشخصية، إلا أن هناك عناصر مشتركة تجمع بينها، حيث يعد قياس أو تقييم الشخصية أسلوبًا لجمع المعلومات عن فرد ما، ويشمل ذلك الملاحظة الموضوعية والمنظمة للسلوك تحت شروط محددة مرتبطة بمثيرات معينة (Stimuli). وللحصول على بيانات دقيقة عن الفرد، يجب فهم تأثير مكونات الموقف المختلفة المتعلقة بالمثير على سلوك الفرد. ويتحدد هذا الموقف الذي يواجهه الفرد ويستجيب له من خلال خصائصه الفيزيائية، والتعليمات المقدمة له، والمتطلبات المحددة للاستجابة.

ونظرًا لاختلاف المتغيرات المرتبطة بعملية جمع البيانات الشخصية، فإن هناك تباينًا واضحًا في طرق وأساليب جمع هذه البيانات. فخصائص كل من: الموقف (Situation)، والمثير (Stimulus)، والتعليمات (Instructions)، والاستجابة المطلوبة (Required Response)، وكذلك طرق تقدير الدرجات (Scoring) وتحليلها وتفسيرها، تعتبر من خصائص عملية قياس أو تقييم الشخصية، إذ تُسند إليها أهمية في جمع وتسجيل وتحليل وتفسير البيانات. لذلك، من الضروري اتباع أسلوب منظم لجمع هذه البيانات ضمن إطار نظرية محددة أو فرضيات واضحة، لأن اختلاف نظريات الشخصية يؤدي

إلى اختلاف في طرق ملاحظة سلوك الأفراد، مما يشير إلى وجود علاقات تفاعلية بين نظريات الشخصية وطرق وأساليب القياس أو التقييم. ويرى Arndt (1974) أن نظرية الشخصية التي لا تساهم في تحديد وتفسير وسائل جمع البيانات تُعد نظرية غير عملية، كما أن البيانات التي تُجمع دون ارتباط بنظرية محددة تكون ذات قيمة قليلة أو عديمة المعنى. لذا ينبغي أن يستند تصنيف مقاييس الشخصية إلى إطار نظري واضح يأخذ بعين الاعتبار الفروقات بين النظريات المختلفة وتوجهات البحث النفسي في هذا المجال.

تتباين طرق وأساليب قياس أو تقييم الشخصية من حيث مكونات عملية جمع البيانات، حيث يمكن أن يكون الموقف الاختباري (Testing Situation) مشابهًا لمواقف الحياة الواقعية، أو محاكاة لها، أو موقفًا مختبريًا محددًا للمثيرات. كما يمكن أن يتضمن الموقف مثيرًا واحدًا أو عدة مثيرات، تكون محددة البنية (Structured) أو غير محددة البنية (Nonstructured) أو (ambiguous). وبعض هذه المواقف يتطلب تسجيل استجابة واحدة، وبعضها الآخر يتطلب تسجيل عدة استجابات، والتي قد تكون تقارير لفظية ذاتية (Self-Report) أو استجابات فسيولوجية، أو أداءً معينًا. أما تقدير الدرجات فقد يكون موضوعيًا باستخدام الحاسوب أو يدويًا، أو ذاتيًا اعتمادًا على أحكام الأخصائيين النفسيين. في الحالة الأولى يتم تفسير الدرجات من خلال الصفحات النفسية (Profiles) التي تُستخلص باستخدام برامج الحاسوب، وفي الحالة الثانية يتم تفسير البيانات بشكل إكلينيكي (Clinically) من قبل المختصين.

يمكن تصنيف مقاييس الشخصية بناءً على بعدين أساسيين، كل منهما ثنائي القطب كما في دراسات (Campbell, 1957: Mehrans & Lehman, 1969).

أ- بنية المثيرات (Stimulus Structure): محددة Structured أو غير محددة Nonstructured.

ب- هدف الاختبار (Test Purpose): إما غامضاً Disguised أو واضحاً Nondisguised (علام، 2000، ص583).

4- مناهج قياس الشخصيـــة:

4-1- الطرق الموضوعية:

أ- تعريف الاختبار الموضوعي للشخصية: يعرِف كاتل (Cattell, 1950) الاختبار الموضوعي للشخصية بأنه موقف يستخدم للتنبؤ بالسلوك في مجال معين، بحيث يكون ذا دلالة على عدد كبير من المواقف السلوكية الأخرى. وهو اختبار يشير إلى خصائص تتجاوز ما يبدو أنه يقيسه ظاهريًا،

كما يتميز بكونه أداة جاهزة البناء، قابلة للنقل والتطبيق في أي مكان، ويمكن استحضارها بدقة، وتُصحح تصحيحًا موضوعيًا، منخفضة التكلفة، قابلة للتقنين، ومختصرة مقارنة بالموقف السلوكي الحقيقي الذي تسعى إلى التنبؤ به. وفي تعريف لاحق، يوضح كاتل (Cattell, 1957) أن الاختبار الموضوعي هو اختبار يُقاس فيه سلوك المفحوص بهدف استنتاج خصائص شخصيته، دون أن يكون على وعي بالاتجاهات التي قد يؤثر بها سلوكه في عملية التفسير أو التقويم.

ولا يكون المفحوص في الاختبارات الموضوعية على علم بالجوانب التي سيُفسَّر أو يُقيَّم أداؤه على أساسها، كما لا يعتمد هذا النوع من الاختبارات على التقدير الذاتي للفرد أو على ملاحظات مباشرة من قبل الآخرين بشأن شخصيته (Schuerger, 1986). ويُعد الاختبار الموضوعي موقفًا تنبيهيًا يمكن إعادة تقديمه بدقة، يتضمن طريقة محددة مسبقًا لتقدير الاستجابات وتصحيحها، دون أن يُفصح عنها للمفحوص. وجميع الاختبارات الموضوعية تُعد مقاييس تجريبية، غير أن ليس كل مقياس تجريبي يصنف بوصفه اختبارًا موضوعياً (Cattell, 1950).

ب- خصائص الاختبارات الموضوعية للشخصية: يتسم الهدف في الاختبارات الموضوعية بكونه مستترًا ومقنَّعاً (Disguised, subtle)، إذ صُممت للحد من تشويه البيانات الشخصية، سواء كان ذلك بصورة متعمدة أو لا شعورية، ولتقليل محاولات التزييف من جانب المفحوص. وقد تتضمن هذه الاختبارات مواد لفظية، إلا أنها تركز أساسًا على أداء الفرد وسلوكه الفعلي، فهي تقيس ما يقوم به المفحوص فعلاً أكثر مما تقيس ما يصرح به عن سلوكه (Cattell & Scheier, 1961).

وتتجلى الموضوعية في هذا السياق في استجابة المفحوص من خلال أنماط سلوكية غير لفظية، مع إمكانية إخضاع الدرجات الناتجة لمختلف المعالجات الإحصائية، دون الحاجة إلى أحكام إنسانية أو تقديرات ذاتية في عملية التصحيح. ومن أبرز الخصائص المشتركة للاختبارات الموضوعية ما يأتي:

- يكون المفحوص فيها موجهًا نحو أداء مهمة (Task-oriented) أكثر من كونه موجهًا نحو التقرير اللفظي عن ذاته (Report-oriented)، كما هو الحال في الاستخبارات، حيث يُطلب منه تنفيذ عمل محدد بدلًا من وصف سلوكه المعتاد.

- يتسم الهدف من الاختبار بالاستتار، ولا يدرك المفحوص أي جوانب من أدائه سيتم على أساسها احتساب الدرجة.

- تقدم للمفحوص مهام ذات بناء محدد أو مفصل (Structured)، ويُعد هذا الجانب الفارق الجوهري بينها وبين المهام المستخدمة في الأساليب الإسقاطية.

- يدرك المفحوص الأداء المطلوب على أنه يتضمن حلًا صحيحًا، ويُنظر إلى كثير من هذه الاختبارات بوصفها مقاييس استعدادات، مما يدفع المفحوص إلى السعي لتقديم الإجابة الصحيحة، وهو ما يميزها عن الاختبارات الإسقاطية (Anastasi, 1988).

وتقيس الاختبارات الموضوعية مفاهيم نظرية، وسمات، وقدرات، ومتغيرات افتراضية، إضافة إلى متغيرات أخرى لا يمكن ملاحظتها بصورة مباشرة. ويمكن تطبيقها بصورة فردية أو جماعية، غير أن تقنين الموقف الكلي للاختبار يُعد أمرًا ضروريًا بغض النظر عن أسلوب التطبيق (Hundelby, 1973). ويميل كثير من المفحوصين، ولا سيما المتعلمين منهم، إلى الاعتقاد بأن الحصول على درجة مرتفعة في الاختبارات النفسية يُعد مؤشرًا مرغوبًا للأداء الجيد، إلا أن هذا الافتراض لا يصح دائمًا في الاختبارات الموضوعية للشخصية؛ إذ قد يُفسَّر الأداء المرتفع في بعض الاختبارات، مثل اختبار إغلاق الجشطلت، على أنه مؤشر على اضطرابات نفسية كالدهان.

وتتصف الاختبارات الموضوعية للشخصية بعدد من الخصائص المحددة، من أهمها ارتفاع درجة البناء والتفصيل، واستدعاؤها لاستجابات غير لفظية، قد تكون لاإرادية (Autonomic) مرتبطة بالجهاز العصبي المستقل، أو حركية (Motor)، فضلًا عن اعتمادها على التصحيح الموضوعي وعدم حاجتها إلى تفسير إكلينيكي مباشر (Eysenck, 1952) (عبد الخالق، ، ص487).

4-2- الطرق الإسقاطيـــة:

4-2-1- تعريف الإسقاط: الطرق الإسقاطية هي وسائل غير مباشرة لقياس الشخصية في جوانبها السوية وغير السوية. وتعتمد هذه الطرق - على تعددها - على مفهوم الإسقاط (Projection). والإسقاط - عند "فرويد" - عملية دفاعية لا شعورية يعزو بها الفرد دوافعه وإحساساته ومشاعره إلى الآخرين أو إلى العالم الخارجي، والهدف منها الدفاع ضد القلق والدوافع اللاشعورية، ويترتب عليها خفض التوتر. أما الإسقاط كما يستخدم في الطرق الإسقاطية، فيشير إلى منبه غامض غير محدد يُقدم إلى الفرد، ويُطلب منه تأويله وإعطاء معنى له. وتعكس استجابات المبحوث دوافعه وحاجاته الخاصة ورغباته ونزعاته، وإدراكاته وتفسيراته الذاتية. ويمكن أن تكون هذه المنبهات غير المتشكلة ذات أنواع شتى من مثل: بقع حبر، وكلمات، وجمل ناقصة، وصور، ومنظر السحب، وخط اليد، وأشكال هندسية... إلخ. ويُفترض أن يكشف المبحوث في استجابته لهذه المنبهات وأمثالها عن بناء شخصيته

الخاصة وقيمه ومثله العليا ومشاعره الداخلية ودوافعه ورغباته، واستعداده للمرض العقلي والاضطراب النفسي... وغير ذلك.

4-2-2- تصنيف الطرق الإسقاطية:

اقترح "لندزي" تصنيف الطرق الإسقاطية - تبعاً لنمط الاستجابة المطلوب من المبحوث - إلى خمسة أنواع كما يلي:

أ- طرق التداعي (Association): المنبه فيها كلمة أو جملة أو بقعة حبر، يستجيب لها المبحوث بكلمة أو عبارة أو مدرك، ومن أمثلتها اختبار تداعي الكلمات، واختبار رورشاخ.

ب- طرق التكوين (Construction): تنتج الاستجابة في هذا النوع من نشاط معرفي بنائي إنشائي مركب، كأن يكوّن المبحوث قصة اعتماداً على صورة (اختبار تفهم الموضوع)، أو يكوّن قصة مصورة، ويعتمد تحليل الاستجابات هنا على تحليل المضمون.

ج- طرق التكملة (Completion): يُعطى المبحوث منبهاً ناقصاً غير مكتمل (جملة، قصة)، ويطلب منه تكملته، كاختبار "ساكس" لتكملة الجمل.

د - طرق الاختيار أو الترتيب (Choice or Ordering): يقدم للمبحوث عدد من المنبهات كالصور أو الجمل، ويطلب منه إعادة ترتيبها، أو يحدد تفضيلاته لها، أو اختيار أحد البدائل منها، أو تنظيمها وفقاً لنظام محدد، وأمثلتها اختبار تنظيم الصور، واختبار "سوندي".

هـ- الطرق التعبيريـــة (Expressive): كاختبارات الرسم بالخطوط أو بالألوان، وطرق اللعب، والتمثيلية النفسية (Psychodrama)، ويمكن أن تُستخدم هذه الطائفة من الاختبارات في كل من التشخيص والعلاج، فهي تكشف عن متاعب الشخص، ويمكن أن تجعل الشخص أيضاً يتخفف منها عن طريق التنفيس.

ثانيـاً: قياس الميول من منظور تاريخي:

لعل أول محاولة للبحث في مجال الميول كانت الدراسة التي أجراها ثورنديك (Thorndike E.) ونشرت في مجلة (Popular Science Monthly) عام 1912. تناولت هذه الدراسة موضوع استمرار الميول وعلاقتها بالقدرات. استندت الدراسة إلى عينة تضم 100 طالب جامعي حيث طُلب من كل منهم ترتيب ميوله كما يتذكرها منذ المرحلة الابتدائية والثانوية وحتى التحاقه بالجامعة وكذلك ترتيب قدراته خلال هذه المراحل. تم حساب ارتباطات بين هذه التراتيب واستنتجت الدراسة أهمية الميول المبكرة وعلاقتها الواضحة بالقدرات مما يدل على أن كلاهما يعكس الآخر مع الإشارة إلى أن هذه

العلاقة معقدة. وكانت المحاولة المبكرة الثانية استبيانًا أعده كيلي (Kelley) عام 1914 تضمن أسئلة تهدف إلى استيضاح تفضيلات الأفراد الترفيهية لكن هذا الاستبيان لم يحظَ باهتمام واسع من الباحثين.

أما النشأة العملية لاستبيانات الميول كما نعرفها اليوم فبدأت بعد الحرب العالمية الأولى بفترة قصيرة في معهد (Carnegie Institute of Technology) بولاية بيتسبرغ الأمريكية. فقد عقد المعهد حلقة بحث حول الميول قام خلالها الأساتذة وطلاب الدراسات العليا ببناء ما يقرب من ألف فقرة تهدف إلى التمييز بين ميول العاملين في مختلف المهن. كان لهذه الحلقة دور كبير في تشجيع الباحثين على بناء استبيانات مقننة للميول على أسس علمية منظمة حيث تم إعداد العديد من هذه الاستبيانات. وقاد علماء النفس الصناعي، وعلى رأسهم بنجهام (Bingham)، دراسات على المشكلات النفسية التطبيقية من بينها قياس الميول.

في عام 1931 أصدر فراير (Fryer) كتابه بعنوان (Measurement of Interests)، الذي لخص فيه التطورات السابقة في هذا المجال. أشار فيه إلى محاولة مور (Moore) عام 1921 لقياس الميول الميكانيكية والاجتماعية لدى المهندسين، ومحاولة ريم (Ream) عام 1924 لتمييز الناجحين من غير الناجحين في المهن التجارية بناءً على ميولهم، بالإضافة إلى بحوث فرايد (Freyd) عام 1923 التي استخدمت مقاييس مور في محاولة التمييز بين الميول الميكانيكية والاجتماعية لدى أفراد من عامة الناس. شهد العقد الثاني من القرن العشرين بداية بناء استبيانات ميول مقننة بفضل جهود سترونج (E.K. Strong)، الذي عين أستاذًا في جامعة ستانفورد عام 1924 بناءً على توصية تيرمان (Terman)، الذي كان يجري بحوثًا على الطلاب ذوي الذكاء العالي. تأثر سترونج بنتائج حلقة البحث المشار إليها فقام بإعداد استبيان الميول المهنية عام 1927، حيث صدرت الطبعة الأولى له في نفس العام وتلاه دليل الاستخدام والوثائق المرجعية عام 1928، كما ظهرت أول معلومات حول تصحيحه آليًا عام 1930. ونظرًا لشيوع استخدام هذا الاستبيان أُجريت عليه العديد من الدراسات الطولية والتعديلات لتوسيع نطاقه حتى صدرت طبعته الحديثة التي تعتمد على التصحيح الحاسوبي عام 1974. من التطورات التاريخية المهمة الأخرى في قياس الميول كان استبيان كيودر (Kuder) عام 1939 المعروف بـ(Kuder Preference Record)، والذي يعتمد على القياس الموضوعي للميول بخلاف استبيان سترونج. كما ساهم علماء النفس في سلاح الجو الأمريكي أثناء الحرب العالمية الثانية في تطوير بحوث القياس الموضوعي للميول. في عام 1943 أعد ثورب (Thorpe) استبيان الميول المهنية (Occupational Interest Inventory)، الذي اختلف في أسلوبه ومنهجيته عن

الاستبيانات السابقة. وفي عام 1948 أعد جيلفورد (Guilford) ومعاونوه استبيان (Guilford's Interest Survey) والذي استخدم منهج التحليل العاملي للحصول على درجات منفصلة للميول المهنية وغير المهنية.

ونظرًا لاعتماد معظم مقاييس الميول على الجانب اللفظي، أعد جايست (Geist) استبيان الصور (Geist Picture Inventory)، لكنه لا توجد أدلة كافية تدعم صدق هذه الاستبيانات غير اللفظية. بعد ذلك ازداد الاهتمام بقياس الميول غير المهنية أي تلك المتعلقة بالأنشطة الحياتية المختلفة، مثل: استبيان كيودر لمسح الميول العامة الذي نُشر عام 1964، بالإضافة إلى بناء استبيانات مخصصة لخريجي المدارس الثانوية الذين لا يستمرون في الدراسة وإنما يلتحقون بسوق العمل، كمثال استبيان (Minnesota Occupational Interest Survey). ولا تزال الدراسات والبحوث مستمرة حتى الوقت الحاضر في هذا المجال، لا سيما في بحث علاقة الميول بجوانب السلوك المعرفية وغير المعرفية (علام، 2000، ص469).

1- تعريف الميول (Interest):

الميول (Interest) هي شعور إيجابي يتولد لدى الفرد تجاه نشاطات أو موضوعات أو أفكار معينة تمنحه المتعة والاهتمام، وتُعد اختبارات الميول أحد فروع اختبارات الشخصية في علم القياس النفسي. تكتسب هذه الاختبارات أهمية كبيرة في مجالات التوجيه والإرشاد المهني والتربوي، إذ تُسهم في تحديد الاتجاهات المهنية والتعليمية المناسبة للأفراد وتعد من الركائز الأساسية في تطوير الإرشاد المهني. يعتمد اختبار الميول على مجموعة من البنود أو الأسئلة تتعلق بأنشطة واهتمامات وأوصاف لأشخاص يُطلب من المفحوص تحديد مدى تفضيله لها. ورغم أن الميول ليست العامل الوحيد في تحديد الاتجاه المهني أو التعليمي فإنها تُعد من العوامل الجوهرية التي ينبغي أخذها بعين الاعتبار. وفي الولايات المتحدة الأمريكية، تحظى هذه الاختبارات بإقبال واسع حيث يُقدّر عدد تطبيقاتها السنوية بأكثر من ثلاثين مليون اختبار مما يعكس أهميتها ومكانتها في الميدان النفسي.

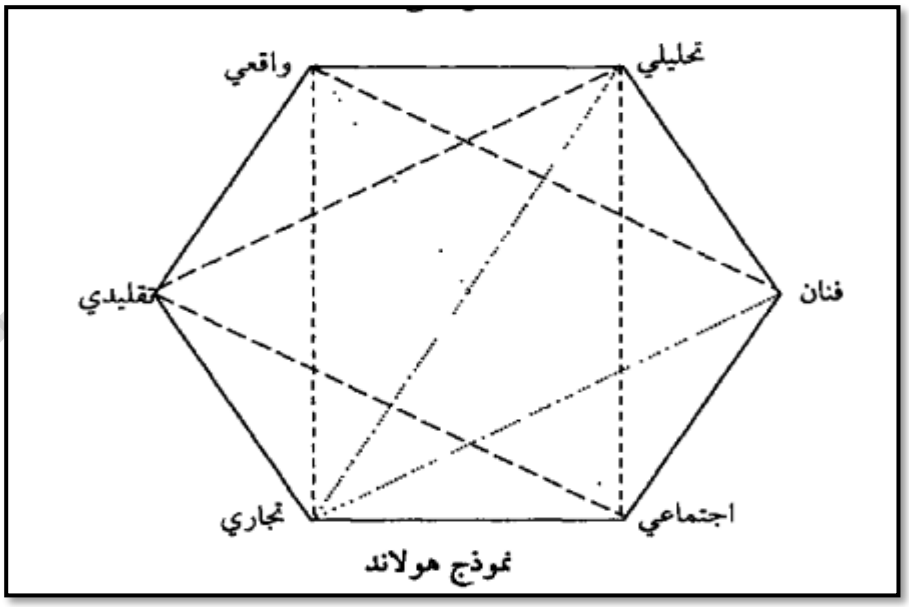
يرتبط قياس الميول ارتباطاً وثيقاً بقياس الشخصية لأن الميول تُعد انعكاساً لخبرات الحياة ولسمات الفرد الشخصية. وقد ذهب التحليل النفسي إلى أن الميول المهنية تتأثر بالدوافع اللاشعورية؛ فالإعلاء (Sublimation) يحوّل الدوافع العدوانية أو الغريزية إلى أنشطة مهنية مقبولة اجتماعياً، كميل ذوي النزعات السادية إلى الجراحة أو أصحاب الميول الجنسية المكبوتة إلى الفنون والشعر.

كما ربطت مدرسة التحليل النفسي بين اختيار المهنة وفكرة التوحد (Identification)، حيث يُلاحظ أن الأبناء الذين يتوحدون مع آبائهم يميلون إلى اختيار المهنة نفسها. غير أن هذا التفسير يُقابَل ببعض التحفظ إذ قد يكون الدافع الحقيقي هو المكانة الاجتماعية أو العائد الاقتصادي للمهنة، لا مجرد التوحد النفسي (ربيع، 2014، ص221).

2- قياس الميول (Interest Measurement):

يرتبط قياس الميول ارتباطًا وثيقًا بقياس الشخصية (Personality Measurement)، لأن الميول تُعد نتاجًا للخبرات المكتسبة وتعكس حاجات الفرد وسماته الشخصية. وقد تطورت نماذج قياس الميول من خلال أعمال عدد من علماء النفس الذين سعوا إلى الربط بين الاتجاهات المهنية وأنماط الشخصية. من أبرز هذه النظريات نظرية "هولاند" (Holland's Theory of Vocational Personalities and Work Environments)، التي قدّمها جون هولاند (John L. Holland) سنة 1959، وطُوّرت لاحقًا في إصدارها الأحدث المعروف باسم Self-Directed Search (SDS, 5th Edition, 2017) (Holland, 2017). ترتكز النظرية على ستة أنماط للشخصية، هي: الواقعي (Realistic)، التحليلي أو الاستقصائي (Investigative)، الفني (Artistic)، الاجتماعي (Social)، الريادي أو التجاري (Enterprising)، والتقليدي (Conventional). ويرى "هولاند" أن التطابق بين نمط شخصية الفرد ونمط البيئة المهنية التي يعيش فيها يؤدي إلى رضا مهني وتوافق نفسي أكبر.

شكل رقم 05 يوضح نموذج هولاند لأنماط الشخصية المهنية.



(ربيع، 2014، ص226)

أما اختبار سترونج للميول المهنية (Strong Vocational Interest Blank – SVIB) فقد طوره إدوارد سترونج (Edward K. Strong, Jr.) سنة 1927 بهدف قياس الفروق بين الأفراد في تفضيلاتهم وميولهم نحو الأنشطة والمهن. يعتمد الاختبار على مقارنة ميول الفرد بميول أشخاص ناجحين في مهن معينة لتحديد المهنة المناسبة له. وقد تم تطويره لاحقًا في صورة أكثر شمولًا تحت اسم Strong-Campbell Interest Inventory (SCII)على يد ديفيد كامبل (David P. Campbell) سنة 1974، ثم أُعيدت مراجعته في صورته الحديثة المعروفة باسم Strong Interest Inventory (SII, 2021 Edition) (Yazak-1, 2024). يتضمن هذا الاختبار مجموعة واسعة من البنود التي تقيس تفضيلات المفحوص نحو المهن، والمواد الدراسية، والأنشطة اليومية، والصفات الشخصية، وهو موجه أساسًا للمراهقين والراشدين.

وفي سياق موازٍ يعد اختبار "كودر" للميول المهنية العامة (Kuder Preference Record – Vocational, KPR-V) من الأدوات الكلاسيكية التي وضعها فريدريك كودر (Frederic Kuder) سنة 1938، وتم تطويره إلى نسخته الحديثة Kuder Career Interests Assessments (KCIA, 2013) (Yazak-2, 2024). يقيس الاختبار عشرة مجالات من الميول: الميكانيكي، الحسابي، العلمي، الإقناعي، الفني، الأدبي، الموسيقي، الاجتماعي، الكتابي، والخلوي (العملي). ويهدف إلى تحديد الميول الغالبة لدى المراهقين لتوجيههم أكاديميًا ومهنيًا. كما طوّر "كودر" لاحقًا قائمة كودر لمسح الميول المهنية (Kuder Occupational Interest Survey – KOIS) سنة 1956، وهي أداة تقيس ميول الأفراد نحو مجالات دراسية ومهنية متعددة، وما تزال تُستخدم بنسخها الحديثة في برامج الإرشاد المهني.

أما اختبار "جوهانسون" للتقييم المهني (Johansson Vocational Evaluation Test)، الذي وضعه كارل جوهانسون (Carl Johansson) في خمسينيات القرن العشرين، فيستخدم لتقدير الميول المهنية للأفراد متوسطي التعليم من خلال قياس اهتماماتهم بمجالات الحياة اليومية والمواد الدراسية والمهن العملية مما يجعله مناسبًا لبرامج التدريب والتأهيل المهني (ربيع، 2014، ص239).

3- بعض المشكلات المتعلقة بالإستبيانات المقننة للميول:

تواجه الاستبيانات المقننة التي تستخدم لقياس الميول، والاتجاهات، وسمات الشخصية، عدداً من المشكلات المنهجية والسلوكية، من أهمها ما يلي:

أ- تزييف أو تحيز الاستجابـات (Faking): تعد هذه المشكلة من أبرز القضايا التي تناولتها البحوث النفسية وخاصة منذ دراسة كرونباخ (1950). ففي حين يصعب على الفرد تزيف إجاباته في اختبارات الذكاء أو الاستعدادات الخاصة يمكنه بسهولة القيام بذلك في الاختبارات الوجدانية كاستبيانات الميول أو الشخصية من خلال إعطاء صورة غير واقعية عن ذاته.

ويزداد احتمال التزيف عندما يكون الهدف من التطبيق الانتقاء المهني أو الوظيفي حيث يسعى الفرد لتقديم انطباع إيجابي عن نفسه مقارنةً بحالات التوجيه التربوي أو المهني. كما قد يتعمد البعض تزيف استجاباتهم لأغراض أخرى كالتهرب من الخدمة العسكرية أو تجنب العقوبات. وقد يكون هذا التزيف أحياناً شبه شعوري ناتج عن دوافع خفية أو نزعات انفعالية خصوصاً لدى المراهقين الذين يعيشون صراعات داخلية مع السلطة أو المجتمع.

ب- الاتجاه العقلي في الاستجابة (Response Sets): تتأثر الاستجابات الذاتية أيضاً بأنماط عقلية شبه شعورية تجعل الأفراد يميلون إلى الاستجابة بطريقة معينة بغض النظر عن مضمون الفقرات. ويعرف هذا النمط باسم الاتجاه العقلي أو "مجموعة الاستجابة"، حيث يستجيب الفرد وفق أسلوب ثابت نسبياً يعكس أسلوبه الشخصي في التعامل مع المواقف الاختبارية أكثر مما يعكس السمة المقاسة. وقد أشار جاكسون وميسك (1958) إلى أن هذا النسق من الاستجابة مستقر نسبياً عبر الزمن ويستحق الدراسة في حد ذاته. وتشير الدراسات إلى وجود نمطين رئيسيين لتباين الاتجاه العقلي:

الأول هو النمط الانتقالي، ويظهر عندما يتأثر الفرد بموقف اختباري أو استبيان محدد، أما الثاني فهو النمط المتسق، ويتسم بدرجة من العمومية والاستقرار حيث يؤثر في استجابات الفرد عبر مواقف واختبارات متعددة. كما يمكن التمييز بين نوعين من هذا الاتجاه العقلي المستقر:

أحدهما يعكس فروقاً فردية بسيطة تنشأ نتيجة عوامل النمو والتنشئة، مثل أسلوب استخدام اللغة أو العادات التعبيرية، وهي جوانب لا تمثل أهمية كبيرة من الناحية النفسية. أما النوع الآخر فيعكس جوانب أساسية في بنية الشخصية ويُعد التباين المرتبط به مؤشراً لخصائص شخصية مستقرة نسبياً.

ويطلق على هذا النمط الثاني اسم أسلوب أونسق الاستجابة (Response Style)، وهو يشير إلى ثبات الطريقة التي يستجيب بها الفرد لفقرات الاستبيانات بغض النظر عن محتواها أي أنه يعكس نمطاً عاماً في كيفية الإجابة أكثر من مضمون الإجابة نفسها. ويتميز أسلوب الاستجابة بدرجة من الاستقرار النسبي عبر الزمن، كما يساهم في تعميم نمط الاستجابات في المواقف الاختبارية وغير

الاختبارية على حد سواء، ولذلك يعد من الضروري دراسته بوصفه ظاهرة قائمة بذاتها (Jackson & Messick, 1958).

وفيما يلي سيتم عرض الأنواع الرئيسة لأساليب الاستجابة (Response Sets) التي تناولتها معظم أدبيات القياس النفسي والتربوي:

أ- الإذعان أو الموافقة (Acquiescence): يعني هذا الاتجاه العقلي ميل الفرد إلى الموافقة على فقرات الاستبيان بغض النظر عن محتواها فيجيب مثلاً بـ"نعم" أو "موافق" دوماً. وتؤدي هذه النزعة إلى تشويه دلالات الدرجات الناتجة عن الاستبيان لأنها تعكس الميل العام للموافقة أكثر مما تعكس السمة المقاسة. ويقترح سميث (Smith, 1961) أن التحكم في هذه المشكلة ممكن جزئياً من خلال صياغة الفقرات بأسلوب متوازن (إيجابية وسلبية)، بينما أشار إيليوت (Elliott, 1961) إلى أن درجة الإذعان تتأثر بثلاثة عوامل متداخلة: شكل الفقرة (عبارات موجبة، عبارات سالبة، اختيار من متعدد)، علاقة محتوى الفقرة بالسمة الشخصية، والمستوى العقلي للفرد المستجيب.

ب- مراعاة القبول الاجتماعي (Social Desirability): يعد هذا النمط من الاستجابة شكلاً من أشكال التحيز الاجتماعي، حيث يحاول الفرد إظهار نفسه بصورة مقبولة اجتماعياً من خلال اختيار الإجابات التي تعكس التكيف، والمساعدة، والالتزام بالمعايير الأخلاقية. وقد بينت دراسات كراون ومارلو (Crown and Marlowe, 1964)، أما إدواردز (Edwards, 1957) يرى أن هذا الاتجاه يرتبط بحاجة الفرد إلى حماية نفسه والامتثال للمعايير الاجتماعية.

وللحد من تأثير الاتجاه العقلي في الاستجابات، اقترح إدواردز (Edwards) استخدام الاستبيانات ذات الاختيار الجبري (Forced-Choice Inventories). في هذا النوع من الاستبيانات تُعرض على المفحوص أزواج من العبارات المقبولة اجتماعياً والمتكافئة في درجة مقبوليتها ومحتواها وصدقها، بحيث تكون كل عبارة في الزوج مرتبطة بدرجة متماثلة بـ محكات إمبيريقية مناسبة، حيث يُطلب من الفرد أن يختار العبارة التي يراها أقرب إلى شخصيته، فإذا كانت العبارة المختارة مرتبطة بالمحك يحصل على درجة واحدة، أما إذا اختار عبارة غير مرتبطة بالمحك فيُمنح صفراً. ويمكن توسيع هذا الأسلوب بحيث تُقدَّم عدة بدائل للفقرة الواحدة بدلاً من الاقتصار على زوج من العبارات.

يتضح تطبيق هذا النوع من الفقرات في استبيانات الميول، وخاصة في استبيان كودر (Kuder) الذي استخدم صيغة الاختيار الجبري. وعلى الرغم من أن هذه الصيغة توفر معلومات أكثر دقة وتفصيلاً مقارنة بالاستبيانات التقليدية، إلا أنها ليست بمنأى عن التزييف إذ يصعب إيجاد عبارتين

متماثلتين تماماً في جميع الجوانب. كما أن بعض الأفراد الراغبين في إظهار صورة إيجابية عن أنفسهم قد يتمكنون من تمييز البديل الأكثر قبولاً اجتماعياً. إضافة إلى ذلك فإن تطبيق الاستبيانات ذات الاختيار الجبري يتطلب وقتاً أطول للحصول على عدد مكافئ من الاستجابات مقارنة بالاستبيانات غير الجبرية.

وفيما يلي مثال توضيحي على فقرة من الاستبيانات ذات الاختيار الجبري (Forced-Choice Inventories) كما اقترحها إدواردز (Edwards):

- يرجى اختيار العبارة التي تصفك بدقة أكبر، علمًا أن كلتا العبارتين مقبولتان اجتماعيًا بنفس الدرجة:

أ- أحب مساعدة الآخرين في حل مشكلاتهم.

ب- أستمتع بتنظيم الأنشطة الجماعية والتخطيط لها.

في هذا المثال كلتا العبارتين إيجابيتان ومقبولتان اجتماعيًا، لكن إحداهما (أ) ترتبط أكثر بميل اجتماعي أو إرشادي، بينما الأخرى (ب) ترتبط بميل قيادي أو إداري. يطلب من المفحوص اختيار العبارة التي تنطبق عليه أكثر، ولا يمكنه اختيار كلتيهما أو تجنب الإجابة. وبناءً على العلاقة بين كل عبارة والمحك التجريبي (Empirical Criterion)، يحصل الفرد على درجة واحدة إذا اختار العبارة المطابقة للمحك، وصفر إذا اختار الأخرى.

ففي استبيان كودر للميول المهنية (Kuder Occupational Interest Survey)، تستخدم مثل هذه الفقرات بكثرة، حيث يُطلب من المفحوص أن يختار من بين عدة أنشطة مقبولة اجتماعيًا النشاط الذي يفضّله أكثر، مثل: أ- العمل مع الأطفال؛ ب- إصلاح الأجهزة الميكانيكية؛ ج- كتابة القصص.

وبهذا يمكن تحديد مجالات الميول المهنية دون أن يتأثر المفحوص برغبته في تقديم صورة اجتماعية مثالية عن نفسه.

ج- الاستجابات الشاذة (Atypical Responses): يميل بعض الأفراد إلى اختيار استجابات غير مألوفة أو مخالفة للنمط السائد بغض النظر عن مضمون الفقرة، مثل اختيار استجابات لفقرات استبيان معين على ميزان تقدير يتراوح بين: أهتم جداً أو أوافق جداً، ولا أهتم، أو لا أوافق على الاطلاق. وهنا يلجأ الأفراد إلى اختيار استجابات تختلف عن الاستجابات الشائعة لجماعة مرجعية معينة. وقد تناول سيشرست وجاكسون (Sechrest and Jackson, 1963) هذا النوع من النزعات بالتحليل المفصل وكيفية قياسها وتفسيرها، موضحين أن اختيار استجابات شاذة قد يعكس رغبة الفرد في التميز أو رفض المعايير الاجتماعية للجماعة المرجعية.

د- التحيز للذكور (Male Bias): تعرضت استبيانات الميول المهنية للنقد بسبب تحيزها للذكور حيث كانت توجه الإناث نحو المهن التقليدية كالتدريس أو التمريض. وقد دعت منظمات حقوق المرأة إلى مراجعة هذه المقاييس لتشجيع الإناث على اختيار المهن العليا.

وبالفعل أُعيد بناء المقاييس لتصبح موحدة بين الجنسين غير أن هذا التعديل - كما يرى كامبل (Campbell, 1977)- قد قلل من صدق المقاييس لأن ميول الذكور والإناث تختلف فعلاً في الواقع النفسي والاجتماعي. ولذا تم في بعض المقاييس الحديثة الدمج الجزئي بين معايير الجنسين لتقليل التحيز دون إلغائه بالكامل (علام، 2000، ص 479).

ثالثاً: الاتجـاهــات (Attitudes):

تعد الاتجاهات مؤشرات يمكن من خلالها التنبؤ بسلوك الفرد في مواقف مستقبلية. فعلى سبيل المثال، قد يؤثر اتجاه الطالب نحو الكتب المدرسية في قدرته على تعلم القراءة، كما أن اتجاهه نحو المدرسة وبرامجها قد ينعكس على سلوكه وأدائه الدراسي. وبالمثل فإن اتجاه الفرد نحو المؤسسة التي يعمل بها أو نحو المشرفين على عمله أو الأدوات التي يستخدمها قد يؤثر في جودة أدائه المهني. وكما يميل الأفراد إلى ممارسة أنشطة معينة أو يفضلونها فقد تكون لديهم أيضًا اتجاهات إيجابية أو سلبية بدرجات متفاوتة نحو أشخاص أو جماعات أو موضوعات أو أفكار أو عادات محددة. ومن هنا يمكن التمييز بين الميول والاتجاهات من حيث الإطار المرجعي الذي يشير إليه كل منهما رغم أنهما مترابطان إلى حد كبير. فقد يكون الفرد مثلًا يميل إلى تنسيق الزهور (نشاط) ولديه اتجاه إيجابي نحو الزهور بصفة عامة (أشياء)، بينما قد يميل إلى التدريس (نشاط) مع امتلاكه اتجاهًا سلبيًا نحو الأطفال (جماعة من الأفراد).

ويعرف ألبورت (Allport, 1954) الاتجاه بأنه استعداد نفسي عضوي منظم بالتجربة يؤثر في استجابات الفرد تجاه المواقف والموضوعات المرتبطة به. إلا أن الاتجاهات لا تكون شاملة لكل المواقف أو الموضوعات أو الجماعات أو الأشخاص، إذ قد تثير كل فئة من المثيرات انفعالات مختلفة في شدتها ونوعها. لذا فإن السلوك الوجداني الذي يعكس اتجاهًا معينًا يتأثر جزئيًا بالظروف المحيطة ووقت ظهور الموقف (علام، 2000، ص514).

1- قياس الاتجاهات من منظور تاريخي:

لم يتطور قياس الاتجاهات بمعزل عن علم النفس الاجتماعي إذ لم تظهر محاولات جادة في هذا المجال قبل نهاية العقد الثاني من القرن العشرين، حيث اقتصرت الجهود المبكرة على محاولات

متفرقة. ففي عام 1925 وضع بوجاردس (Bogardus) مقياس المسافة الاجتماعية (Social Distance Scale) لقياس درجة تقبل الأفراد لأشخاص من جنسيات وعرقيات مختلفة في الولايات المتحدة، حيث طُلب من المشاركين تحديد مدى قبولهم لهؤلاء كأصدقاء أو جيران أو أزواج أو مواطنين.

وفي العام نفسه أجرى ألبورت وهارتمان (Allport & Hartman, 1925) دراسة لقياس اتجاهات الطلاب نحو قضايا سياسية متنوعة، وأكدا الحاجة إلى مقاييس مقننة للاتجاهات. وقد صنفا مجموعة من الفقرات المتعلقة بهذه القضايا وطلبا من محكّمين تقدير أهميتها ثم استخدما متوسط التقديرات لتحديد قيم تمثل كل فقرة بحيث يختار الطالب ما يعبر عن رأيه بدرجة أدق. لفت هذا العمل انتباه ثيورستون (L. Thurstone) الذي كان آنذاك يطور أسلوب الحكم المقارن (Comparative Judgment) في إطار التحليل السيكوفيزيائي، بغرض بناء موازين نفسية لخصائص لا يمكن قياسها فيزيائياً كالخصائص الجمالية أو الانفعالية. طوّر "ثيورستون" هذا المنهج وطبّقه على فقرات مقياس ألبورت وهارتمان، ممهداً بذلك لأساس علمي لتوزيع الآراء في مقاييس الاتجاهات بحيث تمثل فروقاً متساوية ظاهرياً بين المواقف.

وفي عام 1928، نشر ثيورستون مقالته الشهيرة بعنوان "الاتجاهات يمكن قياسها"، قدّم فيها أسلوب الفقرات المتساوية ظاهرياً (Equal Appearing Intervals)، الذي يطلب من المستجيب تصنيف الفقرات في تسع أو إحدى عشرة فئة مرتبة تمثل تدريجاً متساوياً في الموقف. ومن خلال التحليل الرياضي أو أسلوب الحكم المقارن يتم استخراج قيم عددية تمثل شدة الاتجاه نحو كل فقرة. وقد أسهم هذا الأسلوب في تطوير مقاييس الاتجاهات الاجتماعية وأساليب القياس النفسي عمومًا، كما كان الأساس الذي استندت إليه مقاييس مثل مقاييس ريمرز ومعاونيه (Remmers et al.) لقياس الاتجاهات نحو مؤسسات ومجالات خاصة مختلفة. وفي عام 1932، قدم ليكرت (R. Likert) مدخلًا جديدًا أكثر بساطة هو أسلوب التقديرات المجمعة(Summated Ratings) والذي يعتمد على تحديد درجة الموافقة أو عدم الموافقة على كل فقرة ضمن مقياس خماسي التدرج (من "أوافق بشدة" إلى "لا أوافق إطلاقاً")، بحيث تُجمع الدرجات للحصول على المجموع الكلي الذي يمثل اتجاه الفرد. وقد استند ليكرت في تطوير طريقته إلى أساليب تحليل المفردات المستخدمة في الاختبارات العقلية.

ثم جاء جتمان (Guttman, 1944) ليقترح أسلوب تحليل المصفوفات المتدرجة (Scalogram Analysis) الذي يرتب الأفراد والفقرات وفقاً لأنماط استجاباتهم بحيث يكون الفرد الذي يوافق على فقرة معينة موافقاً كذلك على جميع الفقرات الأدنى منها رتبة.

وفي عام 1950، قدّم لازارسفيلد (Lazarsfeld) نموذج التحليل البنيوي الكامن (Latent Structure Analysis) وهو نموذج رياضي عام يربط بين احتمال استجابة الفرد لفقرة معينة وموقعها على متغير كامن مثل الاتجاه. وتُعدّ أساليب ليكرت وجتمان والتحليل العاملي تطبيقات خاصة لهذا النموذج عند تحقق شروط معينة. وفي الفترة نفسها قدّم كومبس (Coombs, 1950) نظرية البيانات (Theory of Data) وهي إطار عام لتدريج استجابات التفضيل، في حين استخدم أدورنو ومعاونوه (Adorno et al.) في دراستهم عن الشخصية السلطوية (The Authoritarian Personality) مجموعة متنوعة من الأساليب غير المباشرة مثل الأسئلة الإسقاطية المفتوحة والمقابلات الكلينيكية، بهدف الكشف عن الاتجاهات الحقيقية للأفراد. وهكذا شكلت هذه الجهود المتعاقبة الأساس المنهجي لتطور قياس الاتجاهات وربطته ارتباطاً وثيقاً بعلم النفس الاجتماعي وأساليبه البحثية الحديثة (علام، 2000، ص514).

2- تصنيف مقاييس الاتجاهات:

تهدف جميع مقاييس الاتجاهات إلى استنتاج معلومات حول اتجاهات الأفراد اعتمادًا على أدلة قابلة للملاحظة. ورغم اختلاف طرق تصنيف هذه المقاييس إلا أنها تشترك في عناصر أساسية، ويمكن تمييز ثلاثة أنماط سلوكية (Attitude Universes) يطلب فيها من الفرد تقديم استجابة لفظية تعبّر عادةً عن رأيه أو مشاعره:

أ- الاتجاهات اللفظية المستدعاة (Elicited Verbal Attitudes): هي الاستجابات اللفظية المطلوبة من الفرد عند سؤاله مباشرةً عن رأيه أو شعوره تجاه موضوع معيّن. هذه الفئة تمثّل الغالبية العظمى من الموازين المستعملة لقياس الاتجاهات.

ب- الاتجاهات اللفظية التلقائية (Spontaneous Verbal Attitudes): تمثل التعبيرات اللفظية غير المستدعاة والتي يظهرها الفرد طواعيةً في حديثه الاعتيادي مع الآخرين، وتُعدّ مؤشراً على مواقف أكثر تلقائية وأقل ضبطًا من جانب المستجيب.

ج- الاتجاهات المتمثلة في الفعل (Action Attitudes): تتعلق بالسلوك الظاهر- لفظيًا أو غير لفظي- الموجه نحو موضوع الاتجاه، أي ما يفعله الفرد فعليًا تجاه الشيء محل الاتجاه.

يمكن ترميز وتعيين قيم عددية لاستجابات كل من هذه النطاقات وفق قواعد مناسبة. وعلى الرغم من أن معظم الموازين تركز على الفئة الأولى (الاتجاهات اللفظية المستدعاة)، فإنه بالإمكان تصميم أدوات لقياس الفئتين الثانية والثالثة بواسطة ملاحظة السلوك أو المقابلات المصممة بعناية.

إلى جانب الأساليب المباشرة توجد تقنيات متخفية أو غير مباشرة (Disguised Techniques)، مثل الأساليب الإسقاطية بمختلف أنواعها- وأيضًا طرق فسيولوجية تقيس مؤشرات جسدية للاستثارة.

وبناءً على نوع الدليل (Indicators) المستخدم للاستدلال على اتجاه الفرد، تصنف مقاييس الاتجاهات عادةً كما يلي:

القسم الأول: أساليب بناء الموازين (Scaling Techniques): تضم الطرق التي تستنتج الاتجاه اعتمادًا على استجابات الفرد تجاه فقرات أو صفات محددة، وتعرف أدواتها بـالموازين (Scales). ومن الأمثلة الشائعة لها: موازين ثيورستون (Thurstone Scales)، ليكرت(Likert Scales) ، وجتمان (Guttman Scales). وهناك أساليب خاصة أخرى كـالتمييز الدلالي (Semantic Differential) لأوزجود، وطريقة الفرد (Unfolding) لكومبس، وطريقة (Q) لستيفنسون، والموازين متعددة الأبعاد. تختلف هذه الموازين عن اختبارات الجوانب المعرفية في أن نتائجها لا تُحكم عليها بصواب أو خطأ، بل تقيس درجة امتلاك الفرد لخاصية وجدانية معينة، مثلاً: اتجاه طالب الثانوية نحو التعلم الذاتي أو اتجاه طالب الجامعة نحو التعليم المستمر.

القسم الثاني: استدلال مبني على السلوك الظاهر (Overt Behavior): يتضمن طرقًا تعتمد على ملاحظة السلوك في مواقف حقيقية، ما يعزّز صدق الاستدلال إذ لا يتطلب تعاون المستجيب ولا يجعله يشعر أنه في وضع اختباري خاص. وقد أوضح Sechrest (1969) ضرورة عدم الاعتماد الكامل على موازين الاتجاهات وحدها، موصيًا بدمج ملاحظة السلوك معها، إذ أظهرت مراجعات مثل مراجعات Wiker (1969) أن العلاقة بين ما تقيسه الموازين والسلوك الظاهر غالبًا ما تكون منخفضة. ومن أمثلة أساليب ملاحظة السلوك :لعب الأدوار (Role Playing)، والاختيارات السوسيومترية (Sociometric Choices).

القسم الثالث: أساليب الاستدلال العميق غير المباشر: تشمل طرقًا تقدم المثيرات بطريقة يفهمها المستجيب بمعنى قد يختلف عن المقصود من قِبَل الباحث؛ أي أنها تقيس ردودًا ذات مغزى لدى المستجيب بينما يُفسَّر الباحث الاستجابة ضمن نظام معنى آخر. من أمثلتها الطرق الإسقاطية التي تستخدم لالتقاط مستويات أعمق من الاستجابة لدى الفرد.

القسم الرابع: المقاييس الفسيولوجيـة: تستند إلى مؤشرات جسمية كالاستجابة الجلفانية للجلد، اتساع أو ضيق الحدقة، أو معدل نبض القلب عند عرض مثيرات متعلقة بموضوع الاتجاه. هذه المؤشرات تكشف مستوى الاستثارة أو الشدة، لكن لا تحدد اتجاه هذه الاستثارة تجاه موضوع معين. أشار

Fishbein & Ajzen (1965) إلى أهمية القياسات غير اللفظية الخارجة عن تحكم المستجيب، لكنهما شككا في قدرة مقاييس رد الفعل الفسيولوجي الحالية على تحقيق مقياس اتجاهي واضح.

ونظراً لشيوع وسهولة تفسير طرق القسم الأول (موازين بناء الموازين)، سنناقشها تفصيليًا فيما يلي، مع عرض موجز لبعض الأساليب الأخرى (علام، 2000، ص529).

3- أساليب بنـاء موازين الاتجـاهـات:

3-1- أسلوب الفترات المتساوية ظاهرياً لثيرستون: يعد هذا الأسلوب من المحاولات المبكرة التي قام بها ثيرستون (Thurstone, 1929, 1931) في مجال بناء مقاييس الاتجاهات. فقد قدّم مجموعة من الأساليب المستمدة من تجارب علم النفس الفيزيائي (Psychophysics) التي تهدف إلى بناء موازين لأحكام الأفراد حول خصائص فيزيائية مختلفة، مثل الوزن. إذ كلما كان الفرق الحقيقي بين وزني جسمين ضئيلاً، قل عدد الأفراد القادرين على التمييز الصحيح بينهما. وانطلاقاً من هذا المبدأ حاول ثيرستون تطوير موازين للاتجاهات من خلال مطالبة الأفراد بالمقارنة بين فقرات أو عبارات مرتبطة باتجاه معين، مثنى مثنى، والحكم على أيٍّ منها أكثر إيجابية أو سلبية من الأخرى من حيث شدة الاتجاه (Attitude Intensity). وقد أطلق على هذه الطريقة أسلوب المقارنات الثنائية (Paired-Comparisons). غير أن هذا الأسلوب يصبح صعب التطبيق عندما يكون عدد الفقرات كبيراً، كما هو الحال في معظم مقاييس الاتجاهات؛ فوجود 10 فقرات مثلاً يتطلب إجراء مقارنة ثنائية.

$$45 = \frac{(1-10)10}{2}$$

ولتجاوز هذا التعقيد اقترح ثيرستون أسلوباً آخر أقل جهداً هو أسلوب الفترات المتساوية ظاهرياً (Equal-Appearing Intervals)، والذي أصبح من أكثر الأساليب شيوعاً في بناء مقاييس الاتجاهات.

3-1-1- خطوات بناء مقياس الاتجاه وفق أسلوب ثيرستون:

أ- تجميع مجموعة كبيرة من الفقرات المرتبطة بموضوع الاتجاه المراد قياسه، وليكن مثلاً 100 فقرة، بحيث تعبر عن درجات مختلفة من التفضيل تتراوح بين الموافقة التامة وعدم الموافقة التامة. ويمكن الاستعانة في صياغتها بالأدبيات السابقة والمقابلات الاستطلاعية. ويشترط أن تكون الفقرات:

- واضحة ودقيقة وتعبر عن فكرة واحدة.
- خالية من الغموض والحقائق الموضوعية.

- تسمح باستجابات متفاوتة تبعاً لاختلاف شدة الاتجاه لدى الأفراد.
- تكتب كل فقرة على بطاقة صغيرة.

ولتوضيح ذلك إذا كان الاتجاه المراد قياسه نحو التعليم، يمكن أن تكون بعض الفقرات كالتالي:

- لا يمكن أن نكون مواطنين صالحين بدون تعليم.
- أذهب إلى المدرسة فقط لأنني مجبر على ذلك.
- ينبغي أن يحظى التعليم بمزيد من التمويل.
- لا يجب أن أبذل جهداً كبيراً إلا للنجاح في الامتحانات.
- التعليم مفيد في حياتنا المعاصرة.
- أشك فيما إذا كان التعليم قد أسهم في تطوير الحياة أم لا.

ب- تعرض هذه البطاقات على عدد من المحكمين يتراوح عادة بين 40 و60 محكماً، ويطلب منهم تصنيفها في 9 أو 11 مجموعة وفق درجة التفضيل أو الأفضلية (Favorability Scale). يعطى الرقم (1) لأعلى درجة تفضيل، والرقم (11) لأدناها، بينما يمثل الرقم (6) النقطة المحايدة (Neutral Point). ويفترض أن تكون المسافات بين المجموعات متساوية ظاهرياً في نظر المحكم. ويجب على المحكمين عدم الحكم وفقاً لآرائهم الشخصية، بل وفق إدراكهم لمدى إيجابية أو سلبية كل فقرة تجاه موضوع الاتجاه

- **فمثلاً**، عند تقييم الفقرة "لا يمكن أن نكون مواطنين صالحين بدون تعليم"، يطلب من المحكم أن يحدد مدى إيجابية مضمونها بالنسبة لموضوع التعليم بغضّ النظر عن موقفه الشخصي منها، بوضع علامة (*) على الميزان التالي:

تكرر هذه العملية لكل فقرة ويراعى أن يكون المحكمون على معرفة كافية بالمفهوم أو الموضوع والجمهور المستهدف بالمقياس.

ج- يحسب الوسيط (Median) لكل فقرة من توزيعات المحكمين، ويعد هذا الوسيط قيمة ميزان الفقرة (Item Scale Value) التي تحدد موقعها على متصل الأفضلية (Favorability Continuum). وبذلك تشير إلى شدة الاتجاه (Intensity Scale).

د- يقاس تشتت التوزيع باستخدام مقياس نصف المدى الربيعي، كونه يعتمد على الوسيط ويعبر عن مدى تمركز 50% من الحالات الوسطى. ويعد هذا المقياس أكثر ملاءمة من غيره لأنه يحدد بوضوح درجة التباين في أحكام المحكمين حول كل فقرة. فإذا كانت قيمة نصف المدى الربيعي لفقرة معينة مرتفعة، فإن ذلك يشير إلى وجود تشتت كبير في التقديرات، أي أن الفقرة غامضة أو غير واضحة بالنسبة للمحكمين، الأمر الذي يستوجب استبعادها من النسخة النهائية لمقياس الاتجاه. وباستبعاد الفقرات التي تتسم بتشتت مرتفع، تُحتفظ فقط بتلك التي تتوزع قيم وسائطها (Scale Medians) بشكل متوازن على طول المتصل الافتراضي للاتجاه، بحيث تغطي جميع درجاته من الإيجابية إلى السلبية. فقد نحصل مثلاً على قيم وسيط متتابعة كالتالي: 1.7، 1.8، 2.2، 2.2، 2.9، 3.0، 3.1، 3.1، 3.2، 3.4، 3.5. ومن المرغوب أن تمتد هذه القيم من أحد طرفي المتصل إلى الطرف الآخر، بما يضمن تمثيل جميع درجات الاتجاه (Attitude Levels). غير أنه قد تُلاحظ تجمعات لفقرات متعددة عند نقاط متقاربة على المتصل؛ وفي هذه الحالة، يُفضل اختيار فقرة واحدة من بين تلك الفقرات المتقاربة، شرط أن تظهر أعلى درجة اتفاق بين المحكمين. فعلى سبيل المثال، إذا بلغت قيمة الوسيط لفقرتين 3.15 و3.19، وكان تشتت أحكام المحكمين حول الفقرة الأولى أقل من الثانية، فيُستحسن اختيار الأولى واستبعاد الثانية. وبالمثل، قد نجد الفقرتان اللتان وسيط كل منهما 4، 3 على التوالي. يقعان على المتصل ويبتعدان عن بعضهما البعض وحدة واحدة للفرق الذي يمكن ملاحظته (Just Noticeable Difference : JND)، مما يستدعي اتخاذ قرار حول مدى المتصل وعدد الفقرات التي سيتم الإبقاء عليها. وعادةً ما يكون اختيار 20 إلى 25 فقرة كافياً لتكوين مقياس متوازن، وإن كان بالإمكان تقليص العدد إذا اقتضت الضرورة. وفي حال توفر عدد كبير من الفقرات، يمكن بناء صورتين متكافئتين (Equivalent Forms) من المقياس، بحيث توزع أزواج الفقرات عشوائياً بين الصيغتين، على أن تتساوى كل فقرتين في قيمة الوسيط ونصف المدى الربيعي .كما يراعى التوزيع العشوائي للفقرات داخل كل صيغة على حدة. وتستخدم هذه الصيغ المتكافئة عادة في دراسة تغير الاتجاهات (Attitude Change) عبر الزمن أو المواقف المختلفة.

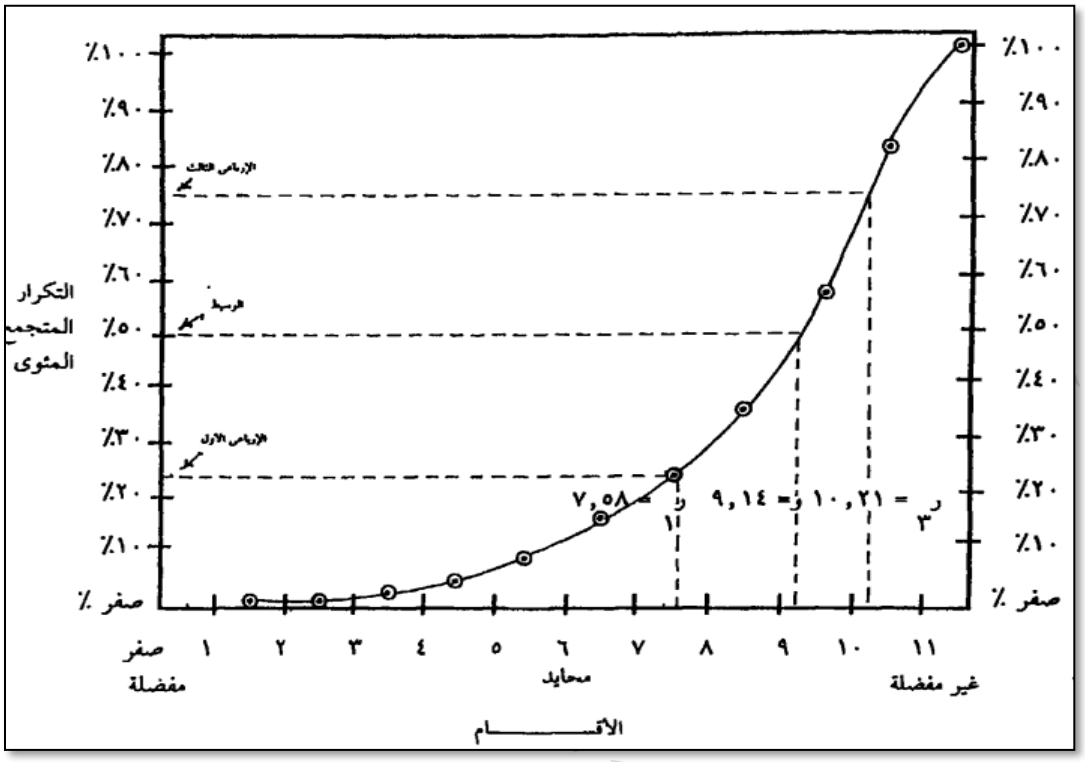
ولتوضيح ذلك عملياً يمكن بيان طريقة حساب كل من الوسيط ونصف المدى الربيعي لفقرة من فقرات مقياس الاتجاه نحو التعليم – ولتكن الفقرة الأولى مثلاً – وفق توزيع أحكام المحكمين الموضح في الجدول أدناه.

جدول رقم 01 يوضح توزيع افتراضي لبيانات التصنيف لفقرة ميزان الاتجاه نحو التعليم

رقم المجموعة أو القسم	عدد المحكمين	النسبة المئوية	النسبة المئوية المتجمعة
1	صفر	صفر	صفر
2	صفر	صفر	صفر
3	1	2	2
4	1	2	4
5	2	4	8
6	4	8	16
7	4	8	24
8	6	12	36
9	11	22	58
10	12	24	82
11	9	18	100
	ن = 50	%100	

ومن خلال الجدول السابق يتضح أن عدد المحكمين بلغ 50 محكماً، وقد قاموا بتصنيف الفقرة ضمن 11 قسماً تمثل درجات التفضيل على متصل الاتجاه. ويلاحظ أن الفقرة لم تدرج في القسمين (1) و(2) وفقاً للعمود الأول من الجدول، أي أن عدد المحكمين في هذين القسمين كان صفراً. تم حساب النسب المئوية لعدد المحكمين في كل قسم (العمود الثالث) بقسمة عدد المحكمين الذين اختاروا كل قسم على العدد الكلي للمحكمين (50)، ثم ضرب الناتج في 100. أما النسب المئوية المتجمعة (العمود الرابع)، فقد تم الحصول عليها من خلال الجمع التراكمي للنسب المئوية المقابلة لكل قسم في العمود الثالث. وباستخدام هذه البيانات، يمكن رسم المنحنى التراكمي الصاعد النسبي (Cumulative Frequency Curve)، بحيث تُوضع الأقسام على المحور الأفقي، وتمثل النسب المئوية المتجمعة على المحور الرأسي، وتحدد النقاط أعلى منتصفات الأقسام المناظرة لنسبها المئوية المتجمعة كما هو موضح في الشكل التالي:

الشكل 06 يوضح المنحنى التكراري المتجمع الصاعد لفقرة ميزان الاتجاه نحو التعليم



ويستدل من الشكل السابق على أن قيمة الوسيط (Median) لهذه الفقرة هي 9.14، في حين بلغ نصف المدى الربيعي:

$$\mathbf{1.32 = \frac{2.63}{2} = \frac{7.58 - 10.21}{2} = \frac{1_{ر} - 3_{ر}}{2} =}$$

وبذلك تكون قيمة مقياس الفقرة (Item Scale Value) مساوية لـ 9.14، بينما يمثل مقدار التشتت (Dispersion) قيمة 1.32.

هـ- يراعى في اختيار الفقرات أن تكون المسافات الفاصلة بينها متساوية تقريباً، وأن تكون قيمة نصف المدى الربيعي لكل فقرة منخفضة، بما يعكس اتساق أحكام المحكّمين حولها. ويمكن تحديد سعة الفترات بوحدة واحدة أو نصف وحدة أو أي قيمة أخرى مناسبة تبعاً لغرض القياس. فعلى سبيل المثال، إذا تقرر اختيار 20 فقرة فقد يتم انتقاء الفقرات التي تقترب قيمة وسيط كل منها من القيم التالية: 2.5، 3، 3.5، 4، 4.5، وهكذا، بحيث تُحقق فترات ميزان الاتجاه تساوياً ظاهرياً على المتصل. وفي بعض الحالات الخاصة التي يُراد فيها تمييز أدق في جزء معين من الميزان مقارنةً

بالأجزاء الأخرى يمكن عندئذ اختيار فقرات تتقارب قيمها في ذلك الجزء لتحقيق دقة أعلى في القياس.

و- بعد الانتهاء من اختيار الفقرات المناسبة ترتب هذه الفقرات ترتيباً عشوائياً وتدرج في قائمة نهائية دون الإشارة إلى قيمة ميزان كل فقرة (Scale Values). وعند تطبيق ميزان الاتجاه يُطلب من الأفراد الذين يجرى عليهم القياس إبداء مدى موافقتهم أو عدم موافقتهم على كل فقرة أو في بعض الحالات اختيار عدد محدد من الفقرات التي تعبر بأفضل شكل عن مواقفهم. يتم بعد ذلك التركيز فقط على الفقرات التي أبدى الفرد موافقته عليها ويُحسب متوسط قيم الميزان لتلك الفقرات (الوسيط)، حيث تُستخدم هذه القيمة لتحديد موقع الفرد على متصل الاتجاه الممتد من التفضيل إلى عدم التفضيل. ومن الناحية النظرية يُفترض أن يُبدي الفرد موافقته على عدد محدود من الفقرات المتجاورة في قيمها على متصل الميزان، بحيث تقترب تلك القيم من موقعه الحقيقي على المتصل. أما إذا لوحظ أن عدداً كبيراً من الأفراد يبدون موافقتهم على فقرات متباعدة في قيم الميزان فقد يشير ذلك إلى أن الميزان متعدد الأبعاد الأمر الذي يستدعي تقسيمه إلى موازين فرعية أحادية البعد لضمان دقته البنيوية.

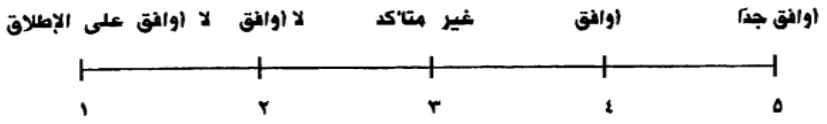
وعلى الرغم من أن هذا الأسلوب في بناء موازين الاتجاه يهدف أساساً إلى تحديد موقع كل فقرة على المتصل الافتراضي للميزان إلا أنه يجب توخي الحذر عند التعامل مع قيم الميزان بوصفها أرقاماً يمكن جمعها أو مقارنتها كما في المقاييس الكمية البحتة. فالدرجات هنا تمثل فترات متساوية ظاهرياً (Equal-Appearing Intervals)، أي أنها وحدات سيكولوجية قد تكون متساوية أو غير متساوية من حيث القيمة العددية الفعلية.

وبناءً على ذلك فإن تقسيم المتصل إلى 11 قسماً يعد تقسيماً اصطلاحياً لا يعكس تساوياً عددياً حقيقياً، ومن ثم لا يصح افتراض أن الدرجة (8) تساوي ضعف الدرجة (4) في معناها النفسي. كما لا يمكن مقارنة الدرجات بين موازين اتجاه مختلفة ضمن بطارية واحدة من المقاييس.

3-2- أسلوب التقدير الجمعي (ليكرت):

يعرف الأسلوب الثاني في بناء موازين الاتجاهات باسم أسلوب التقدير الجمعي (Summated Ratings Method) أو أسلوب ليكرت (Likert Technique). ويعد هذا الأسلوب أبسط وأقل تطلباً للوقت والجهد مقارنةً بأسلوب ثيرستون (Thurstone Method)، مع أنه يؤدي إلى نتائج متقاربة من حيث دقة القياس. ويتميز أيضاً بأنه لا يعتمد على تقييم المحكّمين مما جعله أكثر الأساليب شيوعاً في مجالات القياس والبحوث النفسية والتربوية والاجتماعية.

يرتكز هذا الأسلوب على القياس الرتبي للاتجاهات، حيث تعرض على الفرد قائمة من العبارات أو الفقرات تتعلق بموضوع معين ويُطلب منه إبداء درجة موافقته أو عدم موافقته على كل فقرة وفق درجات متدرجة تعكس شدة اتجاهه. ويفترض أن الأفراد يختلفون في مواقعهم على متصل الاتجاه، بحيث يُتوقع أن يُظهر ذوو الموافقة القوية أو الضعيفة سلوكاً مميزاً في المواقف الاجتماعية ذات الصلة بموضوع الاتجاه. ومع ذلك فإن العلاقة بين الاتجاه والسلوك الفعلي لا تزال موضع نقاش بين علماء النفس. وتُقاس شدة الاتجاه جزئياً من خلال إعطاء أوزان رقمية مختلفة لاستجابات الفرد على كل فقرة. فعلى سبيل المثال، إذا كانت الفقرة هي: "أحب مهنتي". فقد يطلب من الفرد أن يختار استجابته من ميزان رتبي خماسي النقاط على النحو الآتي:



تعطى للاستجابة "أوافق جداً" قيمة رقمية مقدارها (5)، بينما تُعطى للاستجابة "لا أوافق على الإطلاق" القيمة (1)، وهكذا لبقية درجات الاستجابة. وكلما ارتفعت القيمة الرقمية دلّ ذلك على زيادة درجة حب الفرد لمهنته والعكس صحيح. فعلى سبيل المثال، إذا أجاب 100 فرد عن هذه الفقرة فسوف نحصل على 100 استجابة تتراوح أوزانها الرقمية بين (1) و (5). فالأفراد الذين اختاروا "أوافق جداً" يحصلون على الوزن (5)، أما الذين اختاروا "لا أوافق على الإطلاق" فيمنحون الوزن (1). ومع ذلك لا يمكن اعتبار جميع من أبدوا موافقتهم التامة على الفقرة بأن لديهم الدرجة نفسها من المشاعر الإيجابية تجاه مهنتهم؛ لأن فقرة واحدة لا تكفي لتقييم الاتجاه بدقة. ولهذا السبب يعتمد أسلوب التقدير الجمعي (ليكرت) على مجموعة من الفقرات لقياس الاتجاه مما يتيح تمييزاً أدق بين الأفراد من حيث شدة اتجاههم نحو الموضوع المدروس وذلك ضمن ميزان متدرج كما في المثال السابق. فإذا كان الميزان يتكون من خمس درجات واحتوى المقياس على عشر فقرات فإن أعلى درجة ممكنة للفرد ستكون (50) بينما تكون أدنى درجة (10)، أي أن المدى الكلي يصبح (40) بدلاً من (5) للفقرة الواحدة. وكلما زاد عدد الفقرات في الميزان اتسع مدى الدرجات وزاد تباينها مما يؤدي إلى تحسين دقة القياس.

ومن الجدير بالذكر أن هذا الأسلوب لا يقتصر على المقياس الخماسي السابق، بل يمكن استخدام أي نوع من المقاييس الرتبية المتدرجة مثل:

أ- موافق، غير متأكد، غير موافق.

ب- موافق بشدة، موافق، غير متأكد ولكن موافق إلى حد ما، غير متأكد ولكن غير موافق إلى حد ما، غير موافق، غير موافق على الإطلاق.

ج- دائماً، غالباً، أحياناً، نادراً، إطلاقاً.

يفضل ألا يتجاوز عدد أقسام ميزان الاستجابة خمس فئات، حتى يتمكن الفرد من التمييز بينها بدقة واختيار الدرجة التي تعبر بصدق عن مستوى موافقته. ويُلاحظ أن النمطين (1) و (2) من الموازين قد تضمّنا الفئة "غير متأكد" والتي تُستخدم للإشارة إلى أن الفرد لا يمتلك رأياً محدداً تجاه الفقرة ومع ذلك يُسمح أحياناً بإدراجها في بعض موازين الاتجاهات لأغراض معينة. أما النمط الثاني فقد تضمن ست فئات بدلاً من خمس وذلك بهدف إجبار المستجيب على تحديد موقفه بدقة أكبر بحيث يتجه نحو الجانب الإيجابي (الموافقة) إذا اختار "غير متأكد ولكن أميل إلى الموافقة" أو نحو الجانب السلبي (الرفض) إذا اختار "غير متأكد ولكن أميل إلى عدم الموافقة". ويُتبع هذا الإجراء لأن الاستجابة "غير متأكد" البحتة لا تُقدّم فائدة تحليلية كافية عند معالجة البيانات إحصائياً.

ومع ذلك فإن استخدام فئة "غير متأكد " في موازين الاتجاهات يظل مرتبطاً بطبيعة عينة المستجيبين، ومستوى تعليمهم وفهمهم للمواقف المطروحة وكذلك بدرجة الدقة المطلوبة في القياس (علام، 2000، ص539).

3-2-1- خطوات بناء موازين الاتجاهات باستخدام أسلوب ليكرت:

يمكن بناء موازين الاتجاهات باستخدام أسلوب التقدير الجمعي (أسلوب ليكرت) باتباع مجموعة من الخطوات المنهجية كما يلي:

أ- يبدأ بـتجميع عدد كبير من الفقرات التي تتناول موضوع الاتجاه المراد قياسه، مثل: الاتجاه نحو التعليم أو مجال دراسي معين أو موضوعات عامة كـالانفتاح الاقتصادي أو تقنيات الاتصال وغيرها.

ونظراً لأن كل فقرة في ميزان ليكرت تهدف إلى الكشف عن تباين ثابت ومتسق في الاتجاه، فإنه ينبغي أن تتضمن الفقرات مدى واسعاً من شدة التفضيل (Favorableness)، بحيث تعبر عن مواقف متفاوتة من الإيجابية إلى السلبية. كما يجب مراعاة أن الفقرات المحايدة تماماً ليست ذات فائدة حقيقية في مثل: هذه المقاييس في حين أن الفقرات المتطرفة جداً في كلا الاتجاهين (الإيجابي والسلبي) تميل إلى إحداث تباين أقل مقارنةً بالفقرات المعتدلة مما يضعف قدرتها التمييزية بين الأفراد.

وبناءً على ذلك يستحسن أن يتراوح عدد الفقرات الأولي بين 40 و60 فقرة، بحيث تقسم تقسيماً متعادلاً بين الفقرات الموجبة (المعبّرة عن مواقف مؤيدة) والفقرات السالبة (المعبرة عن مواقف معارضة).

ب- يتم تطبيق الفقرات على عينة من الأفراد تكون مماثلة للمجموعة المستهدفة التي سيستخدم فيها ميزان الاتجاه في شكله النهائي لاحقاً. ويمكن أن تضم هذه العينة عددًا من الأفراد يعادل عشرة أضعاف عدد الفقرات حسب (Nunnally, 1978). ويفضل إن أمكن زيادة حجم العينة لتحقيق تمثيل أفضل. وينبغي الحرص على أن تكون ظروف تطبيق المقياس مماثلة قدر الإمكان لظروف التطبيق الفعلية على المجموعة المستهدفة. ويطلب من كل فرد في العينة أن يُعبّر عن رأيه تجاه كل فقرة ويحدد شدة موافقته باستخدام خيارات مثل: موافق جداً، موافق، غير متأكد، غير موافق، أو غير موافق على الإطلاق. ولهذا من المهم أن تكون العينة المختارة تمثل النطاق الكامل للاتجاه المراد قياسه قدر الإمكان لضمان شمولية النتائج ودقتها.

ج- يخصص لكل استجابة من الفرد على كل فقرة درجة أو وزن رقمي بناءً على الأوزان التي تم الإشارة إليها، مع ضرورة عكس القيم الرقمية للفقرات السلبية. مثلاً إذا كانت الفقرة إيجابية مثل: " تعلّمني المدرسة ما يساعدني في الحصول على عمل"، فتُعطى الاستجابة "موافق جداً" الوزن 5، و"موافق" 4، و"غير متأكد" 3، و"غير موافق" 2، و"غير موافق على الإطلاق" 1. أما إذا كانت الفقرة تعبر عن موقف سلبي مثل: "المدرسة مضيعة للوقت"، فتُعكس الأوزان بحيث يصبح "موافق جداً" 1، و"موافق" 2، و"غير متأكد" 3، و"غير موافق" 4، و"غير موافق على الإطلاق" 5. بعد ذلك تُجمع درجات الفرد في جميع الفقرات التي أجاب عنها. على سبيل المثال، إذا أجاب الفرد بـ "موافق جداً" للفقرة الأولى بدرجة 5، و"غير متأكد" للفقرة الثانية بدرجة 3، و"موافق" للفقرة الثالثة بدرجة 4، و"غير موافق" للفقرة الرابعة بدرجة 2، يكون المجموع الكلي لدرجته 14 (5 + 3 + 4 + 2). وبالمثل تحسب الدرجات للفقرة الثانية مع مراعاة عكس الأوزان وفقاً لطبيعة الفقرة.

د- يتم إجراء تحليل للفقرات (Item Analysis) بهدف اختيار الفقرات التي تميز الأفراد بدرجة أفضل على متصل الاتجاه، وذلك من خلال حساب الارتباط بين الدرجات الكلية التي يحققها الأفراد واستجاباتهم لكل فقرة على حدة، حيث تُستبقى الفقرات ذات الارتباط العالي لضمان الحصول على ميزان يتمتع باتساق داخلي جيد. كما يمكن أيضاً حساب معامل تمييز كل فقرة. ويفضل ترتيب قيم

معاملات الارتباط ترتيباً تنازلياً، ثم اختيار عدد متقارب من الفقرات الموجبة والفقرات السالبة ودمجها معاً لتشكيل ميزان الاتجاه في صورته الأولية.

هـ- يتم حساب معامل ألفا لكرونباخ (α) للفقرات المختارة من خلال تحديد تباين درجات كل فقرة بالإضافة إلى تباين الدرجات الكلية وتطبيق الصيغة المعروفة، أو عبر تكوين مصفوفة الارتباطات بين درجات الفقرات والدرجات الكلية وإجراء تحليل المفردات باستخدام الحاسوب. تستخدم هذه المصفوفة أيضاً في الخطوة التالية. إذا كانت قيمة معامل ألفا مرتفعة بما فيه الكفاية (0.80 أو أكثر)، فيُسمح بإدراج هذه الفقرات ضمن الصيغة النهائية لميزان الاتجاه والتي عادةً ما تتكون من حوالي 20 فقرة. وتجدر الإشارة إلى أن قيمة معامل الثبات تعتمد على عدد أقسام ميزان التقدير المستخدم لكل فقرة. تُعتبر هذه الخطوة مؤشراً هاماً على خاصية أحادية البعد (Unidimensionality)، حيث تساعد في التأكد من أن جميع الفقرات تقيس نفس البُعد أو الخاصية وهو ما كان محور اهتمام ليكرت بشكل أساسي.

و- ترتب الفقرات المختارة بشكل عشوائي داخل ميزان الاتجاه، أي في الاستبيان، ليتم استخدامه في قياس الاتجاه المستهدف عند تطبيقه على المجموعة المراد قياسها. بعد ذلك، تُحسب الدرجة الكلية لكل فرد بجمع الدرجات المقابلة لاستجاباته على الفقرات المدرجة في الاستبيان. تجدر الإشارة إلى أن الدرجة الخام الكلية التي يحصل عليها الفرد تكون ذات معنى محدود بمفردها، إذ يتضح معناها بشكل أفضل عند مقارنتها بدرجات الأفراد الآخرين الذين تم تطبيق ميزان الاتجاه عليهم، مما يتيح تحديد موقع الفرد على متصل الاتجاه بالنسبة لباقي الأفراد (علام، 2000، ص541).

3-3- الأسلوب التراكمـي (جتمـان):

تشتمل موازين الاتجاهات المبنية باستخدام الأسلوبين السابقين على فقرات قد تكون غير متجانسة من حيث الأبعاد المختلفة لموضوع الاتجاه المراد قياسه. فمثلاً عند بناء ميزان اتجاه يتعلق بالتعليم قد لا يتم التفريق بين الفقرات التي تعكس الممارسات السائدة في التعليم الإلزامي وتلك التي تتناول فكرة التعليم الذاتي غير الرسمي أو جوانب أخرى متنوعة من الاتجاه نحو التعليم. يؤدي الجمع بين هذه الأبعاد المختلفة في ميزان واحد إلى صعوبة في تفسير الدرجات التي يحصل عليها الأفراد بشكل واضح خاصة مع تعدد أنماط هذه الدرجات كما أوضحنا سابقاً.

ولمعالجة هذه المشكلة اقترح جتمان (Guttman) أسلوباً يعرف بالأسلوب التراكمي (Cumulative Technique)، والذي يمكن من خلاله بناء موازين اتجاهات أحادية البعد (Unidimensional). يُعتبر

ميزان الاتجاه أحادي البعد فقط إذا كان تراكميًا أي أن الفقرات تكون مرتبطة بحيث إذا وافق الفرد على الفقرة رقم (2)، فإنه يوافق بالضرورة على الفقرة (1) التي تسبقها، ومن يوافق على الفقرة (3) فإنه يوافق أيضاً على الفقرتين (1) و(2)، وهكذا دواليك. كمثال إذا طلب من الأفراد إبداء موافقتهم أو عدم موافقتهم على العبارات التالية المتعلقة بالاتجاه نحو منظمة الأمم المتحدة:

(أ) منظمة الأمم المتحدة تخدم جميع الشعوب.

(ب) منظمة الأمم المتحدة هي أملنا في التعايش السلمي.

(ج) منظمة الأمم المتحدة تمد قوة بناءة في العالم.

(د) ينبغي استمرار المشاركة في منظمة الأمم المتحدة.

فإذا كانت هذه الفقرات تشكل ميزاناً تراكمياً فإننا نتوقع أن الفرد الذي يوافق على الفقرة (أ) سيوافق أيضاً على الفقرات (ب)، (ج)، و(د). أما من لا يوافق على الفقرة (أ) لكنه يوافق على الفقرة (ب)، فمن المتوقع أن يوافق على الفقرتين (ج) و(د). بهذا يمكن ترتيب استجابات الأفراد في نمط متسلسل يشبه المثال الموضح في جدول التالي، حيث تُعطى درجة واحدة لكل استجابة "موافق" على أي من الفقرات الأربع.

جدول 02 يوضح نمط الاستجابات على أربع فقرات في ميزان تراكمي

الاستجابة «موافق» على الفقرة				الاستجابة «غير موافق» على الفقرة			
الدرجة	د	جـ	ب	أ	د	جـ	ب	أ
٤	×	×	×	×				
٣		×	×	×	×			
٢			×	×	×	×		
١				×	×	×	×	
صفر					×	×	×	×

3-3-1- خطوات بناء موازين الاتجاهات باستخدام أسلوب جتمان:

عند بناء ميزان اتجاه تراكمي يجب كتابة أو تجميع مجموعة من الفقرات التي تتطلب من المستجيب الاختيار بين "موافق" أو "غير موافق"، ثم ترتيب هذه الفقرات حسب درجة الموافقة. وفي بعض الأحيان يمكن ترتيب الفقرات بشكل عشوائي دون محاولة التأكد من تساوي الفواصل

بينها لأن الميزان التراكمي يُصنف ضمن المستويات الرتبية. ومع ذلك من الضروري التأكد من أن جميع الفقرات تمثل بعداً أحادياً للاتجاه المراد قياسه بحيث تتبع نمطاً مثالياً يشبه ما هو موضح في الجدول السابق. إذا تحقق هذا النمط المثالي فإن مجموعة الفقرات المختارة تحمل معنى واحداً؛ بمعنى أنه إذا عرفنا الدرجة التي حصل عليها فرد يمكننا معرفة الفقرات التي أبدى موافقته عليها دون الحاجة للرجوع إلى الاستبيان. غير أن الحصول على هذا النمط المثالي التام في الواقع العملي غالباً ما يكون صعباً بسبب وجود عدم اتساق في استجابات الأفراد، أي وقوع أخطاء. ويمكن توضيح ذلك من خلال نمط الاستجابات المعروض في الشكل التالي:

شكل 07 يوضح الخطأ في بعض أنماط الاستجابات

الفقرات						
	أكثر شدة			أقل شدة		
الفرد	أ	ب	جـ	د	هـ	الدرجة
١	+	+	−	+	+	٤
٢	+	−	+	+	+	٤
الخطأ	صفر	١	١	صفر	صفر	

يتضح من شكل السابق أن الفرد الأول وافق على الفقرتين (أ) و(ب)، ولم يوافق على الفقرة (ج)، لكنه أبدى الموافقة على الفقرتين (د) و(ه)، وحصل بذلك على الدرجة (4). عدم موافقته على الفقرة (ج) هنا يُعتبر خطأ(Error)، وكذلك عدم موافقة الفرد الثاني على الفقرة (ب) يعد خطأً أيضاً، ولذلك تم وضع دائرة على الفقرتين (ج) و(ب) لكل من الفردين على الترتيب. هذا يشير إلى احتمال أن ميزان الاتجاه قد لا يكون أحادي البعد. في المقابل نمط الاستجابات المعروض في جدول السابق يمثل ميزان اتجاه تراكمي أحادي البعد، حيث يمكن التنبؤ بنمط الدرجات بناءً على الدرجة الكلية للفرد، إذ لا يحتوي نمط الاستجابات على أخطاء (علامات سالبة)، مما يعني إمكانية إعادة تكوين هذا النمط (Reproducible). أما إذا كان عدد الأخطاء في نمط استجابات مجموعة من الأفراد كبيراً، فهذا يشير إلى عدم إمكانية إعادة تكوين النمط (Nonreproducible).

وقد اقترح جتمان (Guttman) صيغة لحساب معامل إعادة تكوين نمط الاستجابات (Coefficient of Reproducibility)، كالتالي:

- معامل إعادة تكوين نمط الاستجابات يُحسب وفق الصيغة التالية:

$$\text{معامل إعادة تكوين نمط الاستجابات} = 1 - \frac{\text{عدد الأخطاء}}{\text{عدد الاستجابات}}$$

حيث يقصد بالأخطاء الابتعاد عن النمط المثالي ويحسب عدد الاستجابات بضرب عدد الأفراد في عدد الفقرات. بالنظر إلى شكل السابق، نجد أن عدد الأخطاء يساوي 2 وعدد الأفراد 2 وعدد الفقرات 6 وباستخدام الصيغة نحصل على:

$$\text{معامل إعادة تكوين نمط الاستجابات} = 1 - \frac{2}{6 \times 2} = 0.50$$

وقد اقترح جتمان أن يكون الحد الأدنى لقيمة هذا المعامل لمجموعة الفقرات التي يتضمنها الميزان التراكمي للاتجاه هو 0.90، ليعتبر الميزان أحادي البعد. ونظراً لاعتماد المثال السابق على فقرتين فقط، انخفضت قيمة المعامل إلى 0.50. لذلك أوصى جتمان باختيار 10 إلى 12 فقرة وتطبيقها على عدد لا يقل عن 110 أفراد مع اختيار هذه الفقرات من نطاق واسع وشامل يخص موضوع الاتجاه المراد قياسه (علام، 2000، ص546).

رابعاً: بعض المشكلات المتعلقة بقياس الجوانب غير المعرفية:

يعد قياس الجوانب غير المعرفية أو الوجدانية للسلوك أكثر تعقيداً وصعوبة من قياس الجوانب المعرفية وذلك لعدة أسباب، من أبرزها ما يلي:

1- غموض المفاهيم وصعوبة الملاحظة المباشرة: لا يوجد اتفاق كامل حول تعريف مكونات الجوانب الوجدانية للسلوك، كما أن هذه المكونات يصعب ملاحظتها مباشرة في أداء الأفراد لأنها تحدث داخل الفرد نفسه. وغالباً ما يواجه المعلم أو المرشد التربوي أو المهني صعوبة في تحديد المؤشرات السلوكية التي يمكن أن تعكس ميلاً أو اتجاهاً معيناً لدى الفرد. لذا يتطلب قياس هذه الجوانب غير المرئية تحليلاً سلوكياً دقيقاً يساعد على ربط السلوك الظاهر بالمكونات الوجدانية الداخلية، مع الإشارة إلى أن الآراء تختلف اختلافاً واضحاً حول إجراءات هذا التحليل ونتائجه.

2- تأثر السلوك الوجداني بعوامل ظرفية ومتغيرة: يحتاج تقييم الجوانب الوجدانية إلى ملاحظة الفرد في مواقف متعددة لأن ملاحظته في موقف واحد لا تكفي للحكم على سلوكه أو أدائه المميز.

فالسلوك الوجداني يتأثر بحالة الفرد المزاجية ومدى ألفته بالموقف والعوامل البيئية المحيطة به. فقد يبدو الشخص عدوانياً في موقف معين نتيجة شعوره بالقلق أو الضيق إلا أنه قد يظهر سلوكاً متكيفاً بعد فترة من الزمن مما يعني أن بعض السلوكيات قد تكون مؤقتة وليست سمة ثابتة في شخصيته.

3- تعدد المكونات وصعوبة قياسها بأداة واحدة: نظراً لتداخل وتعقيد المفاهيم الوجدانية وصعوبة تحديدها بدقة فإنه من غير الممكن الاعتماد على أداة واحدة لقياسها جميعاً. لذا يُفضَّل استخدام مجموعة من المقاييس المتنوعة التي تساهم في تكوين صورة أكثر شمولاً ودقة عن المكون الوجداني المراد قياسه.

4- مشكلات التحقق من الصدق والثبات: يتضمن قياس الجوانب غير المعرفية دراسة الفروق الفردية وكذلك التغيرات التي تطرأ على السلوك عبر الزمن، مما يفرض تحديات إضافية على القائمين ببناء أدوات القياس، خصوصاً فيما يتعلق بالتحقق من صدق المقاييس وثباتها في مختلف الظروف والمواقف (علام، 2000، ص478).

المحـــاضرة السادسة

بنـــاء وتصميم الاختبـــارات النفسيـــة والتربويـــة

الأهداف:

تهدف هذه المحاضرة إلى تعريف الطلبة بالقواعد الإرشادية الأساسية في تطوير المقاييس النفسية، من خلال توضيح الخطوات المنهجية التي تبدأ بتحديد ما يراد قياسه بدقة، ثم توليد بنود مناسبة تعكس هذا البناء، واختيار الشكل أو التنسيق الملائم للمقياس. كما تسعى إلى إبراز أهمية مراجعة البنود من قبل الخبراء، والنظر في إدراج بنود داعمة للصدق، ثم تطبيق المقياس على عينة مناسبة لتطويره. وتهدف كذلك إلى تنمية مهارات الطلبة في تقييم البنود إحصائيًا ومنطقيًا، وصولًا إلى تحقيق الطول الأمثل للمقياس بما يضمن كفاءته ودقته في القياس.

محتوى المحاضرة:

أولاً: قواعد إرشادية في تطوير المقياس

الخطوة الأولى: تحديد ما يراد قياسه بدقة

الخطوة الثانية: توليد مجموعة من البنود

الخطوة الثالثة: تحديد شكل أو تنسيق القياس

الخطوة الرابعة: مراجعة تجمع البنود الأولية من قبل الخبراء

الخطوة الخامسة: النظر في إدراج البنود الداعمة للصدق

الخطوة السادسة: تطبيق البنود على عينة تطوير المقياس

الخطوة السابعة: تقييم البنود

الخطوة الثامنة: تحقيق الحد الأمثل لطول المقياس

مدة المحاضرة: ثلاث ساعات

تمهيد:

يعد تطوير المقاييس النفسية من العمليات الأساسية في مجال القياس والتقويم النفسي والتربوي، حيث يعتمد عليه الباحثون والممارسون في الحصول على بيانات دقيقة وموثوقة حول السمات والخصائص النفسية. وتتطلب هذه العملية اتباع خطوات منهجية منظمة تضمن بناء أدوات قياس تتمتع بدرجات عالية من الصدق والثبات، بدءًا من تحديد البناء المراد قياسه بدقة، مرورًا بصياغة البنود واختيار شكلها المناسب، وانتهاءً بتجريبها وتحليلها وتحسينها. كما أن مراعاة الجوانب العلمية والإجرائية في كل مرحلة من مراحل تطوير المقياس تسهم في رفع جودة الأداة وكفاءتها، مما يعزز من دقة النتائج وإمكانية الاعتماد عليها في اتخاذ القرارات التربوية والنفسية.

أولاً: قواعد إرشادية في تطوير المقياس:

يتم الانتقال إلى التطبيق العملي لتطوير المقاييس من خلال مجموعة من الإرشادات التي تساعد الباحثين في بناء المقاييس وتطويرها.

الخطوة الأولى: تحديد ما يراد قياسه بدقة:

قد يعتقد الباحث أنه يمتلك تصوراً واضحاً للبناء المراد قياسه، إلا أن هذا الوضوح يكون أحياناً مضللاً، ولا يتكشف الغموض إلا بعد مراحل متقدمة من إعداد البنود وجمع البيانات، حيث تصبح التعديلات أكثر كلفة. لذلك ينبغي منذ البداية الإجابة عن تساؤلات أساسية، مثل:

- هل يستند المقياس إلى نظرية قائمة أم إلى توجه فكري جديد؟
- ما مستوى الدقة المطلوب في القياس؟
- هل سيتم التركيز على جميع أبعاد الظاهرة أم على جوانب محددة منها؟

1- دور النظرية في تحقيق الوضوح:

وضوح محتوى المقياس يرتبط بوضوح المفهوم البنائي المراد قياسه. ورغم أهمية الجوانب التقنية في تطوير المقاييس والتحقق من صدقها، فإن الأساس النظري يظل ضرورياً، خاصة عند قياس ظواهر غير مباشرة لا تمتلك معياراً مرجعياً واضحاً. وتسهم النظرية في تحديد حدود الظاهرة ومنع انحراف المقياس نحو مجالات غير مقصودة. لذلك ينبغي الرجوع إلى نظريات العلوم الاجتماعية ذات الصلة قبل الشروع في بناء المقياس. وإذا لم توفر النظريات القائمة إرشاداً كافياً، يمكن تبني توجه جديد، شريطة أن يكون ذلك مبنياً على مراجعة علمية واعية. وفي حال غياب إطار نظري جاهز، يتعين على الباحث صياغة نموذج مفاهيمي أولي يوجه عملية القياس، ولو كان بسيطاً،

كتعريف دقيق للظاهرة، مع توضيح علاقتها بالبناءات الأخرى وطرق قياسها إجرائياً(ديفيليس، 2021، ص147-207).

2- الخصوصية كوسيلة لتحقيق الوضوح:

يعد مستوى التحديد (الخصوصية مقابل العمومية) في قياس المفهوم البنائي عاملاً حاسماً، إذ تشير الأدبيات إلى أن قوة العلاقة بين المتغيرات تزداد عندما تتقارب في مستوى التحديد (Ajzen & Fishbein, 1980). وبناءً على هدف البحث، قد يكون المقياس موجهاً لقياس سلوكيات محددة جداً أو مفاهيم عامة وشاملة.

- **مثال:** مصدر الضبط (Locus of Control)، يوضح هذا المفهوم كيف يمكن أن يختلف القياس من حيث الخصوصية:

- قدم Julian B. Rotter مقياساً عاماً نسبياً لمصدرالضبط الداخلي–الخارجي، حيث يعتمد على بعد واحد وتكون النتائج والمصادر الخارجية فيه عامة (Rotter, 1966).

- طورت Hanna Levenson نموذجاً متعدد الأبعاد يميز بين ثلاثة مصادر للضبط: الذات، والآخرون الأقوياء، والمصير، مما يزيد من دقة تحديد مصادر التأثير مع بقاء النتائج عامة (Levenson, 1973).

- قام Kenneth A. Wallston وزملاؤه بتطوير مقاييس مصدر الضبط الصحي متعددة الأبعاد (MHLC) التي تربط التحكم بنتائج صحية محددة، كالوقاية من المرض (Wallston et al., 1978).

- ثم تم تطوير نسخة أكثر تحديداً (MHLC Form C) تسمح بتكييف البنود وفق حالة صحية معينة، مثل السكري (Wallston, Stein, & Smith, 1994).

3- أهمية التوافق بين مستوى التحديد وسؤال البحث:

تتوقف فاعلية المقياس على مدى توافق مستوى خصوصيته مع هدف الدراسة؛ فالمقاييس العامة (Rotter, 1966). تناسب التنبؤ بالسلوكيات العامة، بينما المقاييس الأكثر تحديداً مثل Wallston. et al, 1994. تكون أنسب لدراسة سلوكيات أو نتائج دقيقة كالصحة. لذا يجب أن يكون قرار مستوى التحديد مقصوداً ومبنياً على الإطار النظري ووظيفة المقياس، وليس نتيجة عرضية بعد صياغة البنود.

4- أبعاد أخرى للخصوصية: لا يقتصر التحديد على النتائج أو مصادر التحكم، بل يشمل أيضاً:

- مجال المحتوى، مثل (القلق مقابل التكيف العام).
- السياق أو البيئة، مثل (بيئات العمل المختلفة).
- المجتمع الإحصائي المستهدف، مثل (الأطفال مقابل البالغين).

وبذلك، فإن تحديد مستوى الخصوصية يعد خطوة أساسية لضمان دقة القياس وملاءمته للسؤال البحثي.

5- توضيح ما ينبغي تضمينه في القياس: يتعين على مطوري المقاييس التأكد من أن المفهوم البنائي المراد قياسه متميز عن غيره من المفاهيم، وأن حدودَه محددة بوضوح. فكما يمكن أن تكون المقاييس عامة أو خاصة من حيث الحالات التي تغطيها، يمكن أيضاً أن تختلف من حيث المفاهيم البنائية التي تتضمنها. فعلى سبيل المثال، يعد قياس القلق العام مشروعاً إذا كان الهدف يشمل أنواعه المختلفة، كقلق الاختبار والقلق الاجتماعي. أما إذا كان الهدف يقتصر على نوع محدد، فيجب استبعاد الأبعاد الأخرى، لأن تداخل البنود مع مفاهيم قريبة قد يؤدي إلى تشويه القياس.

6- مشكلة تداخل المفاهيم البنائية: قد تتضمن بعض البنود مؤشرات تبدو متشابهة لكنها تعكس مفاهيم مختلفة، مما يجعل المقياس حساساً لأكثر من بناء واحد. يظهر ذلك في بعض مقاييس الاكتئاب، كمقياس مركز الدراسات الوبائية (Radloff, 1974)، الذي يتضمن بنوداً ذات طابع جسدي (كضعف النشاط أو الحركة). في بعض الحالات الصحية، كالتهاب المفاصل، قد تؤدي هذه البنود إلى التباس في التفسير، إذ قد تعزى الأعراض الجسدية إلى المرض لا إلى الاكتئاب (Blalock, DeVellis, Brown, & Wallston, 1989).

7- أهمية ملاءمة محتوى المقياس للسياق: بناءً على ذلك، قد يختار الباحث استبعاد الجوانب الجسدية عند تصميم مقياس للاكتئاب إذا كان سيستخدم مع مرضى يعانون من أمراض مزمنة، أو عند دمجه مع مقاييس أخرى تتناول أعراضاً جسدية. ومع ذلك، قد يكون إدراج هذه البنود ضرورياً في سياقات أخرى، خاصة عندما يكون الهدف دراسة الجوانب الجسدية للاضطرابات الانفعالية.

- خلاصة: إن تحديد ما يجب تضمينه في المقياس يتطلب قراراً واعياً يوازن بين نقاء المفهوم البنائي من جهة، وملاءمته لأهداف البحث والسياق التطبيقي من جهة أخرى، مع تجنب التداخل غير المرغوب بين المفاهيم.

الخطوة الثانية: توليد مجموعة من البنود:

1- بناء البنود في ضوء هدف المقياس: بعد تحديد الغرض من المقياس بدقة، يبدأ الباحث في تطوير الأداة عبر توليد عدد كبير من البنود المرشحة للاختيار النهائي. وتعد هذه المرحلة أساسية لضمان جودة المقياس.

2- اختيار البنود بما يعكس المفهوم البنائي: يجب أن تصاغ البنود وفق الهدف المحدد للقياس، بحيث يعكس كل بند المتغير الكامن المراد قياسه. فكل بند يمثل مؤشراً جزئياً لهذا المتغير، وكلما كانت البنود متجانسة ومعبرة عنه بدقة، زادت موثوقية المقياس. ورغم أن تعدد البنود يعزز الثبات، إلا أن كل بند ينبغي أن يظل حساساً للفروق الحقيقية في المتغير الكامن.

3- الاعتبار النظري لمجال البنود: من الناحية المثالية، يتم اختيار البنود عشوائياً من "مجال شامل " (universe of items) يعبر عن جميع تمثلات المفهوم البنائي. إلا أن هذا المجال يكون نظرياً وغير محدود عملياً، مما يجعل الوصول إليه كاملاً أمراً غير ممكن. ومع ذلك، ينبغي استحضار هذا التصور عند بناء البنود.

4- أهمية التفكير الإبداعي ضمن حدود المفهوم: عند صياغة البنود، يجب على الباحث أن يفكر بمرونة وإبداع لإنتاج صيغ متنوعة تعكس المفهوم البنائي، مع الالتزام بحدوده وعدم الخروج عنها. فخصائص المقياس تتحدد مباشرة بجودة بنوده؛ وإذا كانت البنود ضعيفة في تمثيل المفهوم، فإن المقياس ككل سيفشل في قياسه بدقة.

لهذا تعتمد جودة المقياس أساساً على جودة البنود، مما يتطلب توازناً بين الشمول، والدقة، والالتزام النظري بالمفهوم البنائي(ديفيليس، 2021، ص147-207).

5- التمييز بين المفهوم البنائي والفئة: من الضروري أن تعكس البنود مفهوماً بنائياً حقيقياً (متغيراً كامناً) لا مجرد فئة عامة. فالنماذج النظرية تفترض أن البنود هي مظاهر لمتغير كامن مشترك يفسّر الاستجابات عليها، وليس مجرد انتمائها لفئة واحدة. فعلى سبيل المثال، مصطلحات مثل "الاتجاهات" أو "عوائق الامتثال" تمثل فئات واسعة تضم عدة مفاهيم بنائية مستقلة، وليست مفهوماً واحداً. لذا ينبغي أن تركز البنود على بناء محدد، مثل (الاتجاه نحو معاقبة متعاطي المخدرات)، لأن كل بعد (كالخوف من التكلفة أو الألم أو بعد المسافة) قد يمثل متغيراً كامناً مختلفاً. ومن ثم، لا يتوقع أن تعمل البنود التي تنتمي لنفس الفئة العامة بالطريقة نفسها إذا كانت تقيس مفاهيم بنائية مختلفة.

6- التكرار (Redundancy) بين الفائدة والمشكلة: يعد التكرار خاصية مزدوجة في بناء المقاييس؛ فهو مفيد من جهة، وقد يكون ضاراً من جهة أخرى. من الناحية الإيجابية، يسهم التكرار في تعزيز ثبات المقياس، حيث إن استخدام عدة بنود متشابهة في قياس نفس المفهوم يسمح باستخلاص المحتوى المشترك وتقليل التأثيرات غير المرتبطة. وهذا يتماشى مع الأسس النظرية كمعادلة سبيرمان-براون، التي تشير إلى أن زيادة عدد البنود تحسن الثبات، لكن ليس كل تكرار مرغوباً؛ فالتكرار المفيد هو

الذي يرتبط بجوهر المفهوم البنائي، لا بالجوانب الشكلية أو اللغوية. فإعادة صياغة بند بنفس التركيب تقريباً لا تضيف قيمة حقيقية، بينما التعبير عن الفكرة نفسها بصيغ مختلفة يثري القياس ويزيد دقته. لهذا ينبغي أن تبنى البنود حول متغير كامن محدد بوضوح، مع استخدام تكرار وظيفي يخدم قياس هذا المتغير، وتجنب التكرار الشكلي الذي لا يضيف قيمة سيكومترية.

7- التكرار عبر مراحل تطوير المقياس: رغم أن التكرار قد يكون غير مرغوب فيه في النسخة النهائية، إلا أنه أقل إشكالية في المراحل الأولية لتطوير البنود. بل قد يكون من المفيد تضمين صيغ متقاربة لبند واحد (حتى لو اختلفت بكلمة واحدة) في الاختبار الأولي، بهدف مقارنة أدائها واختيار الأفضل منها لاحقاً.

8- مخاطر التكرار الشكلي: التكرار غير المرتبط بالمفهوم البنائي—كالتشابه في الصياغة أو البنية النحوية—لا يعد ميزة، بل قد يؤدي إلى تضخيم تقديرات الثبات. فعلى سبيل المثال، البنود التي تبدأ بعبارة مشتركة قد ترتبط ببعضها بسبب هذا التشابه الشكلي، وليس بسبب المتغير الكامن، وهو ما قد يؤدي إلى تقدير مبالغ فيه لمعاملات كمعامل ألفا كرونباخ.

9- التكرار المرتبط بالخصوصية المفرطة: قد تظهر مشكلة أخرى عندما تكون بعض البنود أكثر تحديداً من اللازم مقارنة ببقية البنود، مما يجعلها تنحرف عن المفهوم البنائي العام. فالبنود شديدة التحديد قد تعكس بُعداً فرعياً ضيقاً وتؤدي إلى تشويه القياس، خاصة إذا كانت متشابهة فيما بينها، حيث قد تكون "عنقوداً" خاصاً بها داخل المقياس (ديفيليس، 2021، ص147-207).

10- أثر التكرار على بنية المقياس: إذا تركز عدد كبير من البنود حول جانب معين، مثل (القلق ضمن مقياس عام للانفعالات)، فقد يؤدي ذلك إلى:

- الإخلال بأحادية البعد؛
- خلق تحيز لصالح هذا البعد؛
- زيادة مساهمته في الثبات مقارنة ببقية الأبعاد، حتى لو لم يكن أكثر أهمية نظرياً.

11- أهمية التوافق مع مستوى التحديد: يتحدد مدى قبول التكرار بحسب طبيعة المقياس:

- في المقاييس العامة، قد يكون التكرار المرتبط ببعد معين مشكلة.
- في المقاييس المتخصصة، مثل (مقياس القلق تحديداً)، فإن تشابه البنود يكون طبيعياً ومطلوباً ما دام يعكس المتغير المستهدف.

12- عدد البنود في المقياس: لا يوجد عدد محدد مثالي للبنود في المرحلة الأولية، لكن ينصح بأن يكون عددها أكبر بكثير من العدد المستهدف في النسخة النهائية. ويعود ذلك إلى أن ثبات الاتساق الداخلي يعتمد على كل من قوة ترابط البنود وعددها، بينما تكون هذه الترابطات غير معروفة في المراحل المبكرة. لذا فإن وفرة البنود تعد بمثابة ضمان أولي ضد ضعف الثبات. غالباً ما يبدأ الباحثون بتجمع بنود يعادل ثلاثة إلى أربعة أضعاف الحجم النهائي؛ فمثلاً، يمكن أن يشتق مقياس من 10 بنود من مجموعة أولية تضم نحو 40 بنداً. ومع ذلك، قد يكون التجمع أصغر في حال صعوبة توليد البنود أو إذا أظهرت البيانات أن عدداً أقل يكفي لتحقيق اتساق جيد.

13- الموازنة بين الشمول وإمكانية التطبيق: رغم أن زيادة عدد البنود أمر مرغوب فيه عموماً، إلا أن المبالغة قد تجعل الأداة غير قابلة للتطبيق في جلسة واحدة. لذلك يمكن تقليص التجمع الأولي وفق معايير محددة، مثل:

- غموض الصياغة؛
- ضعف الصلة بالمفهوم البنائي؛
- التكرار غير المرغوب بين البنود.

14- البدء في عملية كتابة البنود:

تعد مرحلة البدء في صياغة البنود من أصعب مراحل بناء المقياس. في هذه المرحلة، يفضل التركيز على توليد أكبر قدر ممكن من الأفكار المرتبطة بالمفهوم البنائي، دون الانشغال كثيراً بجودة الصياغة. ويمكن الانطلاق من عبارة بسيطة تعبّر عن المفهوم المراد قياسه، ثم إعادة صياغتها بطرق متعددة ومتنوعة. يعتمد هذا الأسلوب على الكتابة السريعة والتفكير الإبداعي، من خلال إنتاج صيغ مختلفة للفكرة نفسها، وتغيير الألفاظ أو السياقات للتعبير عن المفهوم بطرق متعددة. والهدف هنا هو استكشاف جميع الإمكانات الممكنة لتمثيل البناء النظري. بعد توليد عدد كبير من البنود—قد يصل إلى ثلاثة أو أربعة أضعاف العدد المطلوب—تبدأ مرحلة المراجعة والتنقيح، حيث يتم تقييم البنود من حيث وضوحها وقدرتها على تجسيد المفهوم البنائي بدقة.

تركز المرحلة الأولى على الكم والإبداع، بينما تخصص المرحلة اللاحقة للتقويم والاختيار، بما يضمن جودة البنود وملاءمتها لهدف القياس.

15- خصائص البنود الجيدة والسيئة: يصعب حصر جميع معايير جودة البنود، إلا أن الوضوح يظل المعيار الأهم؛ إذ يجب أن يكون البند دقيقاً وغير ملتبس، وأن يتجنب إرباك المجيب أو وضعه في حيرة.

16- الطول وسهولة القراءة: ينبغي تجنب البنود الطويلة لأنها تزيد التعقيد وتضعف الفهم، مع الحفاظ في الوقت نفسه على المعنى دون اختزال مخل. كما يعد مستوى سهولة القراءة عاملاً أساسياً؛ حيث توجد طرق لتقديره كالتي تربط بين طول الكلمات والجمل ومستوى القراءة (Dale & Chall, 1948؛ Fry, 1977). ويقترح (Fry, 1977)، حساب عدد الكلمات والجمل والمقاطع اللفظية في عينة نصية لتحديد مستوى القراءة المناسب، والذي يفضل أن يكون بين الصفين الخامس والسابع عند التعامل مع الجمهور العام. وقد كتبت بعض المقاييس، كمقاييس موضع الضبط الصحي، ضمن هذا المستوى (Wallston et al., 1978).

17- الوضوح اللغوي وتجنب التعقيد: لا يكفي قِصر الجمل لضمان الوضوح؛ إذ ينبغي مراعاة البنية الدلالية والنحوية. فبعض العبارات القصيرة قد تكون غامضة، خاصة عند استخدام كلمات غير شائعة أو تراكيب معقدة. كما يجب تجنب النفي المزدوج لأنه يزيد الالتباس.

- **البنود المزدوجة (Double-barreled):** التي تتضمن فكرتين في بند واحد، مما يربك تفسير الإجابة.
- **الضمائر الغامضة:** التي لا يعرف مرجعها بدقة.
- **المعدلات في غير محلها:** التي تؤدي إلى التباس المعنى.
- **الاختيارات اللغوية غير الدقيقة:** كاستخدام صيغ قد تغير المعنى المقصود.
- **غموض المعنى العام:** قد يكون الغموض على مستوى الجملة كاملة، وليس فقط الكلمات المفردة، مما يؤدي إلى تفسيرات غير مقصودة من قبل المجيبين، وقد يؤثر سلباً على جودة البيانات.

لهذا تعتمد جودة البنود على وضوحها، وبساطتها، وخلوها من الالتباس اللغوي أو التركيبي، مع مراعاة مستوى القراءة المناسب وتجنب الأخطاء الشائعة التي قد تشوه القياس.

18- البنود ذات الصياغة الإيجابية والسلبية: يلجأ بعض مطوري المقاييس إلى استخدام بنود إيجابية وأخرى سلبية في الأداة نفسها، بحيث تعكس الأولى وجود المتغير الكامن، بينما تعكس الثانية غيابه. ويهدف هذا الأسلوب غالباً إلى الحد من تحيز الموافقة (Acquiescence bias)، أي ميل بعض المجيبين إلى الموافقة على البنود بغض النظر عن محتواها. فعلى سبيل المثال، يتضمن مقياس تقدير الذات لـ Morris Rosenberg (1965) بنوداً إيجابية مثل "أشعر بأن لدي صفات جيدة"، وأخرى سلبية

مثل "أشعر أنني عديم الفائدة أحياناً". ويتيح هذا التنوع التمييز بين تقدير الذات الحقيقي وبين نمط الموافقة غير المدروس؛ إذ إن صاحب تقدير الذات المرتفع يوافق على البنود الإيجابية ويعارض السلبية، بينما قد يوافق "المتحيز للموافقة" على كليهما.

19- مشكلة عكس الصياغة: رغم فائدته النظرية، قد يؤدي عكس اتجاه البنود إلى نتائج سلبية، خاصة عند زيادة طول المقياس، إذ يمكن أن يسبب ارتباكاً لدى المجيبين بين معنى الموافقة على العبارة وبين مستوى السمة المقاسة. وقد أظهرت دراسات، مثل (DeVellis & Currey, Callahan, & DeVellis, 2002 ;Callahan, 1993) أن بعض البنود المعكوسة تؤدي أداءً ضعيفاً، بينما يتحسن أداؤها عند إعادة صياغتها بشكل مباشر ومتسق مع بقية البنود. ويرجح أن هذا يعود إلى عدم انتباه بعض المجيبين لعملية العكس، مما يضعف الارتباط الداخلي بين البنود ويؤثر سلباً على الثبات.

الخلاصة: يجب أن يكون بنك البنود (item pool) واسعاً ومتنوعاً، بحيث يثري القياس ويعزز الاتساق الداخلي، مع تجنب البنود المركبة أو المزدوجة التي تربك المجيب. كما ينبغي أن تصاغ البنود وفق قواعد لغوية واضحة، سواء كانت إيجابية أو سلبية، مع مراعاة أن التكرار المرتبط بالمحتوى مفيد في دعم الثبات وليس عيباً في حد ذاته.

الخطوة الثالثة: تحديد شكل أو تنسيق القياس:

توجد أشكال وصيغ متعددة للأسئلة، ولذلك ينبغي على الباحث أن يحدد منذ المراحل الأولى الشكل الذي سيتخذه المقياس. ويجب أن تتم هذه الخطوة بالتوازي مع عملية توليد البنود، بحيث يحدث انسجام وتكامل بين صياغة البنود ونوع الاستجابة المعتمد. فمثلاً، قد يكون من غير المفيد إعداد قائمة طويلة من العبارات التفصيلية إذا كان الشكل النهائي للمقياس يعتمد على بنود قصيرة أو قائمة مراجعة مكونة من كلمات مفردة.

1- مقياس ثرستون (Thurstone):

هناك استراتيجيات متعددة لبناء المقاييس تؤثر في شكل البنود وخيارات الاستجابة، ومن أبرزها مقياس (Louis Leon Thurstone). ويمكن توضيح فكرته من خلال تشبيه الشوكة الرنانة التي تهتز فقط عند توافقها مع تردد معين؛ فإذا تعرضت لتردد مطابق لترددها الطبيعي فإنها تهتز، بينما تبقى ساكنة أمام الترددات الأخرى. وبالمثل، يمكن تصور مجموعة من الشوكات الرنانة مرتبة تدريجياً بحيث تمثل تدرجاً في الترددات، مما يسمح بتحديد التردد الدقيق من خلال الشوكة التي تستجيب. ويطبق مقياس "ثرستون" الفكرة نفسها في القياس النفسي؛ إذ يحاول الباحث بناء بنود تستجيب

بدرجات مختلفة لمستويات محددة من السمة الكامنة. وعندما تتطابق خصائص البند مع مستوى السمة لدى المجيب، يظهر هذا التطابق في الاستجابة. وغالباً ما تفسر الاستجابة الإيجابية على البنود المطابقة للمستوى المناسب، في حين ترفض البنود الأخرى. ويتم تحديد مستويات البنود مسبقاً من خلال حكم الخبراء الذين يصنفونها ضمن فئات أو فواصل متساوية تمثل درجات شدة السمة. وتعد هذه الطريقة من الأساليب القوية في بناء المقاييس، إذ تسمح بتصميم بنود تمثل مستويات مختلفة من السمة بطريقة متدرجة ومتساوية الفواصل، مع إمكانية استخدام خيارات استجابة مثل (موافق/غير موافق). وبعد تطبيق المقياس، تستخدم الاستجابات لتحديد مستوى السمة لدى الفرد بناءً على البنود التي تم الاتفاق معها. ويعد هذا الأسلوب مناسباً لإنتاج مقاييس يمكن التعامل مع نتائجها كبيانات على مستوى القياس الفتري، نظراً لافتراض تساوي الفواصل بين مستويات البنود.

وقد يبدو جزء من مقياس "ثورستون" الافتراضي لقياس تطلعات الآباء للتحصيل التعليمي والوظيفي لأطفالهم كما يلي:

- تحقيق النجاح هو السبيل الوحيد لطفلي أن يسدد الجهود التي بذلتها كوالد.	موافق	غير موافق ....
- الذهاب إلى كلية جيدة والحصول على عمل جيد أمرٌ مهمّ ولكن ليس ضرورياً لسعادة طفلي.	موافق	غير موافق ....
- السعادة ليست لها علاقة بتحقيق الأهداف التعليمية أو المادية.	موافق	غير موافق ....
- زخارف النجاح ذات القيمة المعتادة ليست عائقاً أمام السعادة الحقيقية.	موافق	غير موافق ....

وكما يشير James McKeen Cattell (1978) ، فإن بناء مقياس ثرستون الحقيقي أكثر تعقيداً بكثير مما يبدو في وصفه النظري. فالعثور على بنود "تستجيب" بشكل ثابت لمستويات محددة من السمة يعد أمراً بالغ الصعوبة في التطبيق العملي. وغالباً ما تفوق المشكلات التطبيقية لهذا الأسلوب فوائده، إلا إذا كان لدى الباحث مبرر قوي يتطلب هذا النوع من المعايرة الدقيقة. وعلى الرغم من أن مقياس ثرستون يعد منهجاً مهماً ومفيداً في بعض السياقات، فإنه لن يتم التوسع فيه لاحقاً في هذه المحاضرة. ومع ذلك، تجدر الإشارة إلى أن أساليب نظرية استجابة البند (IRT) اللاحقة تشترك مع هذا النموذج في عدد من الأهداف، لكنها تعتمد إجراءات مختلفة لتحقيقها.

2- مقياس جتمان (Guttman):

مقياس (Louis Guttman) هو مقياس تراكمي يتكون من سلسلة بنود مرتبة بحيث تعكس تدرجاً في شدة السمة. وفق هذا النموذج، إذا وافق المجيب على بند معين، فإنه يفترض أنه وافق أيضاً على جميع البنود الأقل منه شدة، ويتوقف عند نقطة معينة يتجاوز فيها مستوى السمة لديه. فعلى سبيل المثال، في أسئلة التدخين: "هل تدخن؟"، "هل تدخن أكثر من 10 سجائر يومياً؟"، "هل تدخن أكثر من علبة يومياً؟"، فإن الموافقة على بند أعلى تعني ضمنياً الموافقة على البنود السابقة له. ويختلف هذا النموذج عن مقياس ثرستون في أن الأخير يركز على بند واحد "يتوافق" مع مستوى السمة، بينما يركز مقياس جتمان على نقطة التحول من الموافقة إلى عدم الموافقة عبر سلسلة مرتبة من البنود. وقد يبدو جزء من مقياس جتمان الافتراضي لقياس تطلعات الآباء للتحصيل التعليمي والوظيفي لأطفالهم كما يلي:

- تحقيق النجاح هو السبيل الوحيد لطفلي أن يسدد الجهود التي بذلتها كوالد.	موافق	غير موافق ....
- الذهاب إلى كلية جيدة والحصول على عمل جيد أمرٌ مهمّ ولكن ليس ضرورياً لسعادة طفلي.	موافق	غير موافق ....
- السعادة ليست لها علاقة بتحقيق الأهداف التعليمية أو المادية.	موافق	غير موافق ....
- زخارف النجاح ذات القيمة المعتادة ليست عائقاً أمام السعادة الحقيقية.	موافق	غير موافق ....

يمكن أن تعمل مقاييس "جتمان" بشكل جيد في الحالات التي تكون فيها المعلومات موضوعية، أو عندما يكون هناك منطق تراكمي واضح يفترض أن الموافقة على مستوى أعلى تعني بالضرورة تحقق المستويات الأدنى منه داخل نفس التسلسل الهرمي. أما عندما تكون الظاهرة المراد قياسها غير ملموسة أو نفسية بطبيعتها، فإن هذا الافتراض يصبح أقل دقة وأكثر إشكالاً. ففي مثال مقياس الطموح الأبوي الافتراضي، قد لا يكون ترتيب البنود موحداً بين جميع الأفراد؛ فبينما يعني تدخين 20 سيجارة يومياً بالضرورة تدخين أكثر من 10 سجائر، فإن الاستجابات على البنود المتعلقة بالطموح الأبوي لا تتبع دائماً هذا النمط التراكمي الصارم. فقد يوافق المجيب على بند معين مثل (3) لكنه لا يوافق على بند آخر أعلى منه مثل (4)، رغم افتراض مقياس جتمان أن الموافقة على الأعلى تستلزم

الموافقة على الأدنى. ويحدث هذا النوع من التناقض عندما ينظر إلى الظاهرة، مثل (النجاح أو الطموح) بوصفها بنية معقدة قد تجمع بين عناصر متعارضة، مثل كون النجاح عاملاً مساعداً وفي الوقت نفسه مصدراً للضغط أو العائق(ديفيليس، 2021، ص147-207).

وكما هو الحال في مقياس ثرستون، فإن لمقياس جتمان استخداماته، إلا أن نطاق تطبيقه في القياس النفسي يبدو محدوداً نسبياً، وغالباً ما تفوق صعوباته العملية مزاياه. ومن المهم التأكيد أيضاً على أن بعض الافتراضات النظرية التي تقوم عليها هذه النماذج لا تنطبق بالضرورة على جميع أنواع المقاييس، خاصة تلك التي تعتمد على بنى كامنة غير مباشرة. كما أن افتراض وجود علاقة سببية قوية ومباشرة بين المتغير الكامن وكل بند لا يكون مناسباً دائماً في مقياسي ثرستون وجتمان. وقد قدم James S. Nunnally و Ira H. Bernstein (1994) عرضاً مختصراً للنماذج المفاهيمية التي تقوم عليها هذه المقاييس. أما في الحالات التي تكون فيها البنود المرتبة مناسبة بالفعل، فإن نماذج نظرية استجابة البند (IRT) قد توفر بديلاً أكثر دقة، رغم أن تطبيقها قد يكون معقداً ويتطلب جهداً إجرائياً كبيراً.

2-1- المقاييس ذات البنود متساوية الترجيح أو الوزن: تعتمد على نماذج القياس التي تم تناولها سابقاً، والتي تفترض أن البنود تعمل إلى حد ما كـ"أدوات قياس" للظاهرة موضع الاهتمام. أي أن هذه البنود تعد مؤشرات متقاربة في قدرتها على تمثيل المتغير الكامن، دون أن تكون متكافئة تماماً بالمعنى الصارم لمفهوم الاختبارات المتوازية. وهي في جوهرها مؤشرات غير كاملة لمفهوم مشترك، يمكن دمجها ببساطة من خلال الجمع للحصول على درجة كلية تتمتع بدرجة مقبولة من الثبات. ومن المزايا المهمة لهذا النوع من المقاييس أن كل بند يمكن أن يصاغ باستخدام أشكال متعددة من خيارات الاستجابة، مما يمنح الباحث مرونة كبيرة في تصميم الأداة بما يتناسب مع الهدف المحدد للقياس. وهذا التنوع في صياغة البنود وأنماط الاستجابة يتيح إمكانات واسعة لبناء مقياس ملائم من حيث الدقة والوظيفة.

وفيما يلي سيتم تناول بعض القضايا العامة المرتبطة بشكل وصياغة الاستجابة، إضافة إلى عرض مزايا وخصائص بعض الصيغ الشائعة للاستجابة.

2-2- كم عدد فئات الاستجابة؟

- تتكون أغلب بنود المقاييس من عنصرين أساسيين: جذع (Stem) يتضمن العبارة أو السؤال، وسلسلة من خيارات الاستجابة (السلم). فعلى سبيل المثال، قد يكون الجذع عبارة عن فقرة تعبر عن

رأي معين، بينما تكون خيارات الاستجابة على شكل تدريج يعبر عن شدة الاتفاق أو عدمه. ويتمثل السؤال المحوري هنا في عدد فئات الاستجابة المناسبة لكل بند. إذ تختلف الصيغ؛ فبعضها يسمح بعدد كبير جداً أو حتى مستمر من الخيارات، بينما يفرض بعضها الآخر عدداً محدوداً منها. فمثلاً، يمكن استخدام تدريج يشبه "ميزان الحرارة" لقياس الغضب من "لا غضب إطلاقاً" إلى "غضب شديد جداً"، حيث يضع المجيب إشارته على نقطة متصلة. كما يمكن استخدام مقياس رقمي من 1 إلى 100، أو بدائل محدودة مثل: "لا شيء"، "قليل"، "متوسط"، "كثير"، أو حتى خيارين فقط مثل "غاضب/غير غاضب".

2-3- المزايا النسبية لعدد فئات الاستجابة: تعد القدرة على إظهار التباين من أهم خصائص جودة القياس، إذ لا يمكن للمقياس أن يكون مفيداً دون القدرة على التمييز بين الأفراد. ومن الطرق التي تعزز هذا التباين: زيادة عدد البنود أو زيادة عدد فئات الاستجابة داخل كل بند. فعلى سبيل المثال، إذا كان الباحث مضطراً للاعتماد على عدد قليل من البنود، مثل (بندين فقط لقياس الغضب في بيئة العمل)، فإن استخدام تدريج واسع، مثل (0–100) قد يوفر معلومات أكثر دقة من الاستجابات الثنائية (غاضب/غير غاضب). أما إذا كان عدد البنود كبيراً، مثل (50 بنداً)، فقد تكون الفئات البسيطة كافية، لأن تراكم البنود يعوض محدودية تدريج الاستجابة. لكن زيادة عدد الفئات بشكل مبالغ فيه قد تؤدي إلى إرهاق المجيب أو فقدان الدقة، حيث لا يستطيع كثير من الأفراد التمييز فعلياً بين عدد كبير جداً من الدرجات، مما يجعل الفروق الدقيقة، مثل (35 مقابل 37) غير ذات معنى حقيقي، ويزيد من الخطأ بدلاً من تحسين القياس.

2-4- حدود الدقة في الاستجابة: تعتمد قدرة المجيب على التمييز بين الفئات على طبيعة الظاهرة المقاسة وصياغة خيارات الاستجابة. فالتوصيفات الغامضة مثل "قليل"، "متوسط"، "كثير" قد تفسر بطرق مختلفة بين الأفراد. ويمكن تقليل هذا الغموض من خلال ترتيب واضح للفئات أو تقديمها على شكل متصل بصري يساعد على فهم التدرج بشكل أفضل، مثل:

كثيراً	بعض الأحيان	قليلاً	قليلاً جداً	أبداً

وقد يعني هذا ضمنياً أن الخيار "بعض الأحيان" هو أكثر من خيار الاستجابة "قليلاً" بسبب ترتيب هذه البنود. ومع ذلك، إذا كان من الممكن إيجاد صفة لا لبس فيها تمنع المجيبين من وضع افتراضات تستند إلى الموضع أو المكان على طول سلسلة متصلة، فإن هذا يكون أفضل بكثير. وفي

بعض الأحيان، قد يكون من الأفضل أن يكون لديك عدد من خيارات الاستجابة أقل من تلك التي تكون غامضة. لذلك، على سبيل المثال، قد يكون من الأفضل في المثال السابق حذف إما "بعض الأحيان" أو "قليلاً" ويكون لديك أربعة خيارات بدلاً من خمسة. وأسوأ الظروف هي الجمع بين الكلمات الغامضة ومواقع الصفحة الغامضة. فكر في المثال التالي:

مفيد جداً Very Helpful	غير مفيد جداً Not Very Helpful
مفيد إلى حدٍ ما Somewhat Helpful	غير مفيد على الإطلاق Not at All Helpful

ومن الصعب جداً التمييز بين مصطلحات مثل "somewhat" إلى حدٍ ما وليس إلى حدٍ كبير " not very" في ظل أفضل الظروف. ومع ذلك، فإن ترتيب خيارات الاستجابة هذه كما تظهر أعلاه يجعل الأمور أكثر سوءاً. إذا قام المجيب بقراءة العمود الأول ثم أسفل العمود الثاني، فيبدو أن تعبير "somewhat" إلى حدٍ ما يمثل قيمة أعلى من تعبير "not very" ليس إلى حدٍ كبير. ولكن إذا قرأ المجيب الصف الأول ثم عبر الصف الثاني، فإن الترتيب الضمني لهذين الاستجابتين الواصفتين على طول سلسلة متصلة يتم عكسه. ونظراً للغموض الذي يكتنف كلاً من الترتيب اللغوي والترتيب المكاني، يمكن للأفراد أن يعطوا معاني مختلفة للخيارين اللذين يمثلان قيمتين معتدلتين، وكنتيجة لذلك، فإن الثبات سيتأثر(ديفيليس، 2021، ص147-207).

وثمة مسألة أخرى تتمثل في قدرة الباحث واستعداده لتسجيل عدد كبير من القيم لكل بند. إذا تم استخدام طريقة مقياس الحرارة التي تم وصفها سابقاً لتحديد الاستجابات الكمية على مقياس الغضب، فهل سيحاول الباحث بالفعل تسجيل درجات دقيقة لكل استجابة؟ ما هي الدقة المناسبة؟ هل يمكن قياس المنطقة المظللة في حدود ربع بوصة؟ سنتيمتر؟ مليمتر؟ إذا تم استخراج بعض البيانات الخام فقط - مثل الثلث الأدنى أو الأوسط أو الأعلى - من المقياس، فما هي النقطة التي كانت عندها المطالبة بمثل هذه الاستجابة الدقيقة؟

توجد مسألة أخرى تتعلق بعدد بدائل الاستجابة لكل بند: هل يكون العدد فرديًا أم زوجيًا؟ يعتمد ذلك على طبيعة السؤال، ونوع مقياس الاستجابة، وهدف الباحث. فعندما تكون الاستجابات ثنائية القطب، مثل (اتجاه إيجابي قوي مقابل اتجاه سلبي قوي)، فإن العدد الفردي يتيح خيارًا محايدًا في الوسط، مثل "لا أتفق ولا أعارض" أو "غير متأكد"، وهو ما يسمح بقدر من الغموض أو التردد

في الإجابة. في المقابل، فإن العدد الزوجي من الاستجابات لا يتضمن خيارًا محايدًا، مما يدفع المجيب إلى اتخاذ موقف even—لو كان ضعيفًا—نحو أحد القطبين (كأن يختار "إيجابي إلى حد ما" أو "سلبي إلى حد ما").

ولا يعد أي من الشكلين أفضل بشكل مطلق، إذ يتوقف الاختيار على غرض الدراسة. فقد يفضل الباحث استخدام عدد زوجي لتقليل الميل إلى اختيار الإجابة المحايدة بدافع تجنب القرار. فعلى سبيل المثال، في دراسات المقارنة الاجتماعية، قد يكون من المفيد إلزام المشاركين بالتعبير عن تفضيل واضح بين البدائل. وقد استخدم هذا الأسلوب في دراسة المقارنات الاجتماعية لدى المصابين بالتهاب المفاصل (DeVellis et al., 1990).

1- هل تفضل معلومات حول: (أ) المرضى الذين يعانون من التهاب المفاصل أسوأ مما لديك. (ب) المرضى الذين يعانون من التهاب المفاصل أكثر اعتدالاً مما لديك.

2- هل تفضل معلومات حول: (أ) المرضى الذين يعانون من التهاب المفاصل أسوأ مما لديك. (ب) المرضى الذين يعانون من التهاب المفاصل بنفس القدر الذي تعاني منه. (ج) المرضى الذين يعانون من التهاب المفاصل أكثر اعتدالاً مما لديك.

قد يسمح خيار محايد مثل رقم 2 (ب) بالمراوغة (اللبس والغموض) غير المرغوب فيها. وقد يكون من المطلوب والمستحسن أيضاً وجود نقطة محايدة. في دراسة تقييم أي نوع من المخاطر، مثلاً (الملل مقابل الخطر) الذي يفضله الناس، قد تكون نقطة الوسط ضرورية أو حاسمة. قد يغير الباحث فرصة أو شدة الضرر عبر عدة خيارات بين نشاط آمن وممل ونشاط مثير ومحفوف بالمخاطر. ويمكن عندئذ استخدام النقطة التي يكون فيها المجيب ملتبساً إلى أبعد حد بشأن المخاطرة بالنشاط الأكثر إثارة كمؤشر للمخاطرة:

يرجى الإشارة إلى تفضيلك النسبي للنشاط (أ) أو النشاط (ب) من بين البدائل المدرجة أدناه من خلال وضع دائرة حول العبارة المناسبة بعد وصف النشاط (ب).

النشاط (أ): قراءة كتاب إحصائي (لا توجد فرصة لحدوث إصابة شديدة).

1- **النشاط (ب):** القيام برحلة جوية في طائرة ركاب صغيرة (احتمال ضئيل جداً بحدوث إصابة بالغة).

أفضل بشدة النشاط (أ)	أفضل بشكل معتدل النشاط (أ)	لا أفضل أي منهما	أفضل بشكل معتدل النشاط (ب)	أفضل بشدة النشاط (ب)

2- النشاط (ب): القيام برحلة على متن طائرة صغيرة في قمرة القيادة (فرصة طفيفة للإصابة الشديدة).

أفضل بشدة النشاط (أ)	أفضل بشكل معتدل النشاط (أ)	لا أفضل أي منهما	أفضل بشكل معتدل النشاط (ب)	أفضل بشدة النشاط (ب)

3- النشاط (ب): القفز بالمظلة من طائرة مع مظلة احتياطية (فرصة بشكل معتدل لإصابة شديدة).

أفضل بشدة النشاط (أ)	أفضل بشكل معتدل النشاط (أ)	لا أفضل أي منهما	أفضل بشكل معتدل النشاط (ب)	أفضل بشدة النشاط (ب)

4- النشاط (ب): القفز بالمظلة من طائرة بدون مظلة احتياطية (خطر كبير من الإصابة الشديدة).

أفضل بشدة النشاط (أ)	أفضل بشكل معتدل النشاط (أ)	لا أفضل أي منهما	أفضل بشكل معتدل النشاط (ب)	أفضل بشدة النشاط (ب)

5- النشاط (ب): القفز من طائرة بدون مظلة ومحاولة الهبوط على هدف ناعم (إصابة خطيرة مؤكدة تقريباً).

أفضل بشدة النشاط (أ)	أفضل بشكل معتدل النشاط (أ)	لا أفضل أي منهما	أفضل بشكل معتدل النشاط (ب)	أفضل بشدة النشاط (ب)

وبصرف النظر عن المزايا أو الالتزامات الأخرى لهذا النهج أو الأسلوب، فإنه يتطلب بوضوح أن تتضمن خيارات الاستجابة نقطة وسط أو نقطة منتصف.

2-5- أنواع محددة من أشكال وتنسيقات الاستجابة (Response Formats):

تظهر بنود المقاييس في مجموعة واسعة من الأشكال والصيغ، غير أن هناك أنماطًا شائعة الاستخدام أثبتت فعاليتها في تطبيقات القياس النفسي والتربوي. وفيما يلي عرض لأهم هذه الصيغ.

3- مقياس ليكرت (Likert Scale):

يعد مقياس ليكرت (Likert, 1932) من أكثر أشكال البنود شيوعًا في تصميم الاستبيانات ومقاييس الاتجاهات. ويعرض البند عادة في صورة عبارة تقريرية يتبعها مجموعة من بدائل الاستجابة (Response Options)التي تعكس درجات متفاوتة من الاتفاق أو التأييد.

وتبعًا لطبيعة الظاهرة المدروسة وأهداف البحث، قد يتضمن كل بند عددًا فرديًا أو زوجيًا من فئات الاستجابة. ويجب أن تصاغ هذه الفئات بحيث تكون متدرجة بشكل متقارب أو شبه متساوٍ من حيث شدة الاتفاق، أي أن الفرق بين أي فئتين متجاورتين يجب أن يكون مماثلًا تقريبًا لبقية الفروق

بين الفئات. ومن الصيغ الشائعة لمقياس ليكرت استخدام ست فئات استجابة مثل: "غير موافق بشدة"، "غير موافق إلى حد ما"، "غير موافق بشكل متوسط" ، "موافق بشكل متوسط"، "موافق إلى حد ما"، و"موافق بشدة". ويمثل هذا تسلسلًا متصلًا من الرفض الشديد إلى الموافقة الشديدة. وقد يضاف خيار وسط (Midpoint) أو فئة محايدة (Neutral Category)، مثل: "لا أوافق ولا أعارض (Neither Agree nor Disagree) " أو "محايد (Neutral)". وهناك نقاش منهجي حول دلالة هذه النقطة الوسطية؛ إذ قد تعكس حيادًا حقيقيًا أو عدم اهتمام، أو توازنًا بين الاتجاهين. ومع ذلك، فإن كثيرًا من المستجيبين لا يميزون بدقة بين هذه الفروق اللغوية، ويعتبرونها ببساطة خيارًا وسطًا. يستخدم مقياس ليكرت على نطاق واسع في قياس الاتجاهات (Attitudes)، والمعتقدات (Beliefs)، والقيم (Values). ومن المهم أن تكون البنود قوية إلى حد معتدل في صياغتها، بحيث تعكس اختلافات حقيقية في الاتجاه دون إفراط في التعميم أو التطرف. فعلى سبيل المثال، يمكن مقارنة ثلاث صيغ لبند واحد يتعلق باتجاهات الأطباء نحو المرضى:

- يتجاهل الأطباء عمومًا ما يقوله المرضى (Strong Statement).
- في بعض الأحيان لا يولي الأطباء الاهتمام الكافي لتعليقات المرضى (Moderate Statement).
- قد ينسى الأطباء أحيانًا بعض ما يقوله المرضى (Weak Statement).

وتعكس هذه الصيغ مستويات مختلفة من شدة الاتجاه (Item Intensity). ويكمن السؤال المنهجي في أيها أفضل لبناء مقياس ليكرت؟ والإجابة تعتمد على مدى قدرتها على تمثيل الفروق الحقيقية في الاتجاهات بين الأفراد. فالبند الجيد هو الذي يميز بدقة بين الأفراد ذوي الاتجاهات المختلفة. وقد يؤدي استخدام بنود معتدلة جدًا إلى اتفاق مرتفع بشكل غير تمييزي (Ceiling Effect)، حيث يوافق معظم الأفراد بغض النظر عن اختلافاتهم الحقيقية. مثال ذلك عبارة: "سلامة وأمن المواطنين أمر مهم". وعلى الرغم من صحتها العامة، فإنها لا تميز بين الأفراد ذوي الاتجاهات المختلفة. وفي المقابل، قد تؤدي البنود شديدة التطرف إلى رفض واسع حتى من الأفراد الذين يتبنون الاتجاه العام، مما يقلل من فعاليتها التمييزية(ديفيليس، 2021، ص147-207).

وتشير الأدبيات المنهجية إلى أن البنود المعتدلة جدًا قد تكون أكثر إشكالية من البنود المتطرفة، لأنها غالبًا تعكس غياب الرأي أكثر من كونها تمثل اتجاهًا حقيقيًا. وبالتالي قد لا تخدم هدف البحث الذي يركز عادة على قياس وجود الظاهرة وشدتها.

ومن الطرق المفيدة لضبط قوة صياغة البنود (Item Calibration) أن يتخيل الباحث المستجيب النموذجي (Typical Respondent)، وكيف سيستجيب لبنود تختلف في شدتها. وبناءً على ذلك، تصاغ البنود بحيث تستهدف استجابات تقع بالقرب من الفئة الوسطية في المقياس (Central Response Category)، مما يسمح بظهور تباين كافٍ في الاستجابات. هذا التباين يساهم في رفع الارتباط بين البنود (Inter-item Correlation) ويعزز الثبات (Reliability) والصدق (Validity). أما البنود التي تؤدي إلى استجابات متطرفة جدًا من المستجيب النموذجي، فقد تقل قدرتها على التمييز بين فئات الأفراد المختلفة. فيما يلي أمثلة للبنود بصيغ وأشكال استجابة مقياس "ليكرت":

1- تعتبر ممارسة التمارين الرياضية عنصراً أساسياً في نمط حياة صحي.

6	5	4	3	2	1
موافق بشدة	موافق إلى حد ما	موافق بشكل معتدل	غير موافق بشكل معتدل	غير موافق إلى حد ما	غير موافق بشدة

2- ينبغي أن تكون مكافحة تعاطي المخدرات أولوية وطنية عليا أو قصوى.

5	4	3	2	1
غير صحيح تماماً	غير صحيح في الغالب	صحيح وغير صحيح على حد سواء	صحيح في الغالب	صحيح تماماً

4- القياس التفاضلي الدلالي (تباين الدلالات اللفظية):

ترتبط طريقة القياس التفاضلي الدلالي بشكل رئيسي ببحوث الاتجاه التي أجراها على سبيل المثال "أوسقود وزملاؤه" (Osgood & Tannenbaum, 1955). وعادة ما يستخدم تباين الدلالات اللفظية للإشارة إلى واحد أو أكثر من الدوافع أو المحفزات. ففي حالة الاتجاهات، على سبيل المثال، قد يكون الحافز أو الدافع مجموعة من الناس، مثل مندوبي مبيعات السيارات. ويتبع تحديد الدوافع والحوافز المستهدفة قائمة من أزواج الصفات، حيث يمثل كل زوج طرفي نقيض لسلسلة متصلة، محددة بالصفات، مثلاً (أمين / غير أمين أو صادق / غير صادق). وكما هو موضح في المثال أدناه، توجد عدة نقاط (سطور) بين الصفات التي تشكل خيارات الاستجابة:

مندوب مبيعات السيارات:								
أمين (صادق)	-	-	-	-	-	-	-	غير أمين
هادئ (لطيف)	-	-	-	-	-	-	-	مزعج

في الأساس، تمثل الخطوط الفردية (سبعة أو تسعة أرقام شائعة) نقاطًا على طول السلسلة المتصلة المحددة بواسطة الصفات. يضع المجيب علامة أو إشارة على أحد الخطوط للإشارة إلى النقطة الموجودة على طول السلسلة المتصلة التي يتسم بها تقييمه للحافز أو الدافع. على سبيل المثال، إذا كان شخص ما يعتبر مندوب مبيعات السيارات غير أمين للغاية، فقد يختار الخط الأقرب إلى تلك الصفة. ويمكن التعبير عن وجهات النظر والآراء المتطرفة أو المعتدلة من خلال اختيار الخط المناسب لوضع العلامة عليه. وبعد تقييم الدوافع "الحوافز" بالنسبة لزوج الصفات الأول، ينتقل الفرد إلى أزواج صفات إضافية تفصل بينها خطوط.

إن الصفات التي يختارها الفرد قد تكون إما ثنائية القطب أو أحادية القطب، اعتمادًا، كما هو الحال دائمًا، على المنطق الذي تستند إليه الأسئلة البحثية التي يهدف المقياس إلى معالجتها. فكل الصفات ثنائية القطب تعبر عن وجود سمات متناقضة، مثل الودية friendly والعدائية hostile. وتشير أزواج الصفات أحادية القطب إلى وجود سمة واحدة أو عدم وجودها (غيابها)، مثل الودية friendly وغير الودية. وكما هو الحال مع مقياس "ليكرت"، يمكن أن يكون شكل وصيغة الاستجابة التفاضلية الدلالية متوافقًا إلى حد كبير مع النماذج النظرية التي تم عرضها في الفصول السابقة من هذا الكتاب. ويمكن كتابة مجموعات من البنود للاستفادة من نفس المتغير الأساسي. فعلى سبيل المثال، يمكن إضافة البنود التي تستخدم نقاط النهاية التالية: جدير بالثقة / غير جدير بالثقة، عادل / غير عادل، صادق / غير صادق، كنقاط نهاية للعبارة الأولى في المثال السابق بحيث تشكل مقياس «الشرف أو الأمانة». ويمكن تصور مثل هذا المقياس على أنه مجموعة من البنود التي تشترك في متغير كامن مشترك هو (الشرف أو الأمانة) وتتوافق مع الافتراضات التي تمت مناقشتها في الفصل الثاني. وبناءً على ذلك، يمكن إضافة وتحليل العشرات من البنود كل على حدة "لمقياس الأمانة" على النحو المبين في قسم لاحق فيما يتعلق بتقييم البنود.

5- التناظر البصري:

هناك تنسيق آخر لشكل وصيغة البند يشبه إلى حد ما القياس التفاضلي الدلالي وهو مقياس التناظر البصري. يقدم شكل وصيغة الاستجابة هذا للمستجيب خطًا متصلًا بين زوج من التوصيفات التي تمثل نهايات متقابلة لسلسلة متصلة. ويطلب من الشخص الذي يكمل البند وضع علامة عند نقطة على الخط تمثل رأيه أو خبرته أو معتقده أو ما يتم قياسه. إن مقياس التناظر البصري، كما يوحي مصطلح "التناظرية" في الاسم، هو مقياس مستمر أو متصل، حيث يقوم الباحث بتحديد دقة التمايز في تعيين الدرجات إلى نقاط المقياس(ديفيليس، 2021، ص147-207).

وقد تم في وقت سابق مناقشة بعض مزايا وعيوب شكل أو صيغة الاستجابة المستمرة. وهناك مسألة أو مشكلة إضافية لم تتم إثارتها في ذلك الوقت تتعلق بالاختلافات المحتملة في تفسير المساحة المادية من حيث صلتها بالقيم على المتصل. فقد لا تعني العلامة الموضوعة في نقطة معينة على طول الخط نفس الشيء بالنسبة لأشخاص مختلفين، حتى عندما يتم تمييز نقاط النهاية للخط بشكل مماثل لجميع المجيبين. فكر في مقياس تناظر بصري للألم مثل التالي:

لا ألم على الإطلاق ..... أسوأ ألم واجهته أو جربته.

هل تشير الاستجابة في منتصف المقياس إلى الألم الذي يصيب نصف الوقت، أو الألم المستمر بنصف الشدة المحتملة، أو شيء آخر بالكامل؟ جزء من المشكلة في قياس الألم هو أنه يمكن تقييمه على أبعاد متعددة، بما في ذلك التكرار، والشدة، والمدة. كما أن ذكريات أسوأ ألم قد تعرض له شخص ما من المرجح أن تكون مشوهة. والواقع أن المقارنات بين الأفراد تزداد تعقيدًا بفعل حقيقة مفادها أن الأشخاص المختلفين ربما قد تعرضوا لمستويات مختلفة من "أسوأ ألم".

ومن المميزات الرئيسية لمقاييس التناظر البصري أنها قد تكون حساسة للغاية (Mayer, 1978)، وهذا من شأنه أن يجعلها مفيدة بشكل خاص في قياس الظواهر قبل وبعد وقوع بعض الأحداث المتداخلة، كالتدخل أو المعالجة التجريبية، التي تمارس تأثيرًا ضعيفًا نسبيًا. على سبيل المثال، قد لا يؤدي التوبيخ الخفيف أو المتوسط في سياق المعالجة التجريبية إلى إحداث تحول في مقياس لاحترام الذات ذي تدريج خماسي النقاط. ولكن ربما قد يحدث تحول دقيق ولكنه منهجي أو منتظم يؤدي إلى انخفاض القيم على مقياس تناظري بصري بين الأشخاص في حالة "التوبيخ" في هذه التجربة

الافتراضية. وقد تكون الحساسية أكثر فائدة عند فحص ودراسة التغيرات على مر الزمن داخل نفس الفرد وليس بين الأفراد (Mayer, 1978). وقد يكون الأمر كذلك لأنه في الحالة الأولى، لا يوجد خطأ إضافي بسبب الاختلافات الخارجية بين الأفراد.

وهناك ميزة أخرى محتملة لمقاييس التناظر البصري عند تكرارها بمرور الوقت، وهي أنه من الصعب أو المستحيل على الأشخاص ترميز "تشفير" إجاباتهم السابقة بدقة. بالاستمرار في المثال الوارد في الفقرة السابقة، ربما لا يجد شخص ما صعوبة كبيرة في تذكر أي من الخيارات الخمسة المرقمة بالنسبة لبند تقدير الذات الذي كان قد اختاره أو اختارته من قبل استجابة لشكل أو صيغة متعددة الاستجابات مثل مقياس "ليكرت". إلا أنه ما لم يتم اختيار إحدى نقاط النهاية لمقياس تناظر بصري، سيكون من الصعب أن نتذكر بدقة المكان الذي تم فيه وضع علامة على طول خط غير مميز. وقد يكون ذلك مفيدًا إذا كان الباحث يشعر بالقلق من أن المجيبين قد ينحازون إلى الظهور بمظهر متسق مع مرور الوقت. ومن المفترض أن يختار الأشخاص الذين لديهم دوافع ليكونوا متسقين نفس الاستجابة بعد التعرض لتدخل تجريبي كما كان الحال قبل التعرض لمثل هذا التدخل. غير أن شكل أو صيغة التناظر البصري يستبعد هذا الاحتمال. إذا كانت استجابات ما بعد المعالجة التجريبية قد خرجت بشكل ثابت (أي في نفس الاتجاه عادةً) عن الاستجابة السابقة قبل المعالجة بالنسبة للموضوعات التجريبية وبشكل عشوائي لعناصر الضبط والتحكم، فإن اختيار مقياس تناظر بصري قد يكون قد ساهم في الكشف عن ظاهرة دقيقة كانت أساليب أخرى تخفيها.

وغالبًا ما "ستخدم مقاييس التناظر البصري كمقاييس أحادية البند. وهذا له عيب كبير يتمثل في استبعاد أي تحديد للاتساق الداخلي. ومع المقياس أحادي البند، فإنه لا يمكن تحديد الثبات إلا من خلال استخدام طريقة إعادة الاختبار test-retest أو بالمقارنة مع مقاييس أخرى لنفس السمة لها خصائص سيكومترية راسخة. وتعاني الطريقة الأولى من مشكلات تقييمات إعادة الاختبار، ولا سيما استحالة التمييز بين عدم الاستقرار في عملية القياس وعدم استقرار الظاهرة التي يتم قياسها. أما الطريقة الأخيرة فهي في الواقع تقابل (تشبه) صدق المفهوم البنائي. ومع ذلك، نظرًا لأن الثبات شرط ضروري للصدق، يمكن للمرء أن يستنتج الثبات إذا كان هناك دليل على الصدق. ومع ذلك، قد تكون الاستراتيجية الأفضل هي تطوير عدة بنود تناظرية بصرية بحيث يمكن تحديد الاتساق الداخلي.

6- أشكال الاستجابة العددية والعمليات العصبية الأساسية:

تشير دراسة "زورزي، برفيتليس، وأوميليتا (Zorzi, Priftis, & Umiltà, 2002)" المنشورة في مجلة "Nature" إلى أن بعض أشكال الاستجابة العددية قد تعكس طريقة معالجة الدماغ للمعلومات الرقمية. فالأرقام المرتبة تسلسليًا، كما في مقياس "ليكرت"، لا تعبّر فقط عن قيم رقمية بل أيضًا عن مواقع مكانية على خط ذهني. ويقترح الباحثون أن تمثيل الأرقام بصريًا على شكل خط ليس مجرد أداة تنظيمية، بل يرتبط بآليات عصبية أساسية. وقد لاحظوا أن بعض المرضى الذين يعانون من إصابات دماغية تؤثر في الإدراك المكاني يرتكبون أخطاء منهجية في مسائل حسابية بسيطة تقدم بصريًا، حيث يميلون إلى الانحراف في إجاباتهم نحو اليمين، خصوصًا عند تحديد نقطة المنتصف بين قيم رقمية، مثل (3 و9). ويظهر هذا النمط عند العرض البصري فقط، بينما لا يظهر عند تقديم المسائل شفهيًا، كما أن الأفراد غير المصابين لا يظهرون هذا الانحراف.

ويخلص الباحثون إلى أن "خط الأعداد الذهني" قد يعكس أكثر من مجرد تمثيل مجازي، بل قد يستند إلى تنظيم عصبي فعلي للمعلومات العددية. وعلى الرغم من أن هذه النتائج لا تسمح باستنتاجات نهائية، فإنها تقدم أدلة أولية تشير إلى أن تمثيل الأرقام في شكل تسلسل خطي قد يتوافق مع عمليات عصبية مرتبطة بالتقدير الكمي، مما قد يفسر كفاءة استخدام التمثيلات الخطية للأرقام في بعض أشكال الاستجابة.

7- الخيارات الثنائية:

يوفر شكل وصيغة استجابة شائع آخر للأشخاص فرصة الاختيار بين الخيارات الثنائية لكل بند. وكانت الأمثلة السابقة لمقياس «ثرستون» Thurstone ومقياس "جتمان" Guttman تستخدم الخيارات الثنائية ("موافق" و"غير موافق")، على الرغم من أن المقاييس التي تحتوي على البنود متساوية الوزن قد تشتمل أيضاً على خيارات استجابة ثنائية. على سبيل المثال، قد يطلب من الأشخاص التأشير على جميع الصفات الموجودة في قائمة يعتقدون أنها تنطبق على أنفسهم، أو قد يُطلب منهم الإجابة بـ «نعم» أو «لا» على قائمة ردود الفعل العاطفية التي قد تكون لديهم في بعض المواقف المحددة. وفي كلتا الحالتين، يمكن جمع أو دمج الاستجابات التي تعكس بنوداً تشترك في متغير كامن مشترك، مثلاً الصفات ("حزين" و"غير سعيد" و"الكئيب" والتي تمثل الاكتئاب) في درجة واحدة لهذا المفهوم البنائي.

ويتمثل أحد أوجه القصور الرئيسية في الاستجابات الثنائية في أن كل بند لا يمكن أن يكون له سوى أقل قدر من التباين (القابلية للتغير). وبالمثل، فإن أي زوج من البنود يمكن أن يكون له مستوى واحد فقط من مستويين من التغاير (التباين المشترك): الاتفاق أو عدم الاتفاق. إن تباين المقياس الذي يتكون من عدة بنود متساوية الوزن أو الترجيح يساوي تماماً مجموع كافة المفردات في مصفوفة التغاير لكل بند على حدة. وفي حالة البنود ثنائية الاستجابة، فإن كل بند يساهم بقدر ضئيل في هذا المجموع بسبب القيود في التباينات والتغايرات (التباينات المشتركة) المحتملة. والنتيجة العملية لذلك هي وجود حاجة إلى المزيد من البنود للحصول على نفس درجة تباين المقياس إذا كانت البنود ثنائية الاستجابة. ومع ذلك، عادةً ما تكون الإجابة على البنود ثنائية الاستجابة سهلة للغاية. ولذلك، فإن العبء الملقى على عاتق الشخص يكون منخفضاً بالنسبة لأي بند من البنود. على سبيل المثال، يمكن لمعظم الناس أن يقرروا بسرعة ما إذا كانت بعض الصفات هي أوصاف مناسبة لأنفسهم. ونتيجة لذلك، غالباً ما يكون الأشخاص على استعداد لإكمال المزيد من البنود ثنائية الاستجابة أكثر من تلك البنود التي تستخدم شكلاً أو صيغة تتطلب التركيز على الفروق الدقيقة. وبالتالي، فإن الشكل ثنائي الاستجابة قد يسمح للباحث بتحقيق تباين كاف في درجات المقياس عن طريق تجميع المعلومات من المزيد من البنود.

8- الأطر الزمنية للبند: تتمثل مشكلة أخرى تتعلق بشكل وتنسيق البنود في الإطار الزمني الصريح أو الضمني. ويؤكد "كيلي ومكجراث (Kelly & McGrath, 1988)" على أهمية مراعاة السمات الزمنية للمقاييس المختلفة. فبعض المقاييس لا تتضمن إطارًا زمنيًا محددًا، مما يعكس منظورًا زمنيًا عامًا أو دائمًا. على سبيل المثال، تتضمن مقاييس موقع الضبط والتحكم بنودًا تعكس اعتقادات مستقرة نسبيًا حول العلاقات السببية، مثل: "إذا اتخذت الإجراءات الصحيحة، فبوسعي أن أبقى في صحة جيدة"(Wallston et al., 1978) ، وهو ما يتسق مع النظر إلى موضع التحكم باعتباره سمة عامة وليست مرتبطة بمواقف آنية، رغم وجود توجه لاحق نحو مزيد من التحديد في بعض المقاييس (DeVellis et al., 1985).

في المقابل، هناك مقاييس تستهدف ظواهر مؤقتة أو متغيرة عبر الزمن، مثل الاكتئاب، الذي قد يتقلب من فترة لأخرى، وقد أخذت بعض المقاييس هذا البعد الزمني في الاعتبار. فعلى سبيل المثال، يطلب مقياس الاكتئاب التابع لمركز الدراسات الوبائية (Radloff, 1977) من المستجيبين تحديد مدى تكرار خبرتهم بمشاعر معينة خلال الأسبوع الماضي. كما طورت بعض مقاييس القلق

(Spielberger, Gorsuch, & Lushene, 1970) بصيغ مختلفة لقياس كل من الحالات العابرة نسبيًا والسمات المستقرة نسبياً (Zuckerman, 1983).

وعليه، يجب على الباحث اختيار الإطار الزمني بما يتوافق مع طبيعة المتغير النظري، وبشكل واعٍ وليس عشوائيًا: هل الظاهرة تمثل سمة مستقرة ودائمة نسبيًا أم حالة ظرفية تتغير بسرعة؟ وهل الهدف هو رصد تغيرات قصيرة المدى كتغير المزاج بعد موقف معين أم تغيرات طويلة المدى تمتد عبر الحياة كالتحولات في الاتجاهات السياسية مع التقدم في العمر؟

وفي الختام، ينبغي أن تعكس أشكال وتنسيقات البنود، بما في ذلك خيارات الاستجابة والتعليمات، طبيعة المتغير الكامن المست هدف والاستخدامات المقصودة من المقياس.

الخطوة الرابعة: مراجعة تجمع البنود الأولية من قبل الخبراء:

بعد الانتهاء من تحديد الظاهرة بدقة، وصياغة البنود، واختيار شكل الاستجابة المناسب، تأتي مرحلة عرض تجمع البنود على خبراء مختصين في مجال المحتوى، بهدف تعزيز صدق المحتوى للمقياس. تسهم هذه المراجعة أولًا في التحقق من مدى دقة تمثيل البنود للمفهوم البنائي؛ إذ يطلب من الخبراء تقييم صلة كل بند وأهميته، خاصة في حالة المقاييس متعددة الأبعاد. وتمثل تصورات الباحث حول ما يقيسه كل بند فرضيات، بينما تمثل أحكام الخبراء بيانات لاختبارها. وحتى في المقاييس أحادية البعد، تظل هذه المراجعة ضرورية للكشف عن أي معانٍ غير مقصودة قد يستنتجها المستجيبون. عادةً ما يزود الخبراء بتعريف إجرائي للمفهوم البنائي، ويطلب منهم تصنيف البنود من حيث الأهمية أو الصلة (عالية/ متوسطة/ منخفضة)، مع إتاحة المجال للتعليق النوعي. وتعد هذه التعليقات ذات قيمة كبيرة، إذ قد تكشف عن غموض في الصياغة أو تقدم رؤى جديدة لتحسين القياس. كما يقوم المراجعون بتقييم وضوح البنود ودقتها، لأن الصياغة غير الواضحة قد تؤثر سلبًا على الثبات من خلال إدخال عوامل خارجية. لذلك، ينبغي طلب تحديد البنود المربكة واقتراح بدائل لها عند الإمكان. ومن الأدوار المهمة أيضًا للخبراء الإشارة إلى جوانب أو أبعاد لم يتناولها الباحث، مما يساعد على تحقيق شمولية أفضل للمقياس وزيادة صدق المحتوى.

ومع ذلك، يبقى القرار النهائي بيد الباحث، إذ قد لا يكون بعض الخبراء ملمين بأسس بناء المقاييس، وقد يقترحون مثلًا حذف البنود المتشابهة، رغم أن التكرار ضروري لتحقيق الاتساق الداخلي. لذا يجب التعامل مع هذه الآراء بوعي وتمييز. وفي نهاية هذه المرحلة، يكون لدى الباحث تجمع بنود منقح بناءً على مراجعة الخبراء، جاهز للانتقال إلى المرحلة التالية من تطوير المقياس.

الخطوة الخامسة: النظر في إدراج البنود الداعمة للصدق:

تمثل مجموعة البنود جوهر بناء المقياس، إلا أن إضافة بعض البنود المساعدة قد تسهم في فحص صدق المقياس مبكرًا. ويمكن التمييز بين نوعين رئيسيين:

أولًا، بنود الكشف عن المشكلات أو التحيزات في الاستجابة، مثل المرغوبية الاجتماعية (social desirability)، حيث قد يسعى بعض الأفراد لتقديم أنفسهم بصورة إيجابية، مما يشوه استجاباتهم. لذلك، يفيد إدراج مقياس مختصر لهذا البعد، مثل (Strahan & Gerbasi, 1972) في التعرف على البنود المتأثرة به واستبعادها عند الحاجة. كما يمكن الاستفادة من مقاييس أخرى للكشف عن أنماط الاستجابة غير المرغوب فيها، كما في اختبار (Minnesota Multiphasic Personality Inventory).

ثانيًا، بنود تتعلق بـ صدق المفهوم البنائي، حيث يمكن إدراج مقاييس لمتغيرات مرتبطة نظريًا بالمفهوم محل القياس. ويسمح ذلك بفحص العلاقات المتوقعة بين المقاييس منذ المراحل المبكرة، بدلًا من تأجيل التحقق من الصدق إلى ما بعد بناء المقياس. وبذلك، يتيح تضمين هذه البنود الحصول على مؤشرات مبكرة حول جودة المقياس، سواء من حيث دقة الاستجابات أو اتساقها مع الإطار النظري.

الخطوة السادسة: تطبيق البنود على عينة تطوير المقياس:

بعد تحديد البنود المرتبطة بالمفهوم البنائي وبالصدق وإدراجها في الاستبانة، يتم تطبيقها—إلى جانب البنود الجديدة—على عينة من الأفراد. ويفضل أن تكون هذه العينة كبيرة نسبيًا، رغم عدم وجود اتفاق حاسم حول حجمها. ويشير Nunnally (1978) إلى أن نحو 300 مفحوص قد يكون عددًا مناسبًا، مع إمكانية استخدام عينات أصغر تبعًا لعدد البنود وأهداف المقياس؛ فمثلًا، قد تكون عينة أقل من 300 كافية إذا كان الهدف بناء مقياس واحد من نحو 20 بندًا.

تكمن أهمية العينة الكبيرة في الحد من تأثير الصدفة على معاملات الارتباط وأنماط التغاير بين البنود. فالعينات الصغيرة قد تعطي صورة مضللة عن الاتساق الداخلي (مثل معامل "ألفا")، مما يؤدي إلى الاحتفاظ ببنود ضعيفة أو استبعاد بنود جيدة بسبب تقلبات عشوائية. وعند إعادة تطبيق المقياس لاحقًا، قد لا تستمر هذه النتائج، فتظهر معاملات ثبات أقل من المتوقع.

ومن جهة أخرى، يجب مراعاة مدى تمثيل العينة للمجتمع المستهدف، من حيث الحجم والتركيب. فقد يكون عدم التمثيل كميًا (اختلاف في مستوى السمة)، وهو ما يؤثر على المتوسطات دون أن يضر بالضرورة بثبات المقياس. أما المشكلة الأكثر خطورة فهي عدم التمثيل النوعي، حين

تختلف دلالات البنود أو العلاقات بينها لدى العينة مقارنة بالمجتمع، مما قد يؤدي إلى بنية عاملية غير دقيقة. وتبرز أهمية ذلك في السياقات اللغوية والثقافية؛ إذ قد تختلف معاني بعض الألفاظ بين المجموعات، مما يغير طريقة استجابة الأفراد. فمثلًا، قد تحمل كلمة واحدة دلالات مختلفة عبر الثقافات أو حتى داخل المجتمع نفسه، وهو ما ينعكس على نتائج التحليل.

لذلك، فإن اختيار عينة تطوير مناسبة—من حيث الحجم والتمثيل—يعد شرطًا أساسيًا لضمان دقة بنية المقياس وثباته، مع ضرورة توخي الحذر عند تفسير النتائج إذا وجدت مؤشرات على عدم تمثيل العينة بشكل كافٍ(ديفيليس، 2021، ص147-207).

الخطوة السابعة: تقييم البنود:

بعد إعداد البنود الأولية وتطبيقها على عينة مناسبة، تأتي مرحلة تقييم أداء كل بند لاختيار الأنسب لتكوين المقياس، وهي مرحلة محورية في عملية تطويره.

1- الفحص المبدئي لأداء البنود: تتمثل الخاصية الأساسية للبند الجيد في ارتفاع ارتباطه بالدرجة الحقيقية للمتغير الكامن، رغم أن هذه الدرجة لا يمكن قياسها مباشرة. لذلك، يستدل عليها من خلال العلاقات بين البنود وفق نماذج القياس. فكلما ارتفعت الارتباطات بين البنود، دل ذلك على ارتفاع ثباتها، وبالتالي زيادة ثبات المقياس ككل (بافتراض اشتراكها في نفس المتغير الكامن). ومن هنا، تعد قوة الترابط بين البنود مؤشرًا رئيسيًا على جودتها، ويمكن فحص ذلك من خلال تحليل مصفوفة الارتباط.

2- عكس تسجيل الدرجات: عند ظهور بنود ترتبط سلبًا ببنود أخرى، ينبغي النظر في إمكانية تصحيح درجاتها عكسيًا. فالبنود المصاغة في اتجاهات متعاكسة قد تؤدي أحيانًا—عن قصد أو دون قصد—إلى ارتباطات سلبية، خاصة عند دمج مجموعات بنود مختلفة كالسعادة والحزن ضمن مقياس واحد لمفهوم أعم كـ"التأثير".

فعلى سبيل المثال، العبارتان "أنا سعيد" و"أنا حزين" ترتبطان بالمفهوم نفسه، لكنهما متضادتان؛ لذا، إذا كان الهدف أن تعكس الدرجات المرتفعة مستوى أعلى من السعادة، فيجب منح قيمة مرتفعة للاستجابة لبند "سعيد" وقيمة منخفضة لبند "حزين"، أي إجراء عكس في تسجيل درجات الأخير.

وفي بعض الحالات، تصمَّم البنود منذ البداية بطريقة معكوسة، كأن تستخدم نفس أوصاف الاستجابة اللفظية لجميع البنود، بينما ترتب القيم الرقمية تصاعديًا أو تنازليًا وفق اتجاه البند.

1- أشعر بالحزن في كثير من الأحيان.

6	5	4	3	2	1
غير موافق بشدة	غير موافق إلى حد ما	غير موافق بشكل معتدل	موافق بشكل معتدل	موافق إلى حد ما	موافق بشدة

2- في كثير من الأوقات، أكون سعيداً.

1	2	3	4	5	6
غير موافق بشدة	غير موافق إلى حد ما	غير موافق بشكل معتدل	موافق بشكل معتدل	موافق إلى حد ما	موافق بشدة

قد يؤدي هذا الإجراء إلى إرباك المجيب، إذ قد يتجاهل الأفراد الصياغات اللفظية عندما يلاحظون تكرارها عبر البنود. لذلك، قد يكون من الأفضل تغيير ترتيب أوصاف الاستجابة كالانتقال من "غير موافق بشدة" إلى "موافق بشدة" من اليسار إلى اليمين في بعض البنود وعكسه في أخرى. وهناك بديل آخر يتمثل في تثبيت الأوصاف اللفظية والقيم الرقمية لجميع البنود، مع إجراء التعديل اللازم على بعض البنود أثناء ترميز البيانات. غير أن عكس الدرجات يدويًا أثناء الترميز يعد عملية مرهقة وعرضة للأخطاء، إذ يتطلب التعامل مع كل بند معكوس بشكل خاص لكل مفحوص. لذا، فإن الطريقة الأسهل والأدق هي إجراء هذا العكس إلكترونيًا بعد إدخال البيانات. وتتيح برامج الحاسب تنفيذ ذلك بسهولة باستخدام معادلة بسيطة. فمثلًا، إذا استخدم مقياس "ليكرت" من (1) إلى (7)، وكانت القيم الأعلى تدل على الموافقة، فإن البنود الإيجابية تسجل كما هي، بينما تعكس البنود السلبية بحيث تتحول الدرجة (7) إلى (1)، و(6) إلى (2)، وهكذا، لضمان أن تعكس الدرجات المرتفعة دائمًا الاتجاه نفسه للمقياس(ديفيليس، 2021، ص147-207).

الخطوة الثامنة: تحقيق الحد الأمثل لطول المقياس:

1- تأثير طول المقياس على الثبات: في هذه المرحلة، يمتلك الباحث مجموعة بنود ذات مستوى مقبول من الثبات. ويتحدد معامل "ألفا" بعاملين أساسيين: قوة التغاير بين البنود وعددها. فإضافة بنود متوسطة الارتباط مع بقية البنود ترفع قيمة "ألفا"، في حين أن حذفها يؤدي إلى خفضه. بوجه عام، تتميز المقاييس القصيرة بسهولة التطبيق وقلة العبء على المجيبين، بينما تتسم المقاييس الطويلة بارتفاع مستوى الثبات. لذلك، يواجه الباحث مفاضلة بين الإيجاز والثبات، حيث إن تعزيز أحدهما غالبًا يكون على حساب الآخر. ومع ذلك، لا يكون الاختصار ميزة إذا كان الثبات منخفضًا؛ إذ لا

قيمة لمقياس قصير لا يمكن تفسير نتائجه. لذا، تطرح المفاضلة فقط عندما يكون الثبات كافيًا، وعندها يمكن تقليل عدد البنود مع قبول انخفاض طفيف في الثبات لتحقيق مزيد من الإيجاز.

2- آثار استبعاد وإسقاط البنود السيئة: يتوقف تأثير حذف البنود "الضعيفة" على معامل "ألفا كرونباخ" على درجة ضعف هذه البنود وعدد بنود المقياس. فكلما كان عدد البنود أقل، كان تأثير حذف أو إضافة بند واحد على معامل "ألفا" أكبر. فعلى سبيل المثال، إذا كان لدينا أربعة بنود بمتوسط ارتباط متبادل (0,50) فإن معامل "ألفا" يبلغ (0,80). وعند تقليلها إلى ثلاثة بنود بنفس متوسط الارتباط، ينخفض "ألفا" إلى (0,75)، بينما يرتفع إلى (0,83) عند زيادتها إلى خمسة بنود. أما عند وجود 9 أو 10 أو 11 بندًا بنفس متوسط الارتباط (0,50)، فإن "ألفا" يكون تقريبًا (0,90، 0,91، 0,92)، ما يوضح أن زيادة عدد البنود تجعل التغير في "ألفا" أقل حدة.

ويلاحظ أن حذف بند ذي ارتباط منخفض نسبيًا مع بقية البنود يؤدي غالبًا إلى رفع "ألفا"، في حين أن الاحتفاظ ببند ذي ارتباط يساوي أو يفوق المتوسط يسهم في تحسينه. ولتوضيح ذلك، فإن مقياسًا مكونًا من أربعة بنود يحتاج إلى متوسط ارتباط (0,57) بين ثلاثة بنود ليحافظ على "ألفا" بقيمة (0,80). وبالتالي، إذا أدى حذف أحد البنود إلى خفض المتوسط دون (0,57)، فإن ذلك يضر بالثبات بدل تحسينه. أما في حالة مقياس مكون من عشرة بنود، فإن تحقيق "ألفا" (0,80) يتطلب متوسط ارتباط يقارب (0,29)، بينما يحتاج المقياس المكون من تسعة بنود إلى (0,31). وبالتالي، لا يستبعد البند العاشر إلا إذا كان متوسط ارتباطه مع بقية البنود لا يتجاوز (0,20)، وإلا فإن الاحتفاظ به يكون أفضل للمقياس(ديفيليس، 2021، ص147-207).

وتظهر هذه الحسابات أن البند "السيئ" لا يكفي أن يكون ضعيفًا فقط، بل يجب أن يكون أضعف بقدر ملحوظ مقارنة ببقية البنود حتى يكون حذفه مفيدًا فعليًا، ويقل هذا الفارق المطلوب كلما ازداد عدد البنود في المقياس.

3- ضبط وإصلاح طول المقياس: كيف يمكن عمليًا تحسين أو "ترقيع" طول المقياس؟ في البداية، تعد البنود الأقل إسهامًا في الاتساق الداخلي هي المرشحة الأولى للحذف، ويمكن تحديدها بعدة طرق. فبرامج التحليل الإحصائي مثل SPSS وSAS (PROC CORR) فيها خيار (ALPHA) خيار تظهر أثر حذف كل بند على معامل "ألفا"، حيث يُعد البند الذي يؤدي حذفه إلى أقل ضرر أو أكبر تحسن في "ألفا" هو المرشح الأول للاستبعاد. كما يمكن الاعتماد على ارتباط البند بالمقياس، بحيث تستبعد البنود ذات الارتباط الأضعف أولًا. ويوفر برنامج SPSS أيضًا الارتباط المتعدد التربيعي (squared

multiple correlation) لكل بند، الناتج عن انحداره على بقية البنود، وهو مؤشر على تباين البند المشترك (communality) مع باقي البنود. وغالبًا ما تتفق هذه المؤشرات؛ فالبند ذو الارتباط الضعيف بالمقياس يميل أيضًا إلى امتلاك ارتباط متعدد تربيعي منخفض، وقد يؤدي حذفه إلى تحسين أو عدم تغيير معامل "ألفا". ومن المهم الإشارة إلى أن طول المقياس يؤثر في دقة تقدير "ألفا"، إذ يميل معامل "ألفا" إلى الارتفاع مع زيادة عدد البنود، كما تصبح تقديراته أكثر استقرارًا وتقل فترات الثقة حوله، مما يجعل القيم أكثر اتساقًا مقارنة بالمقاييس القصيرة. لذلك يجب مراعاة هذه الحقيقة عند تحديد طول المقياس في مرحلة التطوير. وأخيرًا، ينبغي ترك هامش أمان في قيمة "ألفا"، لأن هذه القيمة قد تنخفض عند تطبيق المقياس على عينات أخرى غير عينة التطوير.

4- تقسيم العينات: إذا كانت عينة التطوير كبيرة بما يكفي، يمكن تقسيمها إلى عينتين فرعيتين، تستخدم إحداهما لتطوير المقياس، بينما تخصص الأخرى للتحقق من النتائج. فعلى سبيل المثال، تستخدم العينة الأولى لحساب معامل "ألفا كرونباخ"، وتقييم البنود، وضبط طول المقياس، وبناء صورته النهائية، ثم تستخدم العينة الثانية للتحقق من هذه النتائج دون إعادة اختيار البنود بناءً على بياناتها.

وبهذا، لا تتأثر نتائج العينة الثانية بتقلبات الصدفة أو التحيزات المرتبطة بعملية اختيار البنود، مما يسمح بتقدير أكثر استقرارًا لمعامل "ألفا كرونباخ" وغيره من المؤشرات. فإذا تكررت النتائج عبر العينتين الفرعيتين، أمكن الاطمئنان إلى أن القيم ليست مشوهة بعوامل عشوائية. كما أن العينات الفرعية، لكونها ناتجة عن تقسيم عشوائي لعينة واحدة، تكون أكثر تجانسًا من عينات مستقلة تمامًا، إذ تشترك في نفس الظروف الزمنية والبيئية والإجرائية لجمع البيانات، بما في ذلك خصائص التطبيق والأدوات المستخدمة. كما تتعرض المجموعتان لنفس مجموعة البنود الأصلية، بما في ذلك البنود التي قد يتم حذفها لاحقًا، مما يضمن قابلية المقارنة بينهما. ورغم هذا التشابه، فإن إعادة التحقق عبر تقسيم العينة يوفّر دليلًا مهمًا على استقرار المقياس؛ إذ تتيح العينة الأولى احتمال تأثر النتائج بعوامل الصدفة أثناء اختيار البنود، بينما تستخدم العينة الثانية لاختبار مدى ثبات هذه النتائج دون هذا التأثير. ويعد التقسيم الأبسط هو تقسيم العينة إلى نصفين، إلا أنه في حال صغر حجمها يمكن اعتماد تقسيم غير متساوٍ، بحيث تستخدم العينة الأكبر لتطوير المقياس، والأصغر للتحقق من صلاحيته (cross-validation) (ديفيليس، 2021، ص147-207).

- تمارين:

افترض أنك تقوم بتطوير مقياس "الخوف من الثعابين" الذي يحتوي على صيغ استجابة "ليكرت" مكون من ستة خيارات، وتم تطبيقه على 300 شخص. على الرغم من أن المزيد من البنود سيكون أمراً مرغوباً فيه لتطوير المقياس الفعلي، إلا أنه يمكنك إكمال الخطوات التالية لهذه التمارين:

1- قم بإنشاء تجمع من عشرة بنود على شكل مقياس "ليكرت".

2- بالنسبة لكل بند قمت بكتابته، قم بتقدير قيم مقياس "ليكرت" التي سيتم إقرارها من قبل "الشخص العادي" (أي، لا يوجد فوبيا من الثعابين أو ساحر الثعابين).

3- اختر بنداً من تجمع البنود تعتقد أنه قد يستخلص استجابة متطرفة من شخص عادي، ثم أعد صياغتها للحصول على استجابة أكثر اعتدالاً.

4- قم بتوليد وتكوين عشرة بنود أخرى من نوع مقياس "ليكرت" لقياس مفهوم بنائي آخر غير الخوف من الثعابين، ثم قم بمزج أو خلط هذه البنود عشوائياً مع البنود العشرة الأصلية واطلب من بعض الأصدقاء أن يشيروا إلى ما يعتقدون حول ما يهدف كل بند من البنود إلى قياسه.

5- ما هي إيجابيات وسلبيات بناء البنود التي يكون لها عدد زوجي مقابل البنود التي يكون لها عدد فردي من خيارات الاستجابة؟

6- باستخدام إما مفهوم الخوف من الثعابين أو المفهوم البنائي الأساسي الذي تقوم عليه المجموعة الثانية المكونة من عشرة بنود، قم بتدوين السلوكيات التي يمكن ملاحظتها والتي قد تعبر عن هذا المفهوم البنائي. بشكل مباشر والتي يمكن استخدامها للتحقق من صدق مقياس يقيس ذلك البناء واشرح كيف يمكنك استخدام البيانات السلوكية للتحقق من الصدق.

7- ما قيمة معامل "ألفا" بالنسبة للمقياس إذا كان متوسط الارتباط المتبادل بين البنود العشرة لمقياس الخوف من الثعابين يساوي (0,30)؟

8- كيف يمكنك استخدام العينات المقسمة لتقدير معامل "ألفا" للمقياس والتحقق من الصدق عبر العينات؟

ملاحظات:

1- الإجابة: ألفا = [10 * 0,30] / [1 + (9 * 0,30)] = 0,81

المحـاضرة السابعة

الخصائص السيكومترية للفقرات

الأهداف:

بعد دراسة موضوع الخصائص السيكومترية لفقرات الاختبار يتوقع من الطالب أن يكون قادراً على توضيح معنى كل من معامل الصعوبة ومعامل السهولة وحسابهما وتفسير نتائجهما، وأن يفسر أثر التخمين على صدق الفقرة ويطبق طرق تصحيحه، كما يصبح قادراً على حساب معامل التمييز وتحليل دلالاته في الحكم على جودة الفقرات. كذلك يتمكن الطالب من تحليل فعالية المموهات أو المشتتات في أسئلة الاختيار من متعدد، والتعرف على معامل ثبات المفردة وعلاقته بالثبات الكلي للاختبار، مما يؤهله لاستخدام هذه المؤشرات في تقويم جودة الاختبارات وبناء أدوات قياس أكثر دقة وعدالــة.

محتوى المحاضرة:

1- معامل صعوبة الفقرة Item Difficulty Index

2- معامل سهولة الفقرة Item Easiness Index

3- معامل تصحيح الصعوبة من أثر التخمين Guessing

4- قوة تمييز فقرات الاختبار Item Discrimination Index

5- طرق أخرى لتقييم تمييز فقرات الاختبار

6- المموهات (المشتتات) Distractions

7- معامل ثبات المفردة

مدة المحاضرة: ثلاث ساعات

تمهيد:

بعد تطبيق الاختبار وتصحيحه تأتي عملية تحليل فقراته للتحقق من مدى صلاحيته وفاعليته كأداة تقويمية، وذلك من خلال التعرف على درجة صعوبة فقراته أو سهولتها، ومعامل تمييزها، إضافة إلى الكشف عن أثر الاختبار في الطلبة. ولضمان ذلك ينبغي مراعاة مجموعة من الضوابط؛ أبرزها أن يخلو الاختبار من الأسئلة التي يجيب عنها جميع الطلبة أو يعجز عن الإجابة عنها الجميع، لأن مثل هذه الأسئلة لا تميز بين المستويات المختلفة ولا تكشف الفروق الفردية في التحصيل خاصة عند الحديث عن الاختبارات معيارية المرجع. كما يستلزم أن تتوزع العلامات بصورة اعتدالية أو شبه اعتدالية مع مراعاة خصائص الطلبة وأهداف الاختبار.

1- معامل صعوبة الفقرة (Item Difficulty Index):

يستخدم معامل صعوبة الفقرة لتحديد مستوى السهولة أو الصعوبة لكل فقرة من فقرات الاختبار، حيث يتم استبعاد الفقرات المتطرفة في السهولة أو الصعوبة الشديدة أو استبدالها بأخرى أكثر ملاءمة. ويرتبط معامل الصعوبة بمعامل السهولة ارتباطًا عكسيًا، إذ أن مجموعهما يساوي (1)، أي أن: معامل السهولة + معامل الصعوبة = 1 (للفقرة الواحدة).

ويتم اختيار الفقرات المناسبة من حيث مستوى الصعوبة بما يتوافق مع طبيعة الطلبة وأهداف الاختبار، ويعرف معامل الصعوبة بأنه نسبة الطلبة الذين أخطأوا في الإجابة عن الفقرة، ويعطى بالمعادلة التالية:

$$\textbf{معامل صعوبة الفقرة} = \frac{\text{عدد الاجابات الخاطئة على الفقرة}}{\text{عدد الاجابات الكلية على الفقرة}} \boldsymbol{X}\ \%\mathbf{100}$$

مثال: في اختبار مادة الرياضيات لطلبة السنة الأولى متوسط، تبين أن 20 طالبًا قد أجابوا إجابة خاطئة عن الفقرة الأولى من الاختبار، بينما بلغ إجمالي عدد الطلبة المفحوصين 30 طالباً.

- المطلوب هو حساب معامل صعوبة هذه الفقرة؟

$$\textbf{معامل صعوبة الفقرة} = \%\mathbf{100}\ \boldsymbol{X}\ \frac{\mathbf{20}}{\mathbf{30}} = \%\mathbf{67}$$

التفسير:

- معامل الصعوبة 0.67 = تقريبًا.
- هذا يشير إلى أن الفقرة صعبة نسبيًا، لأن حوالي 67% من الطلبة أخطأوا في الإجابة عنها.

ـ معامل السهولة المقابل:

معامل السهولة = 1 – 0.6667 = 0.3333.

أي أن حوالي 33% فقط من الطلبة أجابوا إجابة صحيحة، مما يعكس صعوبة الفقرة بالنسبة لمستوى الطلبة. يمكن استخدام هذا المعامل لتقرير ما إذا كانت الفقرة تحتاج إلى مراجعة، تعديل، أو استبدالها في نسخة الاختبار القادم.

أشارت الدراسات السيكومترية والتجريبية إلى أن الاختبار يمكن أن يميّز بين المفحوصين بأقصى درجة ممكنة عندما يكون متوسط مستوى صعوبة الفقرات حوالي 0.50 أي أن حوالي 50% من المفحوصين قادرون على الإجابة عن كل فقرة. لذلك يُفضَّل اختيار فقرات يكون مدى صعوبتها معتدلاً بحيث يكون متوسط صعوبة جميع الفقرات قريباً من (0.50) (صلاح علام، 2022). أما نيتكو (Nitko, 1983) يرى أن مدى معاملات الصعوبة الذي يقع بين القيم 0.20 و0.80 مع متوسط قدره 0.50 يعتبر مقبولاً ويُعطي الاختبار قدرة جيدة على التمييز بين مستويات التحصيل المختلفة للطلاب.

1-1- طريقة أخرى لحساب مؤشر صعوبة الفقرة:

يمكن تحديد صعوبة الفقرة أيضًا من خلال نسبة الطلبة الذين أجابوا عنها إجابة صحيحة وذلك باستخدام المجموعتين المتطرفتين من الطلاب أي الطلبة ذوي الأداء العالي والأداء المنخفض في الاختبار. تعتمد هذه الطريقة على مقارنة مدى تمكن الطلبة من الإجابة الصحيحة على الفقرة بين هاتين المجموعتين لتقدير صعوبة الفقرة بشكل أكثر دقة، بحيث يُظهر الاختبار قدرته على التمييز بين مستويات التحصيل المختلفة.

$$P = \frac{Ru + RL}{T} X100$$

حيث أن:

P = مؤشر صعوبة الفقرة

Ru = عدد أفراد المجموعة العليا الذين اجابوا على الفقرة إجابة صحيحة

RL = عدد أفراد المجموعة الدنيا الذين اجابوا على الفقرة إجابة صحيحة

T = عدد أفراد المجموعتين

مثال: يوضح الجدول أدناه فقرة اختيار من متعدد، وعليك دراسة بياناتها بدقة ثم حساب مؤشر صعوبة الفقرة بناءً على الإجابات الواردة.

جدول 03 معطيات لمثال لحساب مؤشر صعوبة الفقرة

البدائل	المجموعة العليا	المجموعة الدنيا
A	**0**	**0**
***B**	**8**	**4**
C	**3**	**5**
D	**4**	**6**
مجموع الطلبة	**15**	**15**

وتطبيق المعادلة السابقة:

$$P = \frac{8+4}{30} X100 = 0.4 = 40\%$$

(Sivakumar, et al, 2019, p.167)

2-1- معامل الصعوبة لفقرة إنشائية (مقالية):

يعد حساب معامل الصعوبة للفقرة الإنشائية خطوة أساسية لتقييم مدى مناسبة الفقرة لقدرات الطلاب وفاعليتها كأداة تقويمية، و يتم احتسابه باستخدام المعادلة التالية:

$$\boldsymbol{P = \frac{Gu + GL}{T\ x\ S}}$$

حيث أن:

$\boldsymbol{P}$ = مؤشر صعوبة الفقرة الانشائية

$\boldsymbol{Ru}$ = مجموع علامات الطلاب في الفئة العليا على الفقرة.

$\boldsymbol{RL}$ = مجموع علامات الطلاب في الفئة الدنيا على الفقرة.

$\boldsymbol{T}$ = عدد أفراد المجموعتين.

$\boldsymbol{S}$ = الدرجة الكلية المخصصة للفقرة.

- مثال على حساب معامل الصعوبة لفقرة إنشائية: لنفترض أن لدينا فقرة إنشائية تستحق 8 علامات، ونعلم أن:

- مجموع درجات الفئة العليا من الطلاب = 120
- مجموع درجات الفئة الدنيا من الطلاب = 40
- عدد الطلاب في كل فئة = 20 طالب

$$P = \frac{120 + 40}{40\ x\ 8} = 0.5$$

- التفسير:

- معامل الصعوبة يساوي 0.50 الفقرة متوسطة الصعوبة.

- هذا يعني أن الفقرة تمنح فرصة عادلة لكل الطلاب، حيث يتمكن حوالي نصف الطلاب من الحصول على درجات عالية فيها.

- اختيار فقرات ذات صعوبة متوسطة يضمن أن الاختبار قادر على تمييز مستويات التحصيل المختلفة بين الطلاب، ويُسهم في بناء اختبار عادل وفعّال.

- يمكن استخدام هذه الطريقة لمقارنة الفقرات المختلفة وتحديد أيها يحتاج إلى تعديل أو إعادة صياغة لضمان التوازن بين السهولة والتحدي.

2- معامل سهولة الفقرة (Item Easiness Index):

يشير معامل سهولة الفقرة إلى مدى سهولة الفقرة بالنسبة للطلاب، ويعبر عن النسبة المئوية للطلبة الذين تمكنوا من الإجابة عن الفقرة إجابة صحيحة. يساعد هذا المؤشر المعلم أو الباحث على تحديد الفقرات السهلة جدًا والفقرات الصعبة، وبالتالي ضبط مستوى الاختبار بما يتناسب مع قدرات الطلاب، و يمكن استخراج معامل سهولة الفقرة باستخدام المعادلة التالية:

$$\textbf{معامل سهولة الفقرة} = \frac{\text{عدد الاجابات الصحيحة على الفقرة}}{\text{عدد الاجابات الكلية على الفقرة}}\ \textbf{X 100}\%$$

وبالنظر إلى المثال السابق فإن معامل السهولة للفقرة الأولى يساوي:

$$\text{معامل سهولة الفقرة} = \%100\ X\ \frac{10}{30} = \%33$$

يلاحظ أن مجموع معاملي السهولة والصعوبة لأي فقرة في الاختبار يساوي دائمًا الواحد الصحيح .على سبيل المثال، إذا كان معامل الصعوبة = 0.67 ومعامل السهولة = 0.33، فإن:

$$\text{معامل الصعوبة} + \text{معامل السهولة} = 0.67 + 0.33 = 1$$

وبالتالي إذا عرف أحد المعاملين يمكن بسهولة معرفة الآخر ولا يحتاج الحكم على السؤال إلى كلا المعاملين معًا بل يكفي معرفة أحدهما. فعلى سبيل المثال إذا كان معامل السهولة = 0.60 فهذا يعني أن السؤال سهل بالنسبة لـ 60% من الطلاب وصعب بالنسبة لـ 40% منهم. ويواجه المعلم في تصميم الاختبار تحديًا في تحقيق مستوى الصعوبة المرغوب فيه لكل فقرة.

وكقاعدة عامة يفضل أن يتراوح معامل صعوبة الفقرات في الاختبار المرغوب فيه بين 0.50 و0.75 لضمان قدرة الفقرة على التمييز بين مستويات التحصيل المختلفة. كما يُنصح بترتيب الأسئلة بحيث توضع الفقرات السهلة في بداية الاختبار لتشجيع الطلاب ذوي التحصيل الضعيف على الإجابة ووضع الفقرات الصعبة في نهاية الاختبار لتحديد قدرة الطلاب الأقوياء على التعامل مع التحديات وإظهار مستواهم الحقيقي. هذا الترتيب يساهم في جعل الاختبار أكثر عدالـة وفاعلية ويحفز جميع الطلاب على الأداء الأمثل حسب قدراتهم.

3- معامل تصحيح الصعوبة من أثر التخمين (Guessing):

تتأثر فقرات الاختبار بعامل التخمين (Guessing) بشكل خاص عندما تكون من نوع الاختيار من متعدد إذ قد تكون إجابة الطالب صحيحة نتيجة الصدفة أو التخمين وليس نتيجة فهمه أو تحليله للمشكلة المطروحة. وتزداد خطورة هذا التأثير كلما قل عدد البدائل المتاحة في السؤال؛ ففي الاختبار الذي يتيح اختيار إجابة واحدة من خيارين يكون احتمال النجاح عن طريق التخمين 50% بينما إذا تضمن السؤال اختيار إجابة واحدة من ثلاثة بدائل يصبح احتمال النجاح بالتخمين 30%. ويمكن التعبير عن العلاقة بين عدد البدائل واحتمال النجاح بالتخمين بالصيغة الرياضية التالية:

$$\textbf{أفضل معامل الصعوبة في حال مراعاة التخمين} = \frac{(\frac{1}{\text{عدد البدائل}})+1}{2}$$

- أفضل معامل الصعوبة لفقرة اختيار من بديلين = $= \frac{(\frac{1}{2})+1}{2} = 0.75$
- أفضل معامل الصعوبة لفقرة اختيار من ثلاثة بدائل = $\frac{(\frac{1}{3})+1}{2} = 0.665$
- أفضل معامل الصعوبة لفقرة اختيار من خمسة بدائل = $\frac{(\frac{1}{5})+1}{2} = 0.6$
- أفضل معامل الصعوبة في حال مراعاة التخمين للفقرة الانشائية أو المقالية هو 0.5.

❖ بالنسبة للفقرة الانشائية أو المقالية لا يوجد لها معامل صعوبة مصحح.

ولتجنب هذا الخلل الناتج عن التخمين اقترح جيلفورد (Guilford, 1936) قانونًا لتصحيح صعوبة الفقرة من أثر التخمين. ووفق هذا الاقتراح تُعرّف صعوبة الفقرة إحصائيًا على أنها نسبة المفحوصين الذين أجابوا عن الفقرة إجابة صحيحة مع تعديل هذه النسبة لإزالة تأثير التخمين، ويتم تمثيل ذلك بالمعادلة التالية:

$$d_c = \frac{Ld - 1}{L - 1}$$

d_c = النسبة المصححة للاجابات الصحيحة على الفقرة.

d = النسبة المحصلة.

L = عدد بدائل الإجابة لكل فقرة.

في هذا المثال تم إعطاء سؤال لعينة من الأفراد وكانت نسبة الذين أجابوا عنه إجابة صحيحة 50% وإذا علمنا أن السؤال يحتوي على خمس إجابات اختيارية يمكننا حساب النسبة المصححة لأثر التخمين باستخدام المعادلة السابقة، فنجد أن:

$$d_c = \frac{5\ x\ 0.5 - 1}{5 - 1} = 0.37$$

-التفسير:

وبذلك نجد أن نسبة الإجابات الصحيحة بعد تصحيح أثر التخمين انخفضت لتصبح 37% بدلاً من 50% أي أن حوالي 13% من الإجابات الصحيحة كانت نتيجة التخمين أو الصدفة مما يوضح أهمية تعديل صعوبة الفقرة لتقديم تقييم أكثر دقة وموثوقية لمستوى الطلاب الفعلي.

3-1- درجة الكلية للاختبار المصححة من أثر التخمين:

يمكن حساب درجة الاختبار الكلية بعد تصحيح أثر التخمين لتقديم تقدير أدق لمستوى الطلاب الحقيقي بعيدًا عن تأثير الصدفة أو التخمين، وتُحسب باستخدام المعادلة التالية:

$$FS = R - \frac{W}{C - 1}$$

حيث أن:

FS = درجة الاختبار الكلية المصححة من أثر التخمين.

R = عدد الفقرات الذين أجابوا عليها المفحوصين إجابة صحيحة.

W = عدد الفقرات الذين أجابوا عليها المفحوصين إجابة خاطئة.

C = متوسط عدد بدائل الإجابة لكل فقرة.

لنفترض أن لدينا اختبارًا مكونًا من 50 فقرة وكل فقرة تحتوي على 5 بدائل للاختيار من بينها. عند تطبيق الاختبار على عينة من الطلاب تبين أن عدد الفقرات التي أجاب عنها الطلاب إجابة

صحيحة هو 25 فقرة بينما كانت 12 فقرة إجابات خاطئة، ولحساب الدرجة المصححة من أثر التخمين نطبق المعادلة التالية:

$$FS = 25 - \frac{12}{5-1} = 22$$

-التفسير:

هذا يعني أن جزءً من الإجابات الصحيحة للطلاب كان نتيجة التخمين أو الصدفة وليس نتيجة معرفتهم الفعلية بالمادة. بعد تطبيق تصحيح أثر التخمين انخفضت الدرجة من 25 فقرة صحيحة إلى 22 فقرة أي أن حوالي 3 فقرات من الإجابات الصحيحة كانت نتيجة الحظ. هذا التصحيح يوفر تقديرًا أدق لمستوى تحصيل الطلاب الحقيقي ويجعل الاختبار أكثر عدالـة وموثوقية في قياس معرفة الطلاب الفعلية.

4- قوة تمييز فقرات الاختبار (Item Discrimination Index):

يشير معامل تمييز الفقرة إلى قدرة الفقرة على التمييز بين الطلاب ذوي التحصيل العالي والطلاب ذوي التحصيل المنخفض في مجال معرفي معين، ويعد مؤشرًا على صدق الفقرة وفعاليتها كأداة تقييمية. فعلى سبيل المثال إذا كان معامل السهولة لسؤال معين 45%، فهذا يشير إلى أن السؤال قريب من المستوى المثالي (50%) من حيث الصعوبة ويبدو مناسبًا من هذه الناحية. لكن قبل الحكم على جودة السؤال من الضروري تحليل نوعية الطلاب الذين أجابوا عنه، فإذا كانت نسبة الإجابات الصحيحة متساوية تقريبًا بين فئة الممتازين والمتوسطين والضعفاء (15% لكل فئة) فهذا يدل على أن السؤال لم يقم بوظيفته الأساسية وهي التمييز بين الطلاب ذوي التحصيل الجيد وغيرهم. وبالتالي لضمان أن كل سؤال يفي بالغرض من التمييز بين مستويات التحصيل يُستخدم معامل تمييز الفقرة ويتم حسابه باستخدام المعادلة التالية:

$$D = \frac{Ru - RL}{\frac{1}{2}T}$$

حيث أن:

D = معامل تمييز الفقرة.

Ru = عدد الطلبة الذين أجابوا على الفقرة إجابة صحيحة في الفئة العليا.

RL = عدد الطلبة الذين أجابوا على الفقرة إجابة صحيحة في الفئة الدنيا.

$\frac{1}{2}T$ = عدد الطلبة في إحدى الفئتين.

عند حساب معامل تمييز الفقرات يتم أولاً تقسيم الطلاب وفق علاماتهم إلى مجموعتين متساويتين: المجموعة العليا والمجموعة الدنيا. لاستخراج معامل التمييز بدقة يجب الاعتماد على المجموعتين المتطرفتين من الطلاب سواء بأخذ المجموعتين بالكامل أو اختيار مجموعتين تمثلان الأطراف العليا والدنيا. ولتحقيق ذلك يُرتب الطلاب حسب علاماتهم ترتيبًا تنازليًا، ثم تُقسم أوراق الامتحان إلى مجموعتين متساويتين مجموعة الطلاب ذوي العلامات العليا ومجموعة الطلاب ذوي العلامات الدنيا.

فعلى سبيل المثال إذا كان عدد الطلاب في الصف 40 فتقسم المجموعة إلى مجموعتين كل منهما 20 طالبًا. أما إذا كان العدد فرديًا مثل 41 طالبًا فيتم استبعاد الطالب الوسيط (المرتبة 21) من التحليل لضمان توازن المجموعتين. وعند وجود عدد كبير من الطلاب يكفي أحيانًا أخذ ربع الأوراق (25% أو 27% أو 33%) من كل طرف من توزيع العلامات بعد ترتيبها تنازليًا أو تصاعديًا لضمان أن الحساب يعكس تمييز الفقرة بين الطلاب الأقوياء والضعفاء بشكل دقيق وموثوق. هذا الإجراء يضمن أن معامل التمييز يعكس قدرة الفقرة على التفريق بين مستويات التحصيل المختلفة مما يسهم في تحسين جودة الاختبار وموثوقيته.

4-1- حالة لحساب معامل التمييز لفقرة اختيار من بديلين:

لنأخذ مثالاً عمليًا: إذا كان عدد الطلاب في كل مجموعة متطرفة هو 12 طالبًا وأجاب عن الفقرة الأولى 7 طلاب من المجموعة العليا إجابة صحيحة بينما أجاب 3 طلاب من المجموعة الدنيا إجابة صحيحة، فإننا نكون قد حصلنا على البيانات اللازمة لحساب معامل تمييز الفقرة، ويتم حساب معامل التمييز وفق المعادلة التالية:

جدول رقم 04 معطيات لحساب معامل تمييز لفقرة اختيار من بديلين.

	1	2	3	4	5	6	7	8	9	10	11	12
الفئة العليا	أحمد	علي	محمود ✓	عمر	خالد ✓	رشيد ✓	كمال	إلياس ✓	إسحاق	تركي ✓	فتحي ✓	آدم ✓
الفئة الدنيا	سيف	شيث	شيماء	لين	جمانة ✓	زكريا	اياد	يارا ✓	أسماء ✓	حليمة	غنية	خضر

وبالتطبيق في الصيغة الرياضية التالية فإن معامل التمييز يساوي:

$$\text{معامل تمييز الفقرة} = \frac{3-7}{12} = 0.33$$

التفسير:

- قيمة معامل التمييز تساوي *D=0.33* تعني أن الفقرة تميز بشكل متوسط بين الطلاب ذوي التحصيل العالي والطلاب ذوي التحصيل المنخفض.
- أي أن الفقرة تمكنت من تمييز الطلاب الأقوياء عن الأقل قدرة، لكنها ليست فقرة ممتازة جدًا من حيث التمييز لأن القيم المثالية لمعامل التمييز عادة تكون قريبة من القيمة 1.
- هذا الحساب يساعد المعلمين والمصممين على تحديد جودة الفقرة وإعادة النظر في الفقرات التي لا تساهم بفاعلية في تمييز مستويات الطلاب المختلفة.

4-2- حالــة حساب معامل التمييز لفقرة اختيـار من متعدد مع أربع بدائل:

لنفترض أن لدينا فقرة اختيار من متعدد تحتوي على أربعة بدائل للإجابة والبديل الصحيح هو **(B)**. بعد جمع بيانات الأداء من المجموعتين المتطرفتين وجدنا أن 10 طلاب من المجموعة العليا أجابوا بشكل صحيح على الفقرة بينما أجاب 3 طلاب فقط من المجموعة الدنيا بشكل صحيح على الفقرة، ولحساب معامل تمييز الفقرة، نتبع الخطوات التالية:

جدول رقم 05 معطيات لحساب معامل تمييز لفقرة اختيار من متعدد مع أربع بدائل.

البدائل	المجموعة العليا	المجموعة الدنيا
A	**0**	**0**
***B**	**10**	**3**
C	**1**	**3**
D	**1**	**6**
المجموع	**12**	**12**

في هذا المثال تم حساب معامل التمييز للفقرة الاختيارية ذات أربعة بدائل كما يلي:

- عدد الطلاب في كل مجموعة = 12.
- عدد الطلبة في المجموعة العليا الذين اختاروا البديل الصحيح = 10.
- عدد الطلبة في المجموعة الدنيا اختاروا البديل الصحيح = 3.

ومنه يحسب معامل التمييز باستخدام الصيغة الرياضية الآتية:

$$\text{معامل تمييز الفقرة} = \frac{3-10}{12} = +0.58$$

التفسير:

- قيمة معامل التمييز تساوي D=0.58 تعني أن الفقرة تميّز بشكل جيد بين الطلاب ذوي التحصيل العالي والطلاب ذوي التحصيل المنخفض.

- كلما اقتربت قيمة معامل التمييز من 1 دل ذلك على أن الفقرة قادرة على التفريق بفاعلية أكبر بين مستويات التحصيل المختلفة.

- هذا يشير إلى أن الفقرة فعالة وموثوقة كأداة تقييم حيث تمكن الطلاب الأقوياء من الإجابة عنها بشكل صحيح أكثر من الطلاب الأقل قدرة.

وبمعنى آخر الفقرة تظهر قدرة عالية على التمييز بين مستويات الطلاب المختلفة مما يجعلها مناسبة للاختبارات التي تهدف إلى تقييم الأداء بشكل دقيق.

4-3- حالة حساب معامل التمييز لفقرة انشائية (مقالية):

عند التعامل مع الفقرات الإنشائية أو المقالية يختلف حساب معامل التمييز قليلًا عن فقرات الاختيار من متعدد نظرًا لأن النتائج تأتي على شكل درجات مستمرة وليست إجابات صحيحة أو خاطئة فقط. يتيح هذا النوع من الفقرات فرصة أكبر للتعبير عن عمق فهم الطالب ولكنه يتطلب أيضًا دقة أكبر في التقييم لأنه يمكن أن تتفاوت الدرجات بين الطلاب حتى ضمن نفس المجموعة. ولحساب قدرة الفقرة على التمييز بين الطلاب ذوي التحصيل العالي والطلاب ذوي التحصيل المنخفض نحتاج إلى مقارنة مجموع درجات الطلاب في المجموعتين المتطرفتين على الفقرة أي مجموعة الطلاب الأعلى تحصيلًا والأقل تحصيلًا بالنسبة للدرجة القصوى الممكنة للفقرة. هذا الأسلوب يعكس مدى فاعلية الفقرة في التمييز بين مستويات الأداء المختلفة ويحدد ما إذا كانت الفقرة تضيف قيمة حقيقية للاختبار أم لا. بعد هذا التوضيح نستخدم المعادلة التالية لحساب معامل تمييز الفقرة الإنشائية:

$$D = \frac{Ru - RL}{\frac{1}{2}\,T\,x\,S}$$

حيث أن:

D = معامل تمييز الفقرة المقالية أو الانشائية.

Ru = مجموع علامات الطلاب على الفقرة في الفئة العليا.

RL = مجموع علامات الطلاب على الفقرة في الفئة الدنيا.

$\frac{1}{2}\,T$ = عدد الطلبة في إحدى الفئتين.

S = العلامة المخصصة للفقرة أو السؤال المقالي أو الانشائي.

لنفترض أن لدينا سؤالًا مقاليًا مخصصًا له 5 علامات وقد تم جمع درجات الطلاب في المجموعتين المتطرفتين (العليا والدنيا) كما هو موضح في الجدول.

جدول رقم 06 معطيات لحساب معامـل تمييـز لفقرة مقاليــة.

	1	2	3	4	5	6	7	8	9	10	11	المجموع
الفئة العليا	3	4	2	2	0	2	4	5	4	3	2	**31**
الفئة الدنيا	2	2	1	1	3	3	1	2	2	1	1	**19**

ولكي نقوم بحساب معامل تمييز السؤال لمعرفة مدى قدرة الفقرة على التمييز بين الطلاب ذوي التحصيل العالي والطلاب ذوي التحصيل المنخفض، نستخدم المعادلة السابقة:

وبالتطبيق في المعادلة نجد التالي:

$$D = \frac{31 - 19}{11\ x\ 5} = 0.22$$

التفسير:

- قيمة معامل التمييز يساوي D=0.22 تشير إلى أن الفقرة تميز بشكل ضعيف إلى متوسط بين الطلاب ذوي التحصيل العالي والمنخفض.
- هذا يعني أن الفقرة لم تحقق التمييز الكافي بين مستويات التحصيل المختلفة وقد يحتاج المعلم إلى مراجعة صياغة السؤال أو تعديل طريقة تقييمه لضمان فاعليته في الاختبار.

بهذا الأسلوب يمكن استخدام معامل التمييز للفقرة المقالية لتحديد جودة السؤال وتحسين موثوقية الاختبار في قياس الأداء الحقيقي للطلاب.

5- طرق أخرى لتقييم تمييز فقرات الاختبار:

5-1- مؤشر التمييـز باستخدام معامل الارتبـاط الثنائـي النقطـي المتسلسل (Point Biserial Correlation):

إلى جانب طريقة حساب معامل التمييز باستخدام المجموعات الطرفية يمكن استخدام معامل الارتباط الثنائي النقطي المتسلسل (الحقيقي) كطريقة بديلة لتقدير معامل تمييز المفردات (Discrimination Index). تعتمد هذه الطريقة على قياس درجة العلاقة بين درجات كل فقرة في

الاختبار والتي تأخذ قيمة ثنائية (1 للإجابة الصحيحة، 0 للإجابة الخاطئة) والدرجات الكلية للطلاب في الاختبار. في هذا الأسلوب يتم تقسيم توزيع الدرجات الكلية إلى مجموعتين: مجموعة تمثل المستويات العليا ومجموعة تمثل المستويات الدنيا في السمة التي يقيسها الاختبار بحيث تكون نقطة التقسيم في منتصف التوزيع.

2-5- معامل الارتباط الثنائي المتسلسل (Biserial Correlation Coefficient):

يعتمد هذا المعامل على افتراض أن:

- المتغير الكامن (الدرجة الكلية في الاختبار) هو متصل وذو توزيع طبيعي أو اعتدالي.
- المتغير الثنائي الدرجة يمثل درجات كل مفردة أي إما الإجابة صحيحة (1) أو خاطئة (0).

باستخدام هذا الأسلوب يمكن تحديد مدى قدرة الفقرة على التمييز بين الطلاب ذوي التحصيل العالي والمنخفض بشكل أكثر دقة مما يعزز جودة الاختبار وموثوقيته كأداة تقويمية.

وقد قدم أيبل (Ebel, 1965) محكات احصائية لتحديد مستويات قوة المفردات التمييزية على أساس قيم التمييز الخاصة بها، وتظهر هذه القيم والتسميات المقابلة في الجدول الآتي:

جدول 07 يوضح المبادئ التوجيهية لـ(Ebel's, 1965) لتفسير قيم معامل التمييز.

مدى قيم معامل تمييز المفردة	التسمية
1.00 – 0.40	تعتبر الفقرة ذات تمييز عالي وممتاز
0.39 – 0.30	تعتبر الفقرة ذات تمييز جيد وهو الحد الأدنى المقبول
0.29 – 0.20	تعتبر الفقرة ذات تمييز جيد الى حد ما المراجعة ضرورية
0.19 -1.00-	الفقرة ضعيفة وينصح بحذفها

(Crocker & Algina, 2008, 315)

توفر هذه الطريقة طريقة كمية دقيقة لتقييم الفقرات بما يضمن اختيار أسئلة تساهم فعليًا في تمييز مستويات الطلاب المختلفة مما يزيد من موثوقية وجودة الاختبار كأداة تقويمية.

6- المموهات أو المشتتات (Distractions):

قد يكون السؤال ضعيفاً أو غير جيد بسبب خلل في المموهات (المشتتات). فالمموه يعدّ عنصراً أساسياً في فقرات الاختيار من متعدد إذ إن وظيفته الأساسية هي تشتيت انتباه المفحوصين عن الإجابة الصحيحة ودفعهم للتفكير النقدي قبل اتخاذ القرار. وعندما تكون هذه المموهات غير فعّالة أو صيغت

بطريقة ركيكة فإنها لا تؤدي الدور المطلوب منها مما يجعل المفحوص يصل إلى الإجابة الصحيحة بسرعة وسهولة أكبر مقارنة بما لو كانت المموهات قوية وجاذبة.

وتزداد جودة السؤال بقدر ما تكون المموهات واقعية ومقنعة أي قادرة على جذب انتباه بعض المفحوصين الذين لم يتمكنوا من إتقان المحتوى بشكل كامل بحيث تظهر قدرتهم الحقيقية على التمييز بين الإجابة الصحيحة والخاطئة. لذلك يُعتبر تحليل فاعلية المموهات خطوة ضرورية في بناء الاختبار وتقويم مفرداته، ويمكن تقدير هذه الفاعلية بالاعتماد على القانون الآتي:

$$DA = \frac{Pu - Pi}{n}$$

حيث أن:

DA = فاعلية المموه.

Pu = عدد أفراد المجموعة العليا الذين اختاروا ذلك البديل أو المموه.

Pi = عدد أفراد المجموعة الدنيا الذين اختاروا ذلك البديل أو المموه.

n = عدد أفراد المجموعة الواحدة.

وعليه فإن على مصمم الاختبار أن يتأكد من أن كل عمود أو مشتت يحقق الشرطين الآتيين:

- أن يكون المموه جذاباً ومغرياً للطلبة بحيث يختاره بعضهم.
- أن يكون عدد الذين جذبهم المموه في المجموعة الدنيا أكبر منه في المجموعة العليا وعلى الفاحص إجراء التعديلات في الفقرات الاختبارية حتى تصل إلى فقرات اختبارية أكثر ضبطاً وقدرة على تحقيق الأهداف المنشودة.

مثال تطبيقي: لإيضاح كيفية استخراج فاعلية المموهات في فقرات الاختيار من متعدد نفترض وجود فقرة ضمن اختبار معياري المرجع طبق على عينة مكوّنة من 30 طالباً.

ولإجراء التحليل يتم إعداد جدول يضم البيانات الأساسية لكل بديل من بدائل الفقرة، بحيث يظهر:

- الطلاب في المجموعة العليا الذين اختاروا كل بديل.
- عدد الطلاب في المجموعة الدنيا الذين اختاروا كل بديل.
- عدد الطلاب في المجموعة الوسطى الذين اختاروا كل بديل.
- المجموع الكلي لاختيارات كل بديل عبر جميع الطلاب.

هذا الجدول يمثل نقطة الانطلاق في تحليل فقرات الاختيار من متعدد، ومن خلاله يمكن حساب المؤشرات الإحصائية الخاصة بفاعلية المموهات وكفاءتها.

جدول رقم 08 معطيات لحساب معامل فعالية المموهات (المشتتات).

البدائل	المجموعة العليا	المجموعة الدنيا	الإجابات الخاطئة
*A	9	3	-
B	3	4	7
C	2	5	7
D	1	3	4
المجموع	15	15	18

وبتطبيق معادلة حساب فاعلية المموهات السابقة، نجد أن المموهات الثلاثة:

- م ت (B) $= \frac{4-3}{15} = 0.07-$
- م ت (C) $= \frac{5-2}{15} = 0.20-$
- م ت (D) $= \frac{3-1}{15} = 0.13-$

وبما أن معاملات التمييز الخاصة بالبدائل الثلاثة جاءت سالبة فإن ذلك يشير إلى أن عدد الأفراد الذين اختاروا هذه البدائل من المجموعة الدنيا يفوق عدد الذين اختاروها من المجموعة العليا. وهذه النتيجة منطقية ومطلوبة في حالة البدائل الخاطئة إذ يفترض أن تكون أكثر جاذبية للطلبة ضعاف التحصيل وأقل جاذبية للطلبة المتفوقين. وعليه يمكن الاستنتاج بأن البدائل الثلاثة فاعلة وتسير في الاتجاه الصحيح حيث أدت الدور المتوقع منها في تشتيت انتباه الطلبة غير المتمكنين عن الإجابة الصحيحة وبذلك أسهمت في تعزيز جودة الفقرة الاختبارية.

وبالمقابل إذا كانت معاملات التمييز موجبة بالنسبة للبدائل الخاطئة فإن هذا يعني أن عدد الأفراد من المجموعة العليا الذين اختاروا المموه أكثر من عدد الأفراد من المجموعة الدنيا وهو أمر غير مرغوب فيه على الإطلاق. ذلك أن المموه في هذه الحالة لا يقوم بوظيفته بل يصبح مصدر تشويش للطلبة المتفوقين الذين يفترض أن يتجهوا نحو الإجابة الصحيحة. هذا الوضع يدل على أن البديل الخاطئ غير فعّال بل قد يضعف من صدق الفقرة ويمسّ بجودتها، مما يستلزم إعادة صياغته أو استبداله بمموه أكثر ملاءمة وواقعية بحيث يجذب الطلبة ضعاف التحصيل بدلاً من المتميزين.

❖ **معايير صياغـــة المموهـات الجيدة:**

أ- الواقعية والمعقولية: يجب أن تكون المموهات مقنعة وقريبة من الصحة بحيث يختارها الطالب غير المتمكن، وليست بدائل غريبة أو بعيدة جداً عن موضوع السؤال.

ب- التجانس مع الإجابة الصحيحة: على المموهات أن تكون من نفس طبيعة الإجابة الصحيحة، مثلاً: جميعها أرقام، أو جميعها أسماء، أو جميعها مصطلحات علمية، حتى لا يسهل استبعادها بمجرد الشكل أو النوع.

ج- الاعتماد على أخطاء شائعة: أفضل المموهات هي تلك التي تعكس الأخطاء التي يقع فيها الطلاب عادة، لأنها تزيد من قدرتها على جذب ضعاف التحصيل.

د- توازن الطول والشكل: لا ينبغي أن تكون المموهات قصيرة جداً والإجابة الصحيحة طويلة أو العكس، لأن ذلك قد يكشف الصحيح بشكل غير مباشر.

هـ- الانسجام اللغوي: صياغة البدائل يجب أن تكون متشابهة في الأسلوب واللغة، حتى لا يظهر أحدها مميزاً أو مختلفاً فيجذب انتباه المفحوصين.

و- الابتعاد عن التعميم أو الغموض المفرط: المموهات الغامضة جداً قد تربك حتى الطلاب المتميزين، بينما المطلوب هو أن تخدع فقط الطلبة الأقل إتقاناً.

ز- عدم ترك بدائل مهملة (غير وظيفية): يجب التأكد بعد التجريب أن كل مموه يختار من قِبل نسبة مقبولة من الطلبة ($\leq 5\%$)، وإلا فإنه يعتبر غير وظيفي ويضعف من جودة الفقرة.

7- معامل ثبات الفقرة (Item Reliability):

يُقصد به مدى اتساق الفقرة في قياس ما وضعت لقياسه عبر مختلف الأفراد، وهو مؤشر على جودة الفقرة وأثرها في ثبات الاختبار ككل. ويتم تحديد معامل ثبات الفقرة بالاعتماد على كلٍّ من معامل التمييز ومعامل الصعوبة، وذلك وفق المعادلة التالية (Thayn, 2011):

$$IR_i = D\sqrt{pq}$$

حيث أن:

IR_i = معامل ثبات الفقرة.

D = معامل تمييز الفقرة.

p = معامل صعوبة الفقرة.

q = معامل سهولة الفقرة

ويعكس معامل ثبات الفقرة مقدار مساهمة الفقرة الواحدة في تعزيز ثبات الاختبار؛ فكلما ارتفعت قيمة هذا المعامل دل ذلك على أن الفقرة قادرة على قياس السمة المستهدفة بدرجة أكبر من الاتساق. ويُعد ثبات الاختبار الكلي انعكاساً مباشراً لمستوى ثبات فقراته؛ إذ إن زيادة اتساق وثبات الفقرات يؤدي بالضرورة إلى ارتفاع مستوى الثبات الكلي للاختبار.

ومن بين الطرق المستخدمة لتقدير ثبات الفقرة طريقة الاحتمال المنوالي (Mode Probability Method)، والتي تقوم على حساب احتمالية اختيار الإجابة الصحيحة بوصفها القيمة الأكثر شيوعاً (المنوال)، ثم دمج هذه الاحتمالية مع مؤشري التمييز والصعوبة للحصول على تقدير لمعامل ثبات الفقرة.

أما عن المحكات الخاصة بقبول قيمة معامل ثبات الفقرة، فقد أشار Thayn (2011) إلى أن الحد الأدنى المقبول هو (0.10)، بينما تُعد القيمة (0.15) أو أكثر مؤشراً مفضلاً على أن الفقرة تتمتع بدرجة جيدة من الثبات.

الدكتور زيـاد رشيـد

المحـــاضرة الثامنة

التحقق السيكومتري من أدوات القياس

الأهداف:

يهدف هذا الفصل إلى إلقاء الضوء على خصائص الاختبارات والمقاييس الجيدة مع بيان الشروط العامة التي ينبغي أن تتوافر في الاختبار النفسي والتربوي لضمان تحقيق أهدافه. كما يتناول الفصل بالشرح الشروط التجريبية أو السيكومترية الأساسية وفي مقدمتها الصدق (Validity) بمختلف أنواعه والعوامل المؤثرة فيه، إلى جانب عرض مفهوم الثبات (Reliability) وتوضيح العوامل التي قد تؤثر على درجة ثبات الاختبار. ومن خلال ذلك يسعى الفصل إلى تزويد القارئ بالأسس النظرية والتطبيقية التي تجعل من الاختبار أداة علمية دقيقة وفعّالة في القيـاس النفسـي والتربوي.

المحتوى:

1- خصائص الاختبارات والمقاييس الجيدة

2- الشروط العامة للاختبـار النفسـي والتربوي

3- الشروط التجريبيـــة أو السيكومتريـــة (Psychometric Conditions)

3-1- الصدق (Validity)

3-2- العوامل المؤثرة على الصدق

3-2- الثبـات (Reliability)

3-3- العوامل المؤثرة على ثبات الاختبار

مدة المحاضرة: ثلاث ساعات

تمهيد:

تشكل أدوات القيـاس والتقويـم التربـوي والنفسـي أحد الأعمدة الجوهرية في البحث العلمي، إذ تمثّل الوسيلة التي تُترجم الظواهر والسمات النفسية والتربوية إلى بيانات كمية قابلة للتحليل. ونظراً لهذه المكانة المحوري، فإن تصميم هذه الأدوات وإعدادها يتطلب التزاماً صارماً بالمعايير السيكومترية بما يضمن تحقيقها لخصائص الصدق (Validity) والثبات (Reliability) والموضوعية (Objectivity). فالأداة التي تفتقر إلى الصدق – أي التي لا تقيس البُعد أو السمة المستهدفة – تُعد أداة غير صالحة ويؤدي استخدامها إلى نتائج مشوَّهة تفتقر إلى القيمة العلمية مما قد يترتب عليه اتخاذ قرارات غير سليمة في مجالات الانتقاء أو التصنيف أو التشخيص أو العلاج. وبالمثل فإن الأداة التي تفتقر إلى الثبات – أي التي لا تُنتج قياسات متسقة عند إعادة التطبيق في ظروف متماثلة – تفقد قيمتها كمصدر للمعلومات الموثوقة. إضافة إلى ذلك يجب أن تكون أداة القياس خالية من التحيز (Bias-Free) بحيث لا تتأثر نتائجها بعوامل غير مرتبطة بالسمة المقاسة وأن تتسم بدرجة عالية من الحساسية (Sensitivity) لالتقاط الفروق الحقيقية بين الأفراد أو الجماعات. إن ضمان توافر هذه الخصائص يتيح للباحثين والممارسين استخدام نتائج القياس بثقة عالية واتخاذ قرارات مبنية على بيانات دقيقة مما يسهم في تحسين جودة البحث التربوي والنفسي ويعزز من العدالة والموضوعية في تقييم الأفراد وتوجيههم.

وبناءً على ما سبق فإن الإلمام بالخصائص السيكومترية للاختبارات والمقاييس يعد خطوة أساسية لكل باحث أو ممارس في المجال التربوي أو النفسي؛ إذ تشكّل هذه الخصائص المعيار الذي يُحكم من خلاله على جودة الأداة وفاعليتها وفقاً للمعايير الدولية كإرشادات اللجنة الدولية للاختبارات (ITC Guidelines) ومعايير الجمعية الأمريكية لعلم النفس (APA). وفي هذا الفصل سنتناول بالتفصيل أهم هذه الخصائص وآليات تقديرها وأهم المؤشرات الإحصائية الدالة عليها مع بيان تطبيقاتها العملية في بناء وتقويم الاختبارات والمقاييس.

1- خصائص الاختبارات والمقاييس الجيدة:

تعد الاختبارات والمقاييس النفسية والتربوية أدوات محورية في جمع البيانات واتخاذ القرارات العلمية المبنية على أسس كمية. وعند استخدام أي اختبار بغرض الحصول على معلومات دقيقة لدعم قرارٍ ما تبرز أمام الباحث أو الممارس مشكلة أساسية تتعلق باختيار الأداة الأكثر ملائمة. فقد تتوفر عدة بدائل من الاختبارات القابلة للاستخدام لكن السؤال الجوهري يظل قائمًا: أي هذه الاختبارات هو

الأفضل من حيث قدرتها على تقديم معلومات صادقة وموثوقة وملائمة لسياق الاستخدام؟ وما الأسس العلمية التي ينبغي أن نعتمد عليها في عملية الاختيار؟

ووفقًا للمعايير الدولية للجمعية الأمريكية لعلم النفس APA Standards for Educational and Psychological Testing وإرشادات اللجنة الدولية للاختبارات ITC Guidelines يمكن تصنيف الاعتبارات الأساسية لجودة الاختبار في ثلاث مجموعات رئيسية، وهي:

2- الشروط العامة للاختبـار النفسـي والتربوي:

هذه الشروط ترتبط بالتصميم الشكلـي والإجرائـي للأداة وتشمل:

أ- الشمول (Comprehensiveness): تمثيل جميع مكوّنات السمة أو القدرة المستهدفة بحيث يغطي الاختبار مجال المحتوى بشكل متوازن يضمن دقة القياس.

ب- التقنين (Standardization): توحيد جميع إجراءات التطبيق والتصحيح والتفسير لضمان ثبات المعاملة بين جميع المفحوصين بغض النظر عن اختلاف القائمين بالتطبيق أو ظروف الاختبار.

ج- الموضوعية (Objectivity): ضمان خلو عملية التطبيق والتصحيح من التأثيرات الذاتية أو التحيز بحيث تعتمد النتائج على معايير واضحة وقابلة للتحقق من قبل الآخرين.

3- الشروط التجريبيــة أو السيكومتريــة (Psychometric Conditions):

وهي الخصـائص الإحصائية التي تحدد دقة القياس وفاعليته:

أ- الثبـات (Reliability): درجة اتساق القياسات التي يقدمها الاختبار عبر الزمن أو بين أجزائه المختلفة أو عند اختلاف القائمين بالتطبيق.

ب- الصدق (Validity): مدى قدرة الاختبار على قياس البُعد أو الظاهرة التي وضع لقياسها ويشمل ذلك أنواع الصدق المختلفة (المحتوى، المحك، البنائي).

ج- المعايير (Norms): القيم المرجعية التي تسمح بتفسير الدرجات الفردية من خلال مقارنتها بأداء مجموعة معيارية ممثلة.

د- الاعتبارات العملية وقابلية الاستخدام (Practicality & Usability):

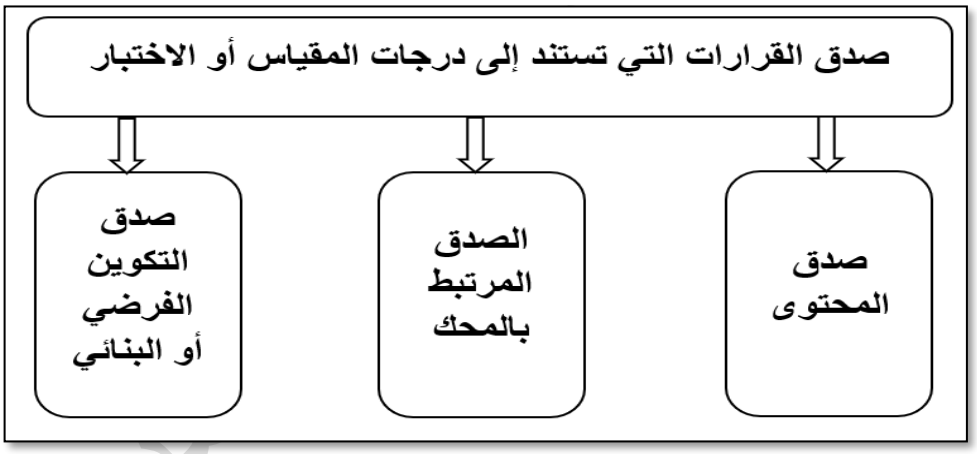
وتتعلق بالجوانب التطبيقية، مثل: وضوح التعليمات، سهولة التطبيق، تكلفة الوقت والموارد، ملاءمة الاختبار للثقافة المحلية، وإمكانية استخدامه مع فئات مختلفة من المفحوصين.

3- الشروط التجريبيـة أو السيكومتريـة (Psychometric Conditions):

3-1- الصدق (Validity):

يعد الصدق من أهم الخصائص السيكومتريـة لأي اختبار نفسي أو تربوي إذ يمثل الدرجة التي يتمكن فيها الاختبار من قياس السمة أو القدرة أو البعد الذي وُضع لقياسه بدقة دون أن يتأثر بعوامل خارجية أو غير ذات صلة. ووفقًا لمعايير (APA) وإرشادات اللجنة الدولية للاختبارات (ITC) يُعد الصدق جوهر جودة القياس حيث يحدد مدى صلاحية التفسيرات والقرارات المستندة إلى نتائج الاختبار. ويقاس صدق المفردة على وجه الخصوص بمدى ارتباطها المباشر بالهدف الذي صممت لقياسه بحيث تعكس كل مفردة جانبًا محددًا من المحتوى أو السمة المستهدفة. ومن الضروري التحقق من صدق الاختبار في صورته النهائية قبل اعتماده وذلك من خلال تطبيقه تجريبيًا على عينات متنوعة وفي مواقف متعددة وجمع بيانات كافية حول الخصائص التي يقيسها أو النتائج التي يتنبأ بها. وللصدق عدة أنواع إلا أنه يمكن تصنيفها في ثلاثة أنواع رئيسية:

شكل 08 يوضح أنواع الصدق الثلاثـة.



3-1-1- صدق المحكمين أو الخبراء (Judgmental Validity):

يعد صدق المحكمين أو الخبراء أحد الأساليب النوعية الأساسية للتحقق من مدى ملاءمة الاختبار لقياس السمة أو المجال المستهدف. ويعتمد هذا النوع من الصدق على التقديرات المتخصصة التي يقدمها خبراء في المجال من خلال فحص محتوى الاختبار وصياغة مفرداته ومدى اتساقها مع الأهداف المرجوة. ووفقًا لمعايير APA Standards for Educational and Psychological Testing وإرشادات اللجنة الدولية للاختبارات (ITC Guidelines) تعد عملية تحكيم الخبراء خطوة محورية

في مراحل البناء الأولية للاختبار وتُسهم في تعزيز صدق المحتوى وضمان ملاءمته للسياق الثقافي والتربوي للمفحوصين (زياد، 2023، ص124).

3-1-2- الصدق الظاهري (Face Validity): الصدق الظاهري يشير إلى الانطباع العام الذي يعطيه الاختبار للمفحوص أو للملاحظ الخارجي حول ما إذا كان يقيس فعلاً ما صمم لقياسه. ويتمثل ذلك في المظهر الخارجي للاختبار من حيث نوع المفردات وصياغتها ووضوحها ودقة تعليماته ومدى موضوعيتها بالإضافة إلى ملاءمة زمن الاختبار للاختبارات الموقوتة (التي تعتمد على السرعة) أو ملاءمة مستويات الصعوبة للاختبارات غير الموقوتة (التي تعتمد على القوة). فعلى سبيل المثال إذا احتوى اختبار للرياضيات في الصف الأول الابتدائي على جميع مفرداته من محتوى المنهج المقرر دون أن يتطلب مهارات لغوية متقدمة فإنه يعد ذا صدق ظاهري جيد. ومع ذلك فإن هذا النوع من الصدق يعتبر أضعف أشكال الصدق من الناحية العلمية لأنه لا يتجاوز المظهر السطحي للأداة ولا يوفر دليلاً إحصائيًا على جودة القياس.

3-1-3- صدق المحتوى (Content Validity): صدق المحتوى يعنى بدرجة تمثيل مفردات الاختبار لكامل المجال أو البعد المستهدف بالقياس بحيث يمكن اعتبار الاختبار عينة ممثلة لمحتوى المادة أو المهارة. فعلى سبيل المثال تستخدم درجات الطالب في اختبار مادة العلوم كدليل على مستوى معرفته في جميع جوانب المادة شريطة أن تكون مفردات الاختبار شاملة لمختلف موضوعاتها. ولضمان صدق المحتوى توصي المعايير الدولية بخطوات منهجية تشمل الآتي:

أ- تحديد مجال المحتوى المراد قياسه بدقة؛

ب- صياغة قائمة تفصيلية بالأهداف التعليمية أو المهارية المطلوب قياسها؛

ج- إعداد مفردات الاختبار بما يتوافق مع الأهداف المحددة مع مراعاة تنوع مستويات الصعوبة.

د- عرض المحتوى والأهداف ومفردات الاختبار على لجنة من المحكمين (الخبراء) في المجال للحصول على آرائهم بشأن مدى اتساقها وشموليتها وتعديل الأداة بناءً على ملاحظاتهم.

تعد هذه العملية من أكثر الإجراءات أهمية في ضمان أن تكون نتائج الاختبار انعكاسًا صادقًا للسمة أو القدرة المستهدفة وتسهم في تعزيز موثوقية القرارات المستندة إليها.

3-1-4- الصدق البيئي (Ecological Validity): الصدق البيئي يعد شكلًا خاصًا من صدق المحتوى ويختلف عن المغالطة البيئية أو مغالطة الاستدلال البيئي (Ecological Fallacy)[3] التي تناقش ضمن الصدق الإحصائي. يرتبط هذا المفهوم بمدى دراسة الأفراد في بيئاتهم الطبيعية مقارنةً بالبيئات المختبرية. فقد يؤدي إخراج الأفراد من بيئتهم المعتادة إلى ظهور سلوكيات تختلف عن سلوكياتهم "الحقيقية". فعلى سبيل المثال قد يظهر الأشخاص آراءً مختلفة حول قضايا اجتماعية عند تواجدهم في المختبر مقارنةً بتعبيرهم عن تلك الآراء في محيط أصدقائهم. وكلما زادت الفجوة بين البيئة الطبيعية وبيئة القياس زاد احتمال وقوع تحيز في النتائج. ولمواجهة تحديات الصدق البيئي يمكن للباحثين اتباع أربع استراتيجيات رئيسية:

أ- إظهار أن العوامل البيئية لا تؤثر على النتائج.

ب- إجراء القياسات داخل البيئة الطبيعية مع مراعاة صعوبة التحكم في المتغيرات الخارجية.

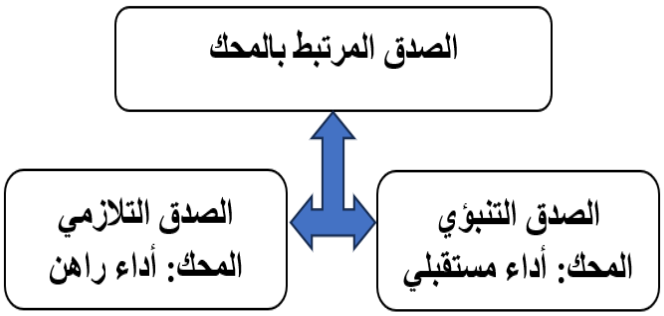
ج- محاكاة العوامل البيئية في بيئة القياس كاستخدام أسئلة تمهيدية في المقاييس.

د- الاعتراف بأن السلوك في المختبر لا يعكس بالضرورة السلوك في البيئة الطبيعية مع تبرير استخدام هذا الأسلوب لفهم السلوك بعيدًا عن المؤثرات البيئية (Garson, 2013, p.23).

3-1-5- الصدق المرتبط بالمحك (Criterion-Related Validity): الصدق المرتبط بالمحك يشير إلى درجـة ارتباط نتائج الاختبار بمحك خارجـي (External Criterion) يمثل السلوك أو الأداء الذي يُراد التنبؤ به أو التحقق منه. ويستخدم هذا النوع من الصدق للتأكد من قدرة الأداة على التنبؤ بأداء مستقبلي أو تفسير أداء حالي في مجال معين. ويعتمد في تقديره غالبًا على حساب معامل الارتباط بيرسون (Pearson Correlation Coefficient) بين نتائج الاختبار ونتائج المحك المعتمد. ووفقًا للمعايير الدولية في القياس النفسي (APA Standards) و (ITC Guidelines) يعد اختيار المحك المناسب ذا أهمية حاسمة، إذ يجب أن يكون المحك نفسه ذا صدق وثبات مرتفعين وأن يتوافق في مضمونه وزمن قياسه مع الهدف من الاختبار. وينقسم هذا النوع من الصدق إلى صدق تنبؤي وصدق تلازمي.

[3] - إذا كانت دراسة تظهر أن الدول التي لديها معدل تعليم عالٍ لديها أيضًا معدل دخل مرتفع، لا يعني ذلك بالضرورة أن كل فرد متعلم في هذه الدول لديه دخل مرتفع.

شكل 09 يوضح أنواع الصدق المرتبط بالمحك



أ- الصدق التلازمي (Concurrent Validity): الصدق التلازمي يستخدم للتحقق من مدى اتفاق نتائج الاختبار المراد التحقق من صدقه مع نتائج اختبار آخر موثوق يقيس نفس السمة أو البُعد وذلك عند تطبيقهما على نفس المجموعة وفي نفس الفترة الزمنية. على سبيل المثال عند تطوير اختبار جديد لقياس الذكاء يمكن تطبيقه على عينة من الأفراد ثم تطبيق اختبار آخر معتمد وذو صدق معروف كاختبار ستانفورد-بينيه (Stanford–Binet) على نفس الأفراد. فإذا كان معامل الارتباط بين نتائج الاختبارين مرتفعًا فإن ذلك يشير إلى ارتفاع الصدق التلازمي للاختبار الجديد.

ب- الصدق التنبؤي (Predictive Validity): الصدق التنبؤي يختبر مدى قدرة نتائج الاختبار الحالي على التنبؤ بأداء الأفراد في المستقبل على محك ذي صلة. يتم ذلك عبر تطبيق الاختبار على مجموعة من الأفراد ثم بعد فترة زمنية مناسبة (قد تمتد لأشهر أو سنوات) يتم الحصول على بيانات أداء هؤلاء الأفراد على المحك المطلوب وحساب معامل الارتباط بين النتائج.

- أمثلة عملية:

- **الاختبارات الأكاديمية:** قد تُستخدم نتائج شهادة البكالوريا للتنبؤ بمستوى النجاح الأكاديمي للطالب في الجامعة وفي تخصصات معينة.
- **اختبارات الذكاء:** يمكن أن تتنبأ بدرجة النجاح الأكاديمي أو القدرة على حل المشكلات المعقدة.
- **امتحانات القبول الجامعي:** مثل اختبار القدرات أو SAT يمكن أن تتنبأ بمدى قدرة الطالب على اجتياز متطلبات الدراسة الجامعية بنجاح.

في هذا النوع من الصدق يعتبر المحك أكثر أهمية من علامة الاختبار نفسها إذ تستخدم نتيجة الاختبار فقط باعتبارها مؤشرًا على الأداء المتوقع في المحك.

3-1-6- الصدق البنائـي أو الفرضي (المفهوم) (Construct Validity):

يهدف الصدق البنائي إلى الإجابة عن سؤالين جوهريين:

- ما السمة أو الصفة التي يقيسها الاختبار تحديدًا؟
- إلى أي مدى يتمكّن الاختبار من قياس هذه السمة بدقة وشمول؟

يمثل الصدق البنائي حجر الأساس في التأكد من أن الأداة تقيس فعليًا البنية أو السمة النفسية الكامنة التي وُضعت لقياسها كالميول والاتجاهات، التفكير الإبداعـي، التفكير الناقـد، الذكـاء، السمات الشخصية وغيرها من البنى النظرية التي لا يمكن ملاحظتها بشكل مباشر.

وتبرز أهمية هذا النوع من الصدق في كونه يجسر الفجوة بين النظرية والتطبيق؛ فهو يضمن أن النتائج المستخلصة من الاختبار تعكس المفهوم النظري الذي بُني عليه لا مجرد سلوكيات أو استجابات سطحية. ولتحقيق الصدق البنائي يعتمد الباحث على دمج دلائل متعددة، منها:

- تحليل طبيعة مفردات الاختبار ومدى اتساقها مع التعريف النظري للمفهوم.
- دراسة تجانس المفردات داخليًا.
- فحص ثبات الاختبار عبر مؤشرات كمية وإحصائية.
- التحقق من علاقات الارتباط أو الفروق مع مقاييس أخرى ذات صلة (أو غير ذات صلة) بما يتفق مع النظرية.

ومن أبرز الطرق التي يمكن للباحث استخدامها للتحقق من الصدق البنائـــي:

1- التحليل العاملي (Factor Analysis) للكشف عن البنية الداخلية للاختبار ومدى مطابقتها للمفهوم المراد قياسه.

2- دراسات الارتباط (Correlation Studies) مع مقاييس أخرى للتحقق من صدق التلازم أو التباين.

3- المقارنة بين المجموعات (Known-Groups Technique) حيث يتم التحقق من قدرة الاختبار على التمييز بين مجموعات يُفترض نظريًا أن تختلف في السمة المقاسة.

مثال تطبيقي:

إذا صمم اختبار لقياس التفكير الناقد لدى طلبة الجامعة يمكننا التحقق من صدقه البنائي من خلال التحليل العاملي للتأكد من أن بنية الاختبار تتطابق مع أبعاد التفكير الناقد النظرية، مثل: (التحليل، التقييم، الاستدلال) ثم دراسة علاقته باختبارات أخرى لقياس التفكير التحليلي أو القدرة الأكاديمية

وأخيرًا فحص أدائه لدى مجموعتين: طلبة برامج جامعية متقدمة مقابل طلبة في بداية الدراسة حيث يتوقع نظريًا وجود فروق لصالح المجموعة الأولى.

3-1-7- التحليل العاملـــي (Factor Analysis):

يعد التحليل العاملي من أهم الأساليب الإحصائية للتحقق من الصدق البنائي للاختبار إذ يهدف إلى الكشف عن مدى قياس الأداة للظاهرة أو البنية النظرية التي صُممت من أجلها.

أ- التحليل العاملي الاستكشافي (Exploratory Factor Analysis – EFA):

يستخدم التحليل العاملي الاستكشافي (EFA) في المراحل الأولى من تطوير الاختبار عندما لا تكون البنية العاملية محددة مسبقًا حيث يهدف إلى استكشاف عدد العوامل الكامنة وطبيعة الارتباط بين المفردات دون فرض نموذج محدد مسبقًا. يقوم الباحث بحساب معاملات الارتباط بين كل مفردة وبقية المفردات مما ينتج عنه مصفوفة ارتباطات تُظهر مدى تجانس المفردات. تقسم المفردات إلى مجموعات (عوامل) بحيث ترتبط مفردات كل مجموعة ببعضها أكثر من ارتباطها بالمجموعات الأخرى. وبهذا يمكن اختزال عدد كبير من المفردات إلى عدد أصغر من العوامل أو المكونات الأساسية التي تمثل السمات المشتركة للظاهرة المقاسة.

ب- التحليل العاملي التوكيدي (Confirmatory Factor Analysis – CFA):

يستخدم عندما يكون لدى الباحث نموذج نظري محدد لبنية الاختبار (عدد العوامل وطبيعة ارتباط المفردات بكل عامل) ويريد التأكد من مدى مطابقة البيانات الفعلية لهذا النموذج. في هذه الحالة يقوم الباحث باختبار مدى جودة التوافق بين البيانات والنموذج النظري باستخدام مؤشرات مطابقة (Fit Indices) مثل: CFI، TLI، RMSEA وغيرها.

وباختصار فإن التحليل العاملي الاستكشافي يبحث عن البنية الكامنة دون فرضيات مسبقة في حين أن التحليل العاملي التوكيدي يختبر مدى صحة فرضية أو نموذج نظري محدد حول هذه البنية.

جدول رقم 09 يوضح الفروق الأساسية بين التحليل العاملي الاستكشافي والتحليل العاملي التوكيدي.

البنـد	التحليل العاملي الاستكشافي(EFA)	التحليل العاملي التوكيدي(CFA)
الهـدف	استكشاف البنيـة الكامنـة للاختبـار دون وجود نموذج محدد مسبقًا.	التحقق من مدى مطابقـة البيانـات لنموذج نظري محدد مسبقًا.
مرحلـة الاستخـدام	المراحل الأولـى من تطويـر الاختبـار أو عند غياب معرفة سابقة بالبنيـة العامليـة.	بعد تحديد نموذج نظري للبنيـة العامليـة للاختبار.
الافتراضات المسبقـة	لا توجد فرضيات محددة حول عدد العوامل أو طبيعة ارتباط المفردات.	توجد فرضيات واضحة حول عدد العوامل والمفردات المرتبطة بكل عامل.
طريقـة التحليل	استخراج العوامل من مصفوفة الارتباطات وتحديد عددها باستخدام محكات إحصائية، مثل: قيمة الجذر الكامن > 1 أو رسم الهضبة أو الحصاة (Scree Plot).	تقدير معاملات النموذج باستخدام تقنيات، مثل: تحليل النمذجـة بالمعـادلات البنائيـة (SEM) وفحص مؤشرات المطابقـة.
النتيجـة	تحديد عدد العوامل والمفردات التي تنتمي لكل عامل، وتوليد نموذج أولـي للبنيـة العامليـة.	تحديد مدى ملاءمـة النموذج النظري للبيانات وتقدير جودة المطابقـة (Fit Indices).
المؤشرات الإحصائية	معاملات تشبعـات العوامـل (Factor Loadings)، ونسبة التباين المفسَّر.	مؤشرات المطابقـة، مثل: CFI، TLI، RMSEA، χ^2.
المرونـة	مرن في اكتشاف العلاقـات، لكن نتائجه استكشافية وقد تختلف بين عينات مختلفة.	أقل مرونـة لأنه يختبر نموذجًا محددًا، لكن نتائجه أدق في التحقق من البنية النظريـة.

3-1-8- دراسات الارتباط (Correlation Studies):

تعد دراسات الارتباط من الأساليب الأساسية للتحقق من الصدق البنائي إذ تقوم على فحص العلاقة بين نتائج الاختبار قيد الدراسة ونتائج مقاييس أو اختبارات أخرى. ويمكن أن يكون الهدف إما التحقق من الصدق التلازمي (Convergent Validity) من خلال إيجاد ارتباطات مرتفعة مع مقاييس مشابهة في البنية أو المفهوم أو التحقق من الصدق التمييزي (Discriminant Validity) من خلال التأكد من وجود ارتباطات منخفضة مع مقاييس مختلفة في البنيـة أو غير ذات صلة بالمفهوم المقاس.

أ-إيجاد معاملات الارتباط مع اختبارات أخرى:

يتم ذلك بحساب معامل الارتباط بين الاختبار المستهدف للتحقق من صدقه البنائي وبين اختبارات أخرى موثوقة ومثبتة الصدق تقيس نفس الظاهرة أو ظواهر مرتبطة بها نظرياً. فعلى سبيل المثال إذا أراد باحث بناء مقياس لقياس الطموح لدى الطلبة فإنه قد يختار تطبيق اختبار آخر يقيس المثابرة نظرًا لاعتبارها إحدى السمات الجوهرية للفرد الطموح. فإذا حصل على معامل ارتباط مرتفع ودال إحصائيًا بين الاختبارين فهذا يعد دليلًا على تحقق الصدق البنائي لمقياس الطموح.

ب- إيجاد العلاقة بين درجات المفردات والاختبار (الاتساق الداخلي):

تعتمد هذه الطريقة على افتراض أن الدرجة الكلية للاختبار تمثل معيارًا صادقاً لقياس الظاهرة المستهدفة وبالتالي فإن مفردات الاختبار الجيدة يجب أن تكون مرتبطة ارتباطًا وثيقًا بهذه الدرجة الكلية. وللتحقق من ذلك يقوم الباحث بحساب معامل الارتباط بين درجة كل مفردة والدرجة الكلية (باستثناء درجة المفردة نفسها عند الحساب) ثم يتم استبعاد أي مفردة يظهر ارتباطها منخفضًا باعتبار أنها لا تقيس البنيـة أو السمة النفسية التي يقيسها الاختبار ككل.

3-1-9- الصدق التمايزي (Differential Validity):

لا يقتصر الصدق البنائي على أن يظهر الاختبار ارتباطًا مرتفعًا مع مقاييس تشبهه بل يتطلب أيضًا ألا يرتبط بقوة مع مقاييس يفترض أن تكون مختلفة عنه. فمثلاً إذا حقق اختبار التفكير الرياضي ارتباطًا قدره (0.60) مع اختبار للقدرة العددية ولكنه أظهر ارتباطًا أعلى من (0.70) مع اختبار للفهم القرائي فإن ذلك يثير الشك حول صدقه. بمعنى آخر ليس مجرد وجود علاقة بين الاختبار ومحك معين دليلًا كافيًا على الصدق؛ إذ ينبغي استبعاد احتمال أن تكون هذه العلاقة ناتجة عن سمة عامة أوسع كالذكاء العام قبل الادعاء بأن الاختبار يقيس قدرة محددة. ويشير الصدق التمايزي إلى الفرق بين:

- العلاقة بين الاختبار والبناء المفترض قياسه الصدق التقاربي (convergent validity).
- العلاقة بين الاختبار والمتغيرات الأخرى التي قد تسبب تداخلاً في النتائج الصدق التباعدي (divergent validity).

وغالبًا ما يفحص هذا النوع من الصدق باستخدام منهج السمات المتعددة–الطرق المتعددة (Multitrait–Multimethod) الذي يقوم على قياس ثلاث سمات أو أكثر بواسطة ثلاث طرق أو أكثر.

فعلى سبيل المثال في اختبارات الشخصية يمكن قياس سمات كالانبساطية، العصابية، والضمير الحي، باستخدام التقارير ذاتية، الأساليب الإسقاطية، تقديرات الأقران. في هذه الحالة يجب أن تكون الارتباطات مرتفعة بين نفس السمة عند قياسها بطرق مختلفة، بينما يجب أن تبقى الارتباطات ضعيفة بين سمات مختلفة (Rust, et al, 2021, p.45).

3-1-10- طريقة المجموعات المعروفة (Known Groups Method):

تعد طريقة المجموعات المعروفة أحد أكثر أشكال الأدلة شيوعًا لدعم الصدق البنائي (Construct Validity) وتستخدم عندما يكون الاختبار قادرًا على تمييز الأفراد المعروف أنهم يمتلكون صفة معينة عن أولئك الذين لا يمتلكونها. باستخدام طريقة المجموعات المعروفة يتم اختيار محك يمكنه تحديد وجود أو غياب صفة معينة ويستخدم السياق النظري وراء البنية (Construct) لتوقع كيفية تصرف المجموعات المختلفة. وبالتالي يدعم صدق الاختبار إذا أظهرت نتائجه هذه الاختلافات المعروفة بين المجموعات. وبالمثل يمكن دراسة المجموعات المعروفة باستخدام مجموعات من الأفراد الذين يختلفون في مستويات أو شدة الصفة نفسها. تقيم طريقة المجموعات المعروفة قدرة الاختبار على التمييز بين هذه المجموعات بناءً على تباين المتوسطات بين المجموعات على الاختبار. فعلى سبيل المثال يجب أن تسجل مجموعة الأفراد المعروف أنهم غير مكتئبين درجات أقل على مقياس الاكتئاب مقارنة بالمجموعة المعروفة بأنها مكتئبة (Portney, et al, 2008, p.121).

3-1-11- اختبار الفرضيات (Hypothesis Testing):

نظرًا لأن البنى النفسية (Construct) تستند إلى أساس نظري يمكن تقييم الصدق البنائي للأداة من خلال اختبار فرضيات محددة مستمدة من النظرية. فمثلاً قام دودز وزملاؤه بتقييم صدق البنية لمقياس الاستقلالية الوظيفية (Functional Independence Measure – FIM) بناءً على الفرضية القائلة بأن الأداة يجب أن تتمكن من تمييز الفروق الوظيفية بين الأشخاص ذوي الحالات السريرية المختلفة، ويعكس مفهوم الاستقلالية الوظيفية (FIM) عبء الرعاية ودرجة المساعدة التي يحتاجها المريض لأداء الأنشطة اليومية، والحركة، والوظائف المعرفية. وانطلاقًا من هذا الإطار النظري صاغ الباحثون ثلاث فرضيات وهي:

1- تنخفض درجات الاستقلالية الوظيفية (FIM) مع التقدم في العمر وزيادة الأمراض المصاحبة.

2- ترتبط درجات الاستقلالية الوظيفية (FIM) بوجهة خروج المريض وفق مستوى الرعاية في المكان، كالمنزل أو مرفق الرعاية المتخصصة.

3- وجود علاقة بين درجات الاستقلالية الوظيفية (FIM) وشدة الحالة لدى المرضى المصابين بالشلل النصفي أو إصابات الحبل الشوكي أو السكتة الدماغية.

باستخدام بيانات أكثر من 11.000 مريض أظهرت النتائج دعمًا متفاوتًا للفرضيات حيث برزت علاقة قوية بين درجات الاستقلالية الوظيفية (FIM) ووجهة خروج المرضى وشدة إصابات الحبل الشوكي والسكتة الدماغية. ويُعد هذا التحليل دليلًا واضحًا على الصدق البنائي للأداة لكنه يسلط الضوء على الحاجة إلى اختبارات إضافية لفهم العلاقة بين درجات الاستقلالية الوظيفية (FIM) والتشخيصات المختلفة والنتائج السريرية المتنوعة (Portney, et al, 2008, p.123).

3-2- العوامل المؤثرة على الصدق:

تتأثر دقة وموثوقية نتائج أي اختبار بعدد من العوامل التي قد تؤدي إلى ارتفاع أو انخفاض مستوى الصدق سواء كانت مرتبطة بالمفحوصين أنفسهم أو بطريقة إدارة (تطبيق) الاختبار أو بخصائص الاختبار ذاته، ويمكن تصنيف هذه العوامل على النحو الآتي:

3-3- عوامل تتعلق بالمفحوص: وتشمل الخصائص أو الظروف الشخصية التي قد تؤثر على أداء الأفراد، مثل:

أ- القلق: وخاصة القلق الاختباري الذي قد يضعف القدرة على التركيز ويؤثر على استدعاء المعلومات أو حل المسائل.

ب- الغش أو السلوك غير النزيه: مما يؤدي إلى تضخيم درجات بعض المفحوصين بشكل لا يعكس قدراتهم الحقيقية.

ج- الدافعية والتحفيز: انخفاض الحافز قد يؤدي إلى أداء أقل من المستوى الفعلي للفرد.

د- الحالة الصحية أو المزاجية: مثل التعب أو المرض أو الحالة النفسية غير المستقرة.

3-4-عوامل تتعلق بإدارة الاختبار: وتشمل الظروف والإجراءات المصاحبة لعملية تطبيق الاختبار، مثل:

أ- الظروف الفيزيقية: الحرارة المرتفعة، الإضاءة الضعيفة، الضوضاء، أو التهوية غير الجيدة، وجميعها قد تؤثر سلبًا على تركيز المفحوصين.

ب- جودة المواد الاختبارية: سوء طباعة أوراق الاختبار أو وجود أخطاء مطبعية قد تربك المفحوص.

ج- صياغـة التعليمات: التعليمات الغامضة أو المضللة قد تؤدي إلى تباين في فهم المطلوب من المفحوصين.

د- طريقـة المراقبـة: الرقابة الصارمة أو العشوائية المفرطة قد تخلق ضغوطًا أو تمنح فرصًا للغش.

3-5- عوامل تتعلق بالاختبـار نفسه: ترتبط بخصائص الأداة وملاءمتها لقياس الظاهرة المستهدفة، وتشمل:

أ- مستوى صعوبـة المفردات: المفردات السهلة جدًا أو الصعبة جدًا تقلل من القدرة على التمييز بين مستويات القدرة المختلفة.

ب- وضوح الصياغـة: المفردات الغامضة أو متعددة المعنى قد تربك المفحوصين.

ج- طول الاختبـار: الاختبارات الطويلة جدًا قد تسبب الملل أو الإرهاق، بينما القصيرة جدًا قد لا تمثل السلوك المستهدف تمثيلًا كافيًا.

د- ثبات الاختبار: انخفاض الثبات يقلل من دقة القياس مما يؤثر بدوره على الصدق.

هـ- التباين في الدرجات: انخفاض التباين يجعل التمييز بين الأفراد أكثر صعوبة.

و- صدق مقياس المحك: في حالة الاعتماد على المحك الخارجي للتحقق من الصدق فإن ضعف صدق المحك نفسه سينعكس على التقدير.

ز- مدى تمثيل الاختبار لمجال السلوك: الاختبار غير الممثل لمجال الظاهرة المستهدفة يفقد جزءًا من صدقه.

ح- تجانس عينة التقنين: العينة غير الممثلة للمجتمع المستهدف تقلل من إمكانية تعميم النتائج.

3-2- الثبـات (Reliability):

يشير مفهوم الثبـات في الاختبـارات إلى مدى خلو الدرجـات من الأخطـاء غير المنتظمـة (العشوائية) التي قد تشوب عملية القياس أي مدى قدرة الاختبار على قياس المقدار الحقيقي للسمة المستهدفة بشكل متسق. ويُعد الاختبار ثابتًا (Reliable) إذا كانت نتائجه متشابهة نسبيًا عند إعادة تطبيقه في ظروف مختلفة أو عند إعادة صياغة بنوده بما يعكس نفس البنية النظرية مع بقاء مستوى القياس دقيقًا ومتسقاً. وبمعنى آخر فإن الثبات يعكس درجة الاتساق والدقة في القياس بحيث لا تتأثر نتائج الأفراد إلا بالسمة المستهدفة وليس بعوامل عرضية أو مؤقتة.

3-2-1- الثبات والدرجـات الحقيقيـة:

وفق النظرية الكلاسيكية للقياس (Classical Test Theory) تتكون الدرجة التي يحصل عليها الفرد في الاختبار من عنصرين رئيسيين:

أ- الدرجـة الحقيقيـة (True Score): وهي الدرجة التي تمثل المستوى الفعلي للفرد في السمة التي يقيسها الاختبار خالية من أي خطأ عشوائي.

ب- درجـة الخطـأ (Error Score): وهي الجزء من الدرجة الذي يعكس تأثير العوامل العشوائية غير المرغوبة، مثل: التعب، الصدفة، الحظ، أو سوء الفهم المؤقت للسؤال. وبذلك يمكن التعبير عن العلاقة على النحو الآتي:

الدرجة الملاحظة (OS)= الدرجة الحقيقية (TS) + درجة الأخطاء العشوائية (ES)

الدرجة الحقيقية (TS)= الدرجة الملاحظة (OS) - درجة الأخطاء العشوائية (ES)

وإذا تمكنا من تقدير درجة الخطأ وإزالتها من الدرجة الملاحظة (Observed Score) فإننا نقترب من معرفة الدرجة الحقيقية للفرد. وكلما كان حجم الخطأ أصغر زاد ثبات الاختبار والعكس صحيح.

مثال 01: حصل طالب في اختبار ا التكافؤ وتحت ظروف متشابهة على العلامات التالية: 72، 68، 67، 65، 73، 80، 75، 70، 67، 72، 63، 68. ما هي العلامة الحقيقية للطالب، وما هي أخطاء القياس في الاختبارات المختلفة وما هو تباين الأخطاء؟

الحل: العلامة الحقيقية (True Score) هي المتوسط الحسابي للعلامات، فإذا عبرنا عن العلامة الحقيقية بالرمز (T) فإن:

$$T = \overline{X} = \frac{\sum X}{n} = \frac{840}{12} = 70$$

- ونكمل الحل كما في الجدول التالي:

جدول رقم 10 يوضح العلاقة بين الدرجة الملاحظة والخطأ ومربع الخطأ

الدرجة الملاحظة X	الخطأ E	مربع الخطأ E^2
72	2+	2
68	2-	2
67	3-	9
65	5-	25
73	3+	9
80	10+	100
75	5+	25
70	0	0
67	3-	9
72	2+	4
63	7-	49
68	2-	4
840	**0**	**242**

بما أن معدل الأخطاء صفر فإن تباين الأخطاء (Error Variance) يساوي:

$$r^{2}{}^{1} = \frac{}{n}[\sum E^2]$$

$$= \frac{242}{12} = 20.17$$

مثال02: كانت العلامات الحقيقية والعلامات الملاحظة (المشاهدة) لخمسة طلبة كما في الجدول التالي (سامي،2011، ص255).

المطلوب: أوجد أخطاء القياس والخطأ المعياري للقياس ثم أوجد معامل الثبات؟

جدول 11 معطيات لحساب أخطاء القيـاس والخطأ المعيـاري للقيـاس.

رقم الطالب	العلامة الملاحظـة X	العلامة الحقيقيـة T	أخطاء القيـاس E
1	36	32	4
2	20	15	5
3	24	30	6-
4	44	43	1
5	20	24	4-
المجموع	**144**	**144**	**0**

الحل:

1- إيجاد أخطاء القياس باستخدام الصيغة التالية: **E= X – T**

2- مجموع أخطاء القياس يساوي صفر.

3- نحسب المتوسط الحسابي لكل من الدرجـات الملاحظـة والدرجـات الحقيقيـة أي:

$$\bar{T} = \frac{144}{5} = 28.8$$

$$\bar{x} = \frac{144}{5} = 28.8$$

4- ثم نحسب التباينات باستخدام المعادلـة التالية:

$$S^2 = \frac{1}{n}\left[X^2 - n\,\bar{X}^2\right]$$

$$s_T^2 = \frac{1}{5}[(32^2 + 15^2 + 30^2 + 43^2 + 24^2) - 5X28.8^2] = 85.34 \approx 85$$

$$s_x^2 = \frac{1}{5}[(36^2 + 20^2 + 24^2 + 44^2 + 20^2) - 5X28.8^2] = 92.16 \approx 92$$

- بنفس الطريقة نحسب تباين أخطاء القياس:

$$s_E^2 = \frac{1}{5}[((+4)^2 + (+5)^2 + (-6)^2 + (+1)^2 + (-4)^2) - 5X0] = 18.8$$

$$\approx 19$$

- ومعامل الثبات ***r*** يساوي:

$$r_{xx} = \frac{S_T^2}{S_X^2} = \frac{85}{92} = 0.92$$

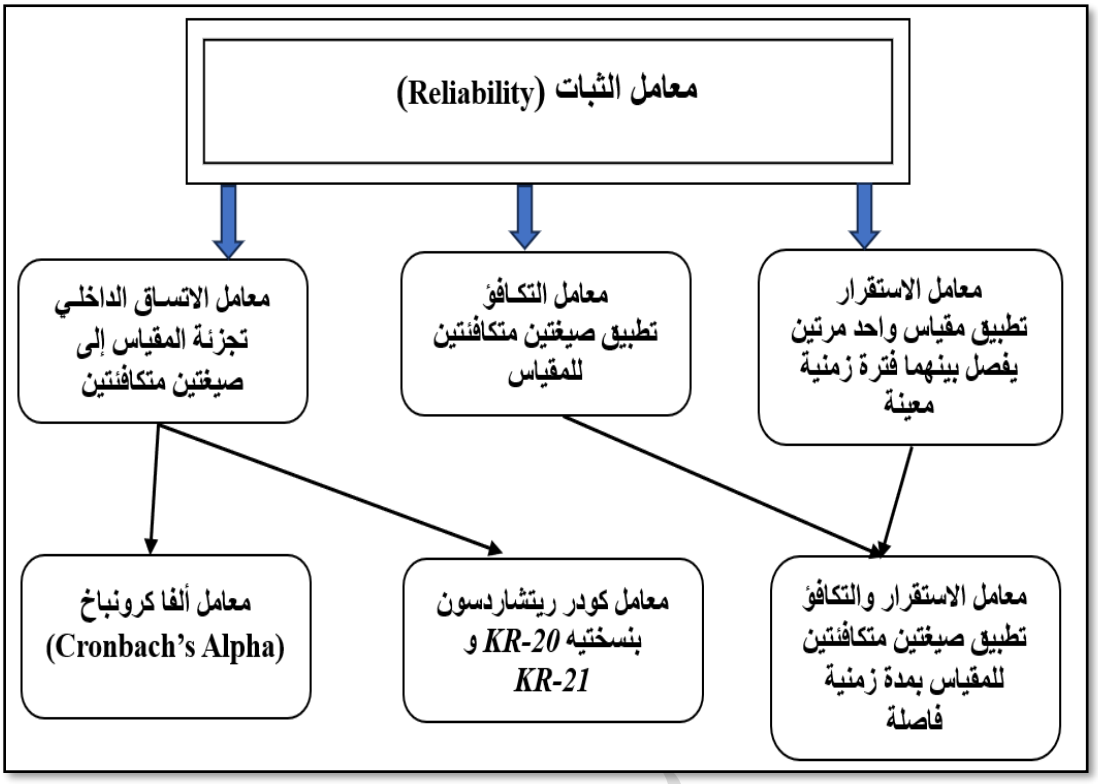
- إذن الخطأ المعياري للقياس يساوي: $S_E = S_x\sqrt{(1-r)}$

- حيث: S_x = الانحراف المعياري.

$$S_E = \sqrt{92.16}\,\sqrt{1-0.92} = 2.72$$

- ويمكن حساب ثبـات الاختبــار بطرق عديد يمكن تلخيصها في الشكل التالي:

شكل رقم 10 يوضح أنواع معامل الثبات



3-2-2- طريقة الاختبار وإعادة الاختبار (Test- Retest Method):

تعد طريقة الاختبار وإعادة الاختبار واحدة من أبرز أساليب قياس ثبات الاختبارات النفسية والتربوية عبر الزمن. تقوم هذه الطريقة على مبدأ بسيط لكنه قوي: النتائج الموثوقة يجب أن تبقى ثابتة عند إعادة القياس في ظروف مماثلة. في هذه الطريقة يُطبق الاختبار على مجموعة من الأفراد أو عينة ممثلة للمجتمع محل الدراسة ويتم تسجيل النتائج بدقة. بعد فترة زمنية محددة وغالبًا ما تكون بين أسبوعين إلى عدة أسابيع حسب طبيعة الصفة المقاسة ثم يُعاد تطبيق نفس الاختبار على نفس الأفراد أو على عينة مماثلة. يتم بعد ذلك حساب معامل الثبات بين النتائج في التطبيقين وعادةً باستخدام معامل الارتباط لبيرسون (Pearson Correlation Coefficient) الذي يقيس مدى الترابط والتوافق بين مجموعتي البيانات.

- مزايا هذه الطريقة:

أ- تقييم الاستقرار الزمني: تساعد في معرفة ما إذا كانت نتائج الاختبار مستقرة على مدى فترة زمنية معينة.

ب- سهولة التطبيق: لا تتطلب أدوات أو أساليب إحصائية معقدة وغالبًا يمكن تطبيقها على أي نوع من الاختبارات.

ج- مؤشر ثبات مباشر: يعطي فكرة واضحة عن قدرة الاختبار على قياس الصفة النفسية أو التعليمية بشكل مستمر ودقيق.

- سلبيات طريقة الاختبار وإعادة الاختبار:

رغم أهمية هذه الطريقة في تقدير ثبات الاختبارات إلا أنها لا تخلو من بعض السلبيات والقيود التي ينبغي أخذها بعين الاعتبار، ومن أبرزها:

أ- تأثير الفترة الزمنية بين التطبيقين: إذا كان الفاصل الزمني قصيرًا فقد يتذكر المفحوصون إجاباتهم السابقة مما يرفع معامل الثبات بشكل مصطنع. أما إذا طال الفاصل الزمني فقد تتدخل عوامل أخرى، مثل: النسيان أو النضج أو التغيرات الطبيعية في الخصائص النفسية والسلوكية مما يؤدي إلى انخفاض معامل الثبات.

ب- تأثير التعلم أو التذكر: قد يكتسب المفحوص خبرة من الاختبار الأول تساعده في الاختبار الثاني وهو ما يجعل النتائج متأثرة بعملية التعلم وليس فقط بالصفة المقاسة.

ج- الوقت والجهد: تتطلب هذه الطريقة جهدًا إداريًا وتنظيميًا كبيرًا لأنها تقوم على تطبيق الاختبار مرتين مع ما يرافق ذلك من متابعة المفحوصين وضبط ظروف التطبيق.

د- تغير ظروف التطبيق: قد تختلف ظروف البيئة أو السياق بين المرتين (مكان الاختبار، حالة المفحوص النفسية، أو حتى ظروف اجتماعية وصحية طارئة)، مما قد ينعكس سلبًا على النتائج ويضعف دقة التقدير.

هـ- اختلاف حجم العينة: في بعض الحالات قد لا يتمكن جميع المفحوصين من المشاركة في التطبيق الثاني مما يؤدي إلى نقص في حجم العينة ويؤثر على دقة معامل الثبات المحسوب.

- ملاحظة هامة: رغم هذه السلبيات تبقى طريقة الاختبار وإعادة الاختبار من أكثر الطرق شيوعًا وموثوقية في تقدير الثبات شريطة مراعاة الضوابط التالية:

أ- أن يكون الفاصل الزمني بين التطبيقين مناسبًا عادةً بين أسبوعين وستة أسابيع.

ب- أن تتم إعادة التطبيق في ظروف مماثلة قدر الإمكان.

ج- استبعاد استخدامها مع اختبارات الذاكرة أو الاختبارات التي تتأثر بدرجة كبيرة بالتعلم والممارسة.

3-2-3- طريقة الصور المتكافئة (Equivalent or Parallel Forms Method):

تعد طريقة الصور المتكافئة من أبرز الأساليب المستخدمة لقياس ثبات الاختبارات حيث تقوم على تطوير صورتين متكافئتين من نفس الاختبار بحيث تتشابهان في جميع الخصائص الأساسية من حيث:

- عدد المفردات ومكوناتها.
- مستوى الصعوبة والسهولة.
- المحتوى والأهداف التي يقيسها الاختبار.
- طول الاختبار وطريقة تطبيقه وتصحيحه وتوقيته.
- تماثل المتوسطات والتباينات الإحصائية لدرجات الأفراد على كلا الاختبارين.

بعد إعداد صورتين للاختبار يتم تطبيقهما على نفس مجموعة الأفراد أو مجموعتين متكافئتين بفاصل زمني قصير لتجنب تأثير النضج أو التعلم ثم يُحسب معامل الارتباط بين الأداء في الصورتين ليكون مؤشرًا على ثبات الاختبار.

- إيجابيات الطريقة:

أ- التقليل من أثر التذكر أو التعلم: نظرًا لاختلاف صياغة المفردات في الصورتين يقل احتمال أن يتأثر أداء المفحوصين بخبرتهم السابقة مع الاختبار.

ب- تقدير أكثر دقة لثبات المحتوى: لأنها تعتمد على صورتين متكافئتين فإنها توفر مؤشرًا عمليًا على مدى اتساق الأداة في قياس نفس البنية أو الصفة.

ج- إمكانية الاستخدام في مواقف مختلفة: يمكن الاستفادة من الصورتين كأدوات بديلة في التطبيق المتكرر، مثلاً: قبل وبعد التدريب أو البرنامج العلاجي.

د- مرونة في التطبيق: تتيح للباحثين أو المعلمين إمكانية التبديل بين الصورتين عند الحاجة دون التأثير الكبير على النتائج.

- سلبيات الطريقة:

أ- صعوبة بناء صورتين متكافئتين تمامًا: إعداد اختبارين متوازيين يتطلب جهدًا كبيرًا وخبرة عالية في تصميم المفردات وضبط مستويات الصعوبة.

ب- زيادة الأعباء على المفحوصين: إعادة الاختبار بصورة أخرى قد يسبب الملل أو الإرهاق خاصة إذا كان الفاصل الزمني قصيرًا.

ج- تكلفة زمنية ومادية مرتفعة: تحتاج هذه الطريقة إلى موارد إضافية لتطوير صورتين متكافئتين بدلًا من اختبار واحد.

د- احتمالية عدم التطابق الكامل: حتى مع الحرص الشديد قد توجد فروق دقيقة بين الصورتين تؤثر على عدالـة القيـاس ودقـة معامل الثبـات.

- ملاحظـة: رغم التحديات فإن طريقة الصور المتكافئة تعد من أكثر الطرق صرامة وموضوعية في تقدير الثبات حيث تتيح فحص استقرار النتائج مع الحد من تأثير الذاكرة والتعلم. ولهذا تُستخدم كثيرًا في الاختبارات واسعة النطاق كالاختبارات المدرسية أو المقاييس النفسية المشهورة متى توفرت الإمكانيات اللازمة لتطوير صورتين متكافئتين فعلاً.

3-2-4- طريقة التجزئة النصفية (Split – Half Method):

تعد هذه الطريقة من أقدم وأكثر طرق تقدير الثبات شيوعاً حيث يقسم الاختبار إلى قسمين متكافئين من حيث: عدد المفردات، مستوى الصعوبة، طبيعة الأهداف، وزمن الإجابة. وغالباً ما يجرى هذا التقسيم إما بطريقة الفردي–الزوجي (Odd–Even) بحيث تجمع إجابات البنود الفردية في جزء والزوجية في جزء آخر أو بطريقة تقسيم الاختبار إلى نصفين متساويين مع مراعاة التوازن. بعد ذلك يحسب معامل الارتباط بين أداء الأفراد على الجزأين ويعتبر هذا المعامل مقياساً لثبات نصف الاختبار فقط. ولأن الهدف هو تقدير ثبات الاختبار كاملاً تستخدم معادلات تصحيحية للحصول على معامل ثبات مناسب، وهناك عدة معادلات تستخدم لتصحيح معامل ثبات نصفي الاختبار منها:

أ- معادلة سبيرمان- براون (Spearman–Brown Prophecy Formula):

تستخدم هذه المعادلـة لتقدير ثبات الاختبار الكامل استناداً إلى معامل الارتباط المحسوب بين نصفي الاختبار. فعند تقسيم الاختبار إلى نصفين متكافئين وحساب معامل الارتباط بينهما فإن هذه القيمة لا تعكس الثبات الحقيقي للاختبار بأكمله لأنها تمثل فقط ثبات نصف الاختبار. لذلك نحتاج إلى إجراء تصحيح يعرف باسم تصحيح سبيرمان–براون للحصول على معامل الثبات الكلي. يتم حساب التصحيح وفق المعادلة التالية:

$$r_{full} = \frac{2(r_{half})}{1 + r_{half}}$$

حيث، أن:

r_{half} = هو معامل الارتباط بين نصفي الاختبار.

r_{full} = معامل ثبات الاختبار ككل بعد التصحيح.

استخدامات المعادلـة:

- شائعة جداً في اختبارات التحصيل الدراسي واختبارات القدرات.
- تساعد الباحث على تقدير الثبات المتوقع إذا زاد أو نقص طول الاختبار.
- تستخدم أيضاً للتنبؤ بمستوى الثبات عند تعديل عدد المفردات (زيادة البنود عادة ترفع الثبات).

حدود المعادلـة:

- تفترض أن نصفي الاختبار متكافئان في جميع الجوانب (الصعوبة، الأهداف، عدد البنود).
- إذا لم يكن الانقسام متكافئاً، قد تعطي نتائج مضللة.
- لا تستخدم لتقدير ثبات الاختبارات متعددة الأبعاد (Multi-dimensional tests).

مثال عددي للتوضيح:

لو كان معامل الارتباط بين نصفي الاختبار $r_{half}=0.70$، فإن معامل الثبات الكلي، يساوي:

$$r_{full} = 2(0.70)/1+0.70 = 1.40/1.70 \approx 0.82$$

إذن ثبات الاختبار بعد التصحيح يساوي (0.82) وهو أعلى من معامل الارتباط بين نصفي الاختبار كما هو متوقع.

ب- معادلـة رولون (Rulon Formula):

تعد معادلة رولون إحدى الطرائق الإحصائية المستخدمة لتقدير معامل ثبات الاختبار الكلي في طريقة التجزئة النصفية (Split-Half Method). وتعتمد هذه الطريقة على دراسة التباين بدلاً من الاعتماد المباشر على معامل الارتباط بين نصفي الاختبار (Tes and Rabu, 2011). وتقوم الفكرة الأساسية على حساب:

1- تباين الدرجات الكلية للاختبار(S^2_total).

$$S_t^2 = \frac{\Sigma X_t^2 - \frac{(X_t)^2}{n}}{n}$$

2- تباين الفروق بين درجات الأفراد في النصف الأول والنصف الثاني (S^2_diff).

$$S_d^2 = \frac{\Sigma X_d^2 - \frac{(X_d)^2}{n}}{n}$$

ثم يطبق القانون التالي للحصول على معامل ثبات الاختبار:

$$r_{11} = 1 - \frac{S_d^2}{S_t^2}$$

مزايـا معادلــة رولون:

- لا تحتاج إلى حساب معامل الارتباط بل تعتمد فقط على التباين.
- تعتبر بديلاً جيداً في حالة الرغبة في تجنب الانحياز الناتج عن اختلاف طريقة تقسيم البنود.
- تبرز مدى مساهمة التباين الناتج عن الفروق بين نصفي الاختبار في تقدير الثبات الكلي.

حدودها:

- تفترض تكافؤ نصفي الاختبار بشكل كامل في الصعوبة، عدد البنود، المحتوى.
- لا تصلح إذا كان الاختبار قصيراً أو يحتوي على أبعاد متعددة.
- قد تتأثر دقتها إذا لم يكن تقسيم البنود تم بطريقة عادلة.

مثال عددي:

- لنفترض أن التباين الكلي لدرجات الاختبار يساوي 100.
- وتباين الفروق بين درجات الأفراد في نصفي الاختبار يساوي 18.

بالتطبيق:

$$r_{11} = 1 - 18/100 = 1 - 0.18 = 0.82$$

إذن معامل الثبات الكلي للاختبار يساوي 0.82 وهو مستوى مرتفع ويدل على اتساق داخلي جيد.

ج- معادلة جتمان (Guttmann):

تعد معادلة جتمان من أبرز صيغ حساب الثبات باستخدام طريقة التجزئة النصفية. وتقوم هذه المعادلة على حساب تباين درجات النصف الأول من الاختبار وتباين درجات النصف الثاني إضافةً إلى تباين الدرجات الكلية للاختبار. وتمتاز عن بعض المعادلات السابقة، مثل – معادلة سبيرمان-براون أو رولون – بكونها تراعي احتمال وجود فروق في تباين درجات النصفين وهو ما قد يحدث فعليًا في كثير من الاختبارات حيث لا يكون النصفان متكافئين تمامًا من حيث الصعوبة أو التباين.
وتكتسب هذه الطريقة أهميتها من كونها من أكثر أساليب تقدير الثبات شيوعًا إذ يتم تطبيق الاختبار مرة واحدة فقط على الأفراد مما يوفر الوقت والجهد مقارنةً بالطرق التي تتطلب إعادة تطبيق الاختبار. وبعد ذلك يجزأ الاختبار إلى نصفين (أو أحيانًا أكثر) بطرائق متعددة كالفصل بين البنود الزوجية والفردية أو تقسيم الاختبار إلى جزأين متساويين في الطول والمحتوى. ويعتبر هذا الإجراء عمليًا وبسيطًا كما أنه يعكس بدرجة معقولة مدى اتساق أداء الأفراد على أقسام مختلفة من نفس الاختبار مما يمنح الباحثين مؤشرًا واضحًا على ثبات الاختبار وموثوقيته، وتعطى بالصيغة الرياضية التالية:

$$r_{tt} = \frac{S_\chi^2 - (S_1^2 + S_2^2)}{S_\chi^2}$$

حيث أن:

r_{tt} = معامل الثبات الكلي للاختبار.

S_χ^2 = التباين الكلي لدرجات الاختبار.

S_1^2 = تباين درجات النصف الأول من الاختبار.

S_2^2 = تباين درجات النصف الثاني من الاختبار.

وتظهر هذه المعادلة أن الثبات يزداد كلما كان التباين الكلي أكبر مقارنة بمجموع تباينات نصفي الاختبار أي أنّها تُصحح الخلل الذي قد ينشأ إذا اختلفت درجة تباين نصفي الاختبار.

3-2-5- طريقة ثبات الاتساق الداخلي (Internal Consistency): تعتمد هذه الطريقة على فحص درجة الاتساق بين مفردات الاختبـار وذلك من خلال قياس قوة الارتباط بين أداء الفرد على أي مفردة من مفرداته والدرجة الكلية للاختبار أو من خلال تحليل العلاقة بين المفردات بعضها مع

بعض. وتهدف هذه المقاربة إلى التأكد من أن المفردات جميعها تعمل في اتجاه واحد لقياس البعد أو السمة النفسية ذاتها وهو ما يعكس مستوى الصدق والثبات الداخلي للأداة.

أ- معادلة كودر-ريتشاردسون (Kuder–Richardson Formula - KR-20): يعد من أشهر مؤشرات ثبات الاتساق الداخلي للاختبــارات التي تتكون من مفردات ثنائية الاستجابة، مثل: (صح/خطأ، نعم/لا). وتعتمد هذه الصيغة على حساب التباين لكل مفردة من مفردات الاختبار بالإضافة إلى التباين الكلي للاختبار. وفي حال عدم توفر بيانات عن تباين المفردات، يمكن اللجوء إلى الصيغة المبسطة (***KR-21***) التي تُعد أقل دقة ولكنها أسهل وأسرع في الحساب؛ لأنها تفترض تجانس صعوبة المفردات ولا تتطلب حساب تباين كل مفردة على حدة، وتعطى بالصيغة الرياضية التالية:

$$KR\text{-}20 = \left(\frac{K}{K-1}\right)\left[1-\frac{\Sigma_{k=1}^{p}(p_k q_k)}{S^2_{\text{Total}}}\right]$$

حيث:

K = إجمالي عدد المفردات على الأداة.

p_k = نسبة الأفراد المجيبين إجابة صحيحة على المفردة.

q_k = نسبة الأفراد المجيبين إجابة خاطئة على المفردة.

S^2Total = تباين الكلي لدرجات الاختبار.

وقد وضع كودر-ريتشاردسون شروطاً لاستخدام هذه المعادلة، وهي:

- أن تكون درجة أسئلة الاختبار (صفر أو واحد).
- ألا يكون عدد الأسئلة المتروكة كبير.
- تقارب مستوى صعوبة الأسئلة.
- تساوي معاملات الارتباط بين درجات الأسئلة. ومن الواضح أن هذه الشروط لا توجد في اختبارات المجال الوجداني ولذلك ننصح بعدم استخدامها في هذا النوع من الاختبارات.

مثال تطبيقي:

لنفترض أن لدينا اختبارًا مكوَّنًا من خمس مفردات (أسئلة)، وقد جرى تطبيقه على عينة صغيرة مكونة من ستة طلاب. يجيب كل طالب عن كل مفردة باستخدام بديلين فقط، إمّا (1) للدلالة على

الإجابة الصحيحة، أو (0) للإشارة إلى الإجابة الخاطئة. ولغرض التوضيح العملي نعرض فيما يلي النتائج التي حصل عليها الطلاب الستة عبر المفردات الخمس، كالتالي:

جدول رقم 12 معطيات لمثال لحساب معامل كودر- ريتشاردسون (KR-20)

الطلبة	المفردة1	المفردة2	المفردة3	المفردة4	المفردة5	المجموع
الطالب1	1	0	1	1	1	4
الطالب2	1	1	1	1	1	5
الطالب3	1	0	1	0	0	2
الطالب4	0	0	0	1	0	1
الطالب5	1	1	1	1	1	5
الطالب6	1	1	0	1	1	4
p_k	0.8333	0.5	0.6667	0.8333	0.6667	**S^2= 2.70**
q_k	0.1667	0.5	0.3333	0.1667	0.3333	
$p_k x q_k$	0.1389	0.25	0.2222	0.1389	0.2222	

1- حساب المتوسط الحسابي للدرجات:

$$\bar{x} = \frac{\sum x}{N} = \frac{4+5+2+1+5+4}{6} = \frac{21}{6} = 3.5$$

2- حساب انحراف كل درجة عن المتوسط، ثم تربيع الفروق:

$x_i - \bar{x}$

- الطالب1: $4 - 3.5 = 0.5$ $\quad (0.5)^2 = 0.25$
- الطالب2: $5 - 3.5 = 1.5$ $\quad (1.5)^2 = 2.25$
- الطالب3: $2 - 3.5 = -1.5$ $\quad (-1.5)^2 = 2.25$
- الطالب4: $1 - 3.5 = -2.5$ $\quad (-2.5)^2 = 6.25$
- الطالب5: $5 - 3.5 = 1.5$ $\quad (1.5)^2 = 2.25$
- الطالب6: $4 - 3.5 = 0.5$ $\quad (0.5)^2 = 0.25$

3- المجموع: $\sum = (x_i - \bar{x})^2 = 13.5$

4- حساب التباين الكلي للدرجات، بالمعادلة التالية:

$$\sigma^2 = \frac{\sum = (x_i - \bar{x})^2}{N-1} = \frac{13.5}{5} = 2.7$$

- وبتعويض النتائج في معادلة كودر- ريتشاردسون نتحصل على التالي:

$$\sum pk\ x\ qk = 0.1389 + 0.25 + 0.2222 + 0.1389 + 0.2222$$

$$\sum pk\ x\ qk = 0.972$$

$$KR\text{-}20 = \frac{5}{4}\ (1 - \frac{0.972}{2.70})$$

$$= 1.25\ (1 - 0.36)$$

$$= 1.25\ (0.64) = 0.80$$

ب- ألفا كرونباخ (Cronbach Alpha): يعتبر ألفا كرونباخ من أكثر المؤشرات الإحصائية شيوعًا لتقييم الاتساق الداخلي للمقاييس ويعد حالة خاصة من معادلة كيودر-ريتشارسون (Kuder-Richardson). يمكن استخدامه لكل من البنود الثنائية (نعم/لا) والبنود متعددة الخيارات كالمقاييس الترتيبية لليكرت (Likert). ويمثل معامل ألفا كرونباخ متوسط المعاملات الناتجة عن تجزئة الاختبار إلى طرق مختلفة وبذلك يعكس معامل الارتباط بين أي جزئين من أجزاء الاختبار. يتم حسابه عبر تحديد تباين كل مفردة من مفردات الاختبار ثم مجموع التباينات ومقارنته مع تباين الدرجة الكلية للاختبار مع اشتراط أن تقيس مفردات الاختبار سمة واحدة فقط. ولتوضيح كيفية تطبيقه عندنا مثلاً بيانات افتراضية لعينة تضم 14 مريضًا في مستشفى إعادة التأهيل تم تقييم وظائفهم باستخدام ستة بنود، وهي: المشي، صعود السلالم، حمل 5 كيلوغرام، الوصول إلى الهاتف، ارتداء القميص، والصعود والنزول من السيارة، مع مقياس من 1 إلى 5 لكل بند حيث يشير الرقم 5 إلى الاستقلالية الكاملة. وبالتالي يكون المجموع الكلي الأقصى هو 30.

يعكس الاتساق الداخلي درجة الارتباط بين هذه البنود الستة وكذلك ارتباط كل بند بالمجموع الكلي. في هذا المثال كانت قيمة ألفا كرونباخ تساوي 0.894 ما يشير إلى ثبات مرتفع للمقياس. تتراوح قيم ألفا كرونباخ بين القيمتين 0 و1، والقيم القريبة من 0.90 تعكس ثبات عالي. كما يمكن استخدام ألفا كرونباخ لفحص مدى توافق البنود الفردية مع المقياس ككل. أظهر بند "المشي" أعلى متوسط درجات بينما سجل بند "الصعود والنزول من السيارة" أدنى متوسط. وأظهرت معاملات

الارتباط بين البنود أن معظمها كانت أعلى من 0.60 باستثناء بند "الصعود والنزول من السيارة" الذي سجل ارتباطات منخفضة (0.354) أو أقل ما يشير إلى أنه قد يمثل بعدًا مختلفًا عن الوظائف الأخرى. ميزة أخرى لألفا كرونباخ وهي إمكانية إعادة حسابه بعد حذف أي بند لتقدير تأثيره على الثبات الكلي. عند إزالة بند "الصعود والنزول من السيارة" ارتفع معامل ألفا كرونباخ إلى 0.932 مما يدل على أن المقياس أصبح أكثر تجانسًا وبالتالي يُستحسن استبعاد هذا البند.

وتشير الأدبيات إلى أن المقياس ذو الاتساق الداخلي الجيد يجب أن تظهر بنوده ترابطًا متوسطًا بين 0.70 و0.90 فالقيم المنخفضة تشير إلى أن البنود قد تقيس سمات مختلفة والقيم العالية جدًا قد تعني تكرار البنود وتقلل صدق المقياس الكلي (Portney, et al, 2008, p.616). ويعطى ألفا كرونباخ بالمعادلة التالية:

$$Cofficient\ Alpha = \left(\frac{K}{K-1}\right)\left[1-\frac{\Sigma_{k=1}^{p} S_k^2}{S_{\text{Total}}^2}\right]$$

حيث:

K = العدد الكلي لمفردات المقياس.

Sk^2 = تباين المفردة K.

S^2= التباين الكلي للمقياس.

مثال تطبيقي:

لنفترض أن لدينا اختبارًا يتكون من خمس مفردات وقد طبق على ستة طلاب. يجيب كل طالب عن كل مفردة على مقياس ليكرت خماسي يتراوح من 1 إلى 5 ، حيث تمثل القيم الأعلى درجةً أكبر من السمة المقاسة)، سجلت استجابات الطلاب كما في الجدول الآتي:

جدول 13 معطيات لمثال لحساب معامل ألفـا كرونبـاخ

	المفردة1	المفردة2	المفردة3	المفردة4	المفردة5	المجموع
الطالب1	4	3	4	5	5	21
الطالب2	3	3	2	3	3	14
الطالب3	2	3	2	2	1	10
الطالب4	4	4	5	3	4	20
الطالب5	2	3	4	2	3	14

10	3	1	2	2	2	الطالب6
S^2= 22.57	**1.77**	**1.87**	**1.77**	**0.40**	**0.97**	***Sk^2***

- الخطوة1: التذكير بصيغة تباين العينة (Sample Variance):

$$\sigma^2 = \frac{\sum = (x_i - \bar{x})^2}{N - 1}$$

الخطوة2: إيجاد التباين لكل مفردة:

❖ حساب تباين المفردة1 (القيم:4؛ 3؛ 2؛ 4؛ 2؛ 2).

- المتوسط: (4 + 3 + 2 +4+ 2 + 2) / 6 = 17 / 6 = 2.833

- الانحرافات عن المتوسط:

2.833 – 4 = 1.167؛ 2.833 – 3 = 0.167؛ 2.833 – 2 = - 0.833؛ 2.833 – 4 = 1.167؛

2.833 – 2 = - 0.833؛ 2.833 – 2 = - 0.833.

- مربعاتها: 1.361؛ 0.028؛ 0.694؛ 1.361؛ 0.694؛ 0.694.

- مجموع المربعات: 4.833.

- تباين المفردة1 = 4.833 / (1-6) = 0.967.

❖ **وبنفس الطريقة نقوم بحساب بقية تباينات المفردات المتبقية والنتيجة في الجدول أعلاه.**

الخطوة3: إيجاد التباين الكلي لدرجات للطلبة:

❖ لحساب تباين المجموع الكلي للدرجات، نقوم بالتالي:

- المتوسط الكلي للمجاميع: (21 + 14+ 10 + 20 + 14 + 10) / 6 = 89 / 6 = **14.83**.

- حساب الانحرافات عن المتوسط ومربعاتها:

- الطالب 1: (14.83 - 21) ^{2} = (6.17 -)^2 = 38.07.

- الطالب 2: (14.83 - 14) ^{2} = (0.83)^2 = 0.69.

- الطالب 3: (14.83 - 10) ^{2} = (4.83)^2 = 23.33.

- الطالب 4: (14.83 - 20) ^{2} = (5.17 -)^2 = 26.73.

- الطالب 5: (14.83 - 14) ^{2} = (6.17 -)^2 = 0.69.

- الطالب 6: (14.83 - 10) ^{2} = (6.17 -)^2 = 23.33.

- حساب مجموع المربعات وقسمتها على (N-1): (38.07 + 0.69 + 23.33 + 26.73 + 0.69 + 23.33) / (6-1) = 112.84 / 5 = **22.57.**

الخطوة4: وبتطبيق معادلة ألفا كرونباخ نتحصل على النتائج التالية:

$$\boldsymbol{\alpha} = \frac{5}{4}(1 - \frac{0.97 + 0.40 + 1.77 + 1.87 + 1.77}{22.57})$$

$$\alpha = 1.25\ (1 - \frac{6.78}{22.57})$$

$$\alpha = 1.25\ (1 - 0.30)$$

$$\alpha = 1.25\ (0.70) = 0.875$$

- **التفسير:** يلاحظ أن قيمة ألفا كرونباخ تقريبا تساوي 0.88 وهذا يدل على ثبات داخلي جيد أي أن القيمة المرتفعة تشير إلى تماسك واتساق جيد بين مفردات الاختبار.

ويلاحظ أنه يتوجب على الشخص التأكد من صدق الاختبار أولاً ثم من ثباته لأن الاختبار الصادق دائما ثابت في حين أن الاختبار الثابت ليس من الضروري أن يكون صادقاً كأن تعطى علامات الطلبة حسب الطول حيث يتضمن ذلك الثبات ولا يتضمن الصدق وإذا لم يكن الاختبار صادقاً أو ثابتاً يتوجب على واضعه إعادته أو تعديله ويتم التأكد من صدق الثبات بتطبيق لاختبار على عينة استطلاعية من خارج عينة الدراسة.

ويلاحظ كذلك أنه تقبل معاملات ثبات منخفضة نسبياً لاختبارات الاتجاهات والميول مثلاً وبشكل عام للمقاييس الانفعالية في حين يتوجب الحصول على معاملات ثبات عالية نسبياً للاختبارات التحصيلية.

3-3- العوامل المؤثرة على ثبـات الاختبـار:

يتأثر معامل ثبات الاختبار بعدة عوامل أساسية تؤثر على دقة وموثوقية القياس، ومن أبرزها:

أ- طـول الاختبـار: يزداد معامل الثبات بزيادة عدد مفردات الاختبار، فالاختبارات الأطول تشمل محتوى أوسع وتغطي الجوانب المختلفة للخاصية المقاسة ما يؤدي إلى رفع صدق المحتوى وتحسين التقديرات الإجمالية للثبات.

ب- تباين درجـات الأفراد: يرتبط معامل الثبات بتنوع الدرجات داخل عينة المفحوصين، كلما كانت المجموعة أكثر تجانسًا يقل التباين في الدرجات مما يقلل من قيمة معامل الثبات. بالمقابل زيادة تباين درجات الأفراد تعزز قدرة الاختبار على التمييز بين مستوياتهم المختلفة وبالتالي تزيد من الثبات.

ج- مستوى القـدرة في مجموعة الاختبـار: ترتبط دقة الاختبار بمستوى القدرة لدى المفحوصين إذ يختلف أداء الاختبار حسب مستوى الصعوبة للمفردات. لذلك ينصح بأن يحتوي الاختبار على أغلبية من المفردات متوسطة الصعوبة لتمييز المستويات المتوسطة للقدرة بدقة إلى جانب عدد محدود من المفردات السهلة جدًا والصعبة جدًا لضمان شمولية التقدير عبر نطاق واسع من القدرات.

بالإضافة إلى ذلك يمكن القول إن تصميم الاختبار بعناية للتوازن بين طول الاختبار وتنوع صعوبته مع مراعاة الخصائص الديموغرافية والعقلية للمجموعة المستهدفة مما يسهم في تحسين الثبات العام للاختبار وضمان موثوقية النتائج في الاستخدامات المختلفة.

الدكتور زيـاد رشيـد

المحـــاضرة التاسعة

نمـــاذج من بطاريـــات التقييـــم والقيـــاس

الأهداف:

في إطار هذه المحاضرة، يهدف المحتوى إلى تمكين الطلبة من التعرف على نماذج متعددة من بطاريات التقييم والقياس في مجال القدرات العقلية، وفهم الأسس النظرية التي تستند إليها هذه المقاييس وكيفية تطورها عبر الزمن. كما يسعى إلى توضيح مقياس ستانفورد–بينيه بوصفه أحد المقاييس الكلاسيكية في قياس الذكاء، مع إبراز مفهوم العمر العقلي ودوره في تفسير الأداء العقلي، والتعرف على مقياس وكسلر للذكاء (النسخة الرابعة) من حيث بنيته ومجالاته اللفظية والأدائية وأبعاده المعرفية المختلفة، إضافة إلى التعرف على اختبار المصفوفات المتتابعة لرافن بوصفه اختبارًا غير لفظي يقيس التفكير المجرد ويتميز بخصائص عابرة للثقافات، وكذلك اختبار أوتس–لينون للقدرة العامة من حيث دوره في قياس الاستعداد الدراسي والقدرات الاستدلالية لدى الطلبة، وأخيرًا مقياس إلينوي للقدرات السيكو-لغوية باعتباره أداة متخصصة في تشخيص العمليات اللغوية والإدراكية لدى الأطفال، بما يتيح تكوين رؤية شاملة حول خصائص هذه البطاريات واستخداماتها في مجالات التشخيص والتوجيه التربوي.

محتوى المحاضرة:

أولاً- نماذج من اختبارات القدرات العقلية

1- مقياس استانفورد – بينيه (Stanford-Binet Scale)

1-1- مفهوم العمر العقلي (Mental Age Concept)

2- مقياس وكسلر للذكـاء النسخة الرابعـة (Wechsler Intelligence Scale- WISC_1V)

3- اختبـار المصفوفـات المتتابعـة لرافن (Raven)

4- اختبار أوتس – لينون للقدرة العامة ((Otis-Lennon Mental Ability Test (OLMAT))

5- مقياس الينوي للقدرات السيكولغوية (Illinois Test of Psycholinguistic Abilites)

مدة المحاضرة: ثلاث ساعات

تمهيـد:

يعد قياس القدرات العقلية من المجالات الأساسية في علم النفس التربوي، لما له من دور محوري في فهم الفروق الفردية بين المتعلمين، وتوجيه عمليات التشخيص والتقويم واتخاذ القرار التربوي المناسب. وفي هذا السياق، ظهرت مجموعة من البطاريات والمقاييس النفسية التي سعت إلى تقديم تقدير دقيق للقدرات المعرفية بمختلف أبعادها، سواء اللفظية أو غير اللفظية أو اللغوية أو الاستدلالية. وقد تطورت هذه الأدوات عبر مراحل تاريخية متتابعة، انعكست فيها التحولات النظرية في فهم الذكاء وبنيته، إضافة إلى التقدم في أساليب القياس والتقنين. ومن هنا تأتي أهمية التطرق إلى نماذج من بطاريات التقييم والقياس، كمقياس ستانفورد–بينيه، ومقياس وكسلر للذكاء، واختبار المصفوفات المتتابعة لرافن، واختبار أوتس–لينون للقدرة العامة، ومقياس إلينوي للقدرات السيكو-لغوية، وذلك لفهم خصائصها وأسسها النظرية ومجالات استخدامها في الممارسة التربوية والإكلينيكية.

أولاً- نماذج من اختبارات القدرات العقلية:

1- مقياس استانفورد – بينيه (Stanford-Binet Scale):

صدرت الصورة المعدلة من مقياس ستانفورد–بينيه على يد ثورندايك وهيجن وساتلير (Thorndike; Hagen & Sattler, 1986)، حيث أدخل هؤلاء تعديلات جوهرية مقارنة بالصورتين الثانية والثالثة. وقد تمثلت هذه التعديلات في إعادة تنظيم الفقرات وفق استراتيجية اعتمدت على اختيار عينات متنوعة وواسعة من المهام المعرفية، بحيث تغطي مجالات متعددة تتطلب أنماطًا مختلفة من التفكير الارتباطي في سياقات متنوعة، بما يسهم في الكشف عن القدرة العقلية العامة العامل العام (g). وتمثل هذه النسخة الصورة الرابعة من المقياس، والتي قام مليكة (1998) بإعدادها وتقنينها في البيئة العربية. ويعد هذا المقياس اختبارًا فرديًا للذكاء يتكون من (15) اختبارًا فرعيًا (بركات، 2018). وتندرج هذه الاختبارات ضمن أربعة مجالات معرفية رئيسية هي: الاستدلال اللفظي، والاستدلال المجرد البصري، والاستدلال الكمي، والذاكرة قصيرة المدى، حيث تسفر عن درجة مركبة كلية تعكس القدرة الاستدلالية العامة (العامل العام). وقد استند اختيار هذه المجالات إلى نموذج هرمي ثلاثي المستويات لبنية القدرات المعرفية، يتمثل في:

المستوى الأول: العامل العام (g).

المستوى الثاني: ويشمل القدرات المتبلورة (Crystallized Abilities)، والقدرات السائلة التحليلية (Fluid/Analytic Abilities)، والذاكرة قصيرة المدى (Short-Term Memory).

المستوى الثالث: ويتضمن مجالات أكثر تخصصًا، هي: الاستدلال اللفظي، والاستدلال الكمي، والاستدلال المجرد البصري.

يوفر المقياس عشرين درجة عمرية معيارية، منها (15) درجة للاختبارات الفرعية، و(4) درجات للمجالات الأساسية، بالإضافة إلى درجة مركبة كلية. كما تتوافر معايير مختلفة لطرق تجميع الاختبارات الفرعية. ويتضمن المقياس كذلك كراسة إجابة تستخدم لتسجيل بيانات المفحوص ودرجاته في الاختبارات الفرعية والدرجات المركبة، إلى جانب الملاحظات الإكلينيكية المرتبطة بالعوامل المؤثرة في أدائه، مستعينًا بمقاييس تقدير مدونة في صفحة الغلاف، فضلًا عن صفحة تحليلية تمكن الفاحص من الانتقال من مستوى الأداء إلى تفسير خصائص وقدرات المفحوص (مليكة، 1998). ويعد مقياس ستانفورد–بينيه في صورته الرابعة من أبرز أدوات قياس الذكاء، لما أحدثه من نقلة نوعية في تقويم القدرات المعرفية، نتيجة ما شهده من تطوير في محتوى المهام، وتنوعها، وأساليب عرضها، إضافة إلى تحديث المعايير المعتمدة، واعتماده على إطار نظري متكامل. وقد أسهم ذلك في تعزيز مكانته في مجال القياس النفسي، سواء على الصعيد النظري أو التطبيقي، وجعله من أكثر المقاييس استخدامًا في التقييم الإكلينيكي على مستوى العالم.

ولا يقتصر هدف المقياس على قياس الذكاء فحسب، بل يمتد ليشمل التعرف على فئات خاصة من التلاميذ، مثل ذوي الإعاقة العقلية، وذوي صعوبات التعلم، وكذلك الموهوبين. كما يُستخدم في الدراسات الطولية لمتابعة تطور القدرات المعرفية أو تدهورها عبر مراحل العمر، بدءًا من سن سنتين وحتى ما بعد السبعين عامًا. وقد أدى هذا التنوع في الأهداف إلى تنوع محتوى الاختبارات الفرعية، بما يتيح تكوين صورة شاملة للنشاط العقلي للفرد. وتصنف درجات الذكاء وفق المعايير التي وضعها Thorndike و Hagen و Sattler (1986) على النحو الآتي:

- (67) فأقل: إعاقة عقلية.
- (68-78): بطيء التعلم.
- (79-88): دون المتوسط.
- (89-110): متوسط.
- (111-120): فوق المتوسط.
- (121-131): مرتفع (ممتاز).
- (132) فأكثر: مرتفع جدًا.

يتكون المقياس من أربعة مجالات رئيسية لقياس القدرة المعرفية، حيث يندرج تحت كل مجال مجموعة من الاختبارات الفرعية، وذلك على النحو الآتي:

- **الاستدلال اللفظي:** ويشمل اختبارات المفردات، والفهم، والسخافات، والعلاقات اللفظية.
- **الاستدلال المجرد البصري:** ويتضمن تحليل النمط، والنسخ، والمصفوفات، وثني قطع الورق.
- **الاستدلال الكمي:** ويضم الاختبار الكمي، وسلاسل الأعداد، وبناء المعادلات.
- **الذاكرة قصيرة المدى:** وتشمل ذاكرة الخرز، وذاكرة الجمل، وذاكرة الأشياء، وذاكرة الأرقام.

1-1- مفهوم العمر العقلي (Mental Age Concept):

ترجع جذور مفهوم العمر العقلي إلى فترة سابقة لجهود ألفريد بينيه، إلا أنه ظهر آنذاك بصورة غير دقيقة وغير محددة إجرائيًا. ويُعد الطبيب الفرنسي أسكويرول (Esquirol) من أوائل من أشاروا إلى هذا المفهوم عام (1838)، حيث ربط بين التخلف العقلي وعدم قدرة الفرد على اكتساب المعارف العامة بالمستوى الذي يبلغه أقرانه في نفس العمر الزمني. ومع ذلك، لم يتمكن من تحديد هذا المفهوم بشكل موضوعي أو إجرائي يعتمد على أدوات قياس دقيقة. كما ظهرت محاولات لاحقة لدى دنكان (Duncan) وميلارد (Millard) عام (1866)، حيث قاما بمقارنة الطفل المتخلف عقليًا بطفل عادي أصغر منه سنًا، وهو ما يعكس جوهر فكرة العمر العقلي. غير أن هذه المحاولات بقيت محدودة لغياب أدوات قياس موضوعية تمكن من إجراء هذه المقارنة بشكل دقيق.

وقد تحقق التحديد الإجرائي الدقيق لمفهوم العمر العقلي على يد ألفريد بينيه، خاصة في مراجعته لمقياس بينيه–سيمون عام (1908)، والتي أكدها لاحقًا في نسخة (1911). ومنذ ذلك الحين، أصبح العمر العقلي مفهومًا أساسيًا في اختبارات الذكاء، سواء من حيث كونه معيارًا أو طريقة للتفسير.

وفيما بعد، قام تيرمان بتطوير هذا المفهوم وربطه بحساب نسبة الذكاء(IQ) ، حيث أتاح ذلك إمكانية التعبير عن مستوى الذكاء بصورة كمية (Delany & Hopkins, 1987).

1-2-الصورة المعدلــة من المقياس:

يتكون المقياس في صورته المعدلة من (129) اختبارًا، وهو يتشابه في طبيعته العامة مع الصورة السابقة من حيث نوعية الاختبارات، إذ يشتمل على مهام في الفهم، والسخافات، ورسم الأشكال، وإعادة الأرقام، والتفريق بين أوجه الشبه والاختلاف، إضافة إلى تعريف الكلمات المجردة.

ويبدأ تطبيق المقياس من عمر سنتين، ويمتد حتى مرحلة الرشد (18 سنة)، مع إدخال تحسينات مهمة شملت دقة تعليمات التطبيق، وتطوير معايير التصحيح، إضافة إلى إعادة تقنينه على عينات أكبر وأكثر تمثيلًا، بما يعزز من دقته وموثوقيته (Thorndike; Hagen & Sattler, 1986).

3-1- الحالات التي يصعب فيها تطبيق مقياس ستانفورد–بينيه:

على الرغم من الأهمية الكبيرة التي يتمتع بها مقياس ستانفورد–بينيه في مجالي القياس النفسي والتربوي، إلا أن هناك بعض الفئات التي يواجه فيها التطبيق صعوبات تحد من دقته أو من ملاءمته، ومن أبرز هذه الحالات ما يأتي:

أ- المكفوفون: تعد صعوبة الاعتماد على المواد البصرية في عدد من اختبارات المقياس من أبرز التحديات التي تعيق تطبيقه على هذه الفئة.

ب- الصم: يمكن الاستفادة من بعض الاختبارات الأدائية غير اللفظية، كذاكرة الخرز وتحليل النمط، كما يمكن تحديد مستوى البداية (المدخل) بالاعتماد على اختبار ذاكرة الخرز.

ج- الأفراد المصابون بالشلل الدماغي: يمكن تطبيق المقياس في هذه الحالة، إلا أن نتائج القياس قد تتأثر بدرجة الإعاقة الحركية وما يصاحبها من صعوبات في الأداء، مما قد ينعكس سلبًا على صدق النتائج. لذلك، يوصى بالتركيز على اختيار الاختبارات الفرعية التي تتناسب مع قدرات الفرد وظروفه الخاصة.

د- الأطفال ذوو اضطراب التوحد: في الحالات التي يعاني فيها الطفل من صعوبات لغوية أو عدم القدرة على التعبير اللفظي، يُفضل الاعتماد على المقاييس الأدائية أو الاختبارات العملية، مثل المتاهات، أو اختيار الاختبارات الفرعية غير اللفظية المناسبة.

2- مقياس وكسلر للذكـاء النسخة الرابعة (Wechsler Intelligence Scale- WISC_1V):

أعد هذا المقياس العالم وكسلر (Wechsler) لقياس الذكاء للصغار (6 – 12) سنة وللكبار سنة. وقد صدر عنه صور عديدة أحدثها الطبعة الرابعة لعام (2003)، وهو مقياس من النوع الفردي ويحتاج تطبيقه ما يقارب (60 – 90) دقيقة، ويحصل المفحوص عليه على درجة كلية بالإضافة إلى (4) درجات فرعية ودرجتين مركبتين. هذا ولم يعد مقياس وكسلر مجرد مقياس يقدم درجات معيارية للذكاء فحسب، بل أصبح في نفس الوقت أداة سريرية تفحص المناطق الحساسة في الوظيفة المعرفية لدى الطفل، ويقدم الجيل الرابع من هذا المقياس إمكانيات أكبر للتقييم مع مختلف الاختلافات الثقافية،

وفي نفس الوقت يسعى إلى سبر أغوار المناطق المعرفية المسؤولة عن الصعوبات التعليمية لدى الأطفال (علي ومصطفى، 2012، ص87).

يعد اختبار وكسلر من المقاييس المتكاملة، إذ بني في ضوء أحدث الاتجاهات في علم النفس المعرفي، مع مراعاة الأبعاد الثقافية في عملية التقييم. كما طور ليشمل نطاقًا أوسع من القدرات التي تناولتها النسخ السابقة، مع إتاحة إمكانية تعديل وتحسين المهام الاختبارية، بما يسمح بتقديم تقييم أكثر عمقًا لقدرات الطفل. ويسهم هذا التطوير في تعزيز فهم قابلية الطفل للتحسن، إلى جانب تمكين الفاحص من تحديد جوانب القوة والضعف في العمليات المعرفية، والكشف عن الفروق الفردية المؤثرة في الأداء على مهام حل المشكلات. ويمهّد ذلك لتقديم تدخلات ذات دلالة إكلينيكية تسهم في تخطيط استراتيجيات تعليمية مناسبة (الخطيب، 2011).

2-1- الخصائص العامة لاختبار وكسلر: يتميز اختبار وكسلر بعدد من الخصائص، من أبرزها:

- تطوير أساليب تقييم التفكير السائل، والذاكرة العاملة، وسرعة معالجة المعلومات.
- تعزيز الدقة في التشخيص الإكلينيكي.
- التأكيد على أهمية التعزيز أثناء عملية التقييم.
- تحسين الخصائص السيكومترية للمقياس.
- تطوير مستويات البداية (القاعدية) ومستويات السقف في الاختبارات الفرعية .
- اعتماد معايير حديثة مستمدة من عينة قومية أمريكية وفق إحصاء عام (2000).

كما تضمن المقياس تحديث المواد الاختبارية واستبدال القديمة منها بأخرى أكثر ملاءمة، وإفراد فصل خاص لتفسير النتائج، إضافة إلى الاستغناء عن بعض الاختبارات الفرعية، مثل: المتاهات، وتركيب الأجسام، وترتيب الصور، وكذلك حذف بعض الاختبارات الإضافية كالمعلومات، والتفكير اللفظي، والحساب، وتكميل الصور. ويتميز المقياس بوضوحه الثقافي، وخفة أدواته، وإطالة عمر استخدامها.

2-2- إجراءات الصدق والثبات:

تم تقنين اختبار وكسلر على عينة ممثلة للمجتمع الأمريكي استنادًا إلى بيانات التعداد السكاني لعام (2000)، حيث بلغت العينة (2200) طفل تتراوح أعمارهم بين (6–12) سنة، وقد روعي في اختيارها التوازن من حيث العمر، والجنس، والمستوى التعليمي للوالدين، والمنطقة الجغرافية، والانتماء الثقافي والعرقي. وقد جمعت بيانات الصدق والثبات باستخدام مجموعة واسعة من الأدوات

المقننة، إضافة إلى تطبيقه على (16) فئة خاصة من الأطفال. كما أُجريت دراسات مقارنة ضمن سلسلة مقاييس وكسلر عبر المراحل العمرية المختلفة. وتضمن الدليل الفني معاملات الثبات (الاتساق الداخلي، وإعادة الاختبار، ومعاملات الثقة)، إلى جانب معاملات الارتباط مع مقاييس أخرى. كما تم التحقق من صدق المقياس من خلال مقارنته بمحكات متعددة، كاختبارات التحصيل، والذاكرة، والسلوك التكيفي، والاختبارات الانفعالية، واختبارات الموهبة، حيث أظهرت النتائج وجود علاقات ارتباطية دالة (عليان والكيلاني، 1988).

2-3- مجالات الاختبار (الاختبارات الفرعية الحديثة):

أُضيفت مجموعة من الاختبارات الفرعية التي تتماشى مع التطورات الحديثة في علم النفس المعرفي والممارسة الإكلينيكية، من أبرزها:

أ- استنتاج الكلمة: يقيس القدرة اللفظية من خلال التعرف إلى مفاهيم تُعرض في صورة تلميحات متتابعة.

ب- استنتاج المصفوفة: يقيس التفكير السائل عبر اختيار الشكل المناسب لإكمال نمط معين.

ج- مفاهيم الصورة: يقيس التفكير المجرد والتنظيم الإدراكي والقدرة على التصنيف دون الاعتماد على الاستجابة اللفظية.

د- سلسلة عدد–حرف: يقيس الذاكرة العاملة من خلال ترتيب الأرقام والحروف وفق قواعد محددة.

هـ- الإلغـاء: يقيس سرعة المعالجة من خلال تمييز أشكال محددة ضمن مجموعة من المثيرات.

و- الدرجات في اختبـار وكسلر: يقدم المقياس مجموعة من الدرجات الفرعية والمركبة، إضافة إلى الدرجة الكلية، وذلك على النحو الآتي (معوض، 2014):

أ- الدرجـات الفرعيـة:

- الدرجة اللفظيـة.
- الدرجة الإدراكية–الحسيـة.
- درجة الذاكرة العاملــة.
- درجة سرعة المعالجـة.

ب- الدرجـات المركبـة:

- **القدرة العامة:** تمثل حاصل الجمع بين المقياسين اللفظي والإدراكي–الحسي، وتعكس مستوى الفهم اللفظي والذكاء الإدراكي.

- **المهارة المعرفية:** تمثل مجموع درجتي الذاكرة العاملة وسرعة المعالجة، وتعكس كفاءة المعالجة المعرفية.

- **الدرجة الكلية:** تمثل مجموع درجات جميع الاختبارات الفرعية، وهي حصيلة الدرجتين المركبتين.

2-4- أمثلة من اختبارات وكسلر (الإصدار الرابع) للأطفال:

يتكون مقياس وكسلر لذكاء الأطفال من قسمين رئيسيين، يضم كل منهما مجموعة من الاختبارات الفرعية التي تقيس جوانب مختلفة من القدرات المعرفية، وذلك على النحو الآتي:

- **القسم اللفظـي (Verbal Part):** يهدف هذا القسم إلى قياس القدرات اللغوية والفهم اللفظي، ويشتمل على الاختبارات الفرعية التالية:

أ- اختبار المعلومات (Information Test): يتضمن مجموعة من الفقرات (حوالي 30 فقرة) تقيس المعرفة العامة، حيث يطلب من المفحوص الإجابة عن أسئلة متنوعة.

ب- اختبار المتشابهات (Similarities Test): يقيس القدرة على التفكير المجرد من خلال تحديد أوجه الشبه بين أزواج من الأشياء.

ج- اختبار المفردات (Vocabulary Test): يهدف إلى قياس الحصيلة اللغوية من خلال تعريف كلمات معطاة.

د- اختبار الاستيعاب (Comprehension Test): يقيس الفهم الاجتماعي والقدرة على الحكم من خلال الإجابة عن مواقف حياتية وأسئلة تفسيرية.

هـ- اختبار الحساب (Arithmetic Test): يقيس القدرة على التفكير العددي وحل المشكلات الحسابية ذهنيًا.

و- اختبار إعادة الأرقام (Digit Span Test): يقيس الذاكرة السمعية قصيرة المدى من خلال إعادة سلاسل رقمية بالترتيب المباشر والعكسي.

- **القسم الأدائـي (Performance Part):** يركز هذا القسم على القدرات الإدراكية-البصرية والمهارات الحركية، ويتضمن الاختبارات الفرعية الآتية:

أ- اختبـار تكميل الصور (Picture Completion Test): يقيس القدرة على إدراك التفاصيل من خلال التعرف على الجزء الناقص في الصور.

ب- اختبـار ترتيب الصور (Picture Arrangement Test): يقيس القدرة على الفهم التسلسلي من خلال ترتيب صور تعكس أحداثًا منطقية.

ج- اختبـار تصميم المكعبـات (Block Design Test): يقيس التنظيم البصري–المكاني من خلال تكوين أشكال باستخدام مكعبات ملونة وفق نموذج معين.

د- اختبـار تجميع الأشيـاء (Object Assembly Test): يقيس القدرة على إدراك العلاقات الكلية من خلال تجميع أجزاء لتكوين شكل ذي معنى.

هـ- اختبـار الترميز (Coding Test): يقيس سرعة المعالجة والتآزر البصري–الحركي من خلال تحويل رموز إلى أشكال وفق مفتاح محدد.

و- اختبـار المتاهـات (Mazes Test): يقيس التخطيط والتنظيم البصري من خلال تتبع المسار الصحيح داخل متاهات ضمن زمن محدد.

3- اختبـار المصفوفـات المتتابعـة لرافن (Raven):

اتسمت الاختبارات التي سبقت ظهور مقياس بينيه بتركيزها على قياس العمليات العقلية الأولية، كالتمييز الحسي وسرعة الاستجابة. غير أن انتشار مقياس بينيه أسهم في ترسيخ أهمية قياس الذكاء باعتباره بعدًا أساسيًا في دراسة الفروق الفردية. ومع ذلك، برزت إشكالية تطبيق هذه الاختبارات على بعض الفئات الخاصة، كالصم، والبكم، والأميين، فضلًا عن الأفراد الذين لا يجيدون لغة الاختبار، مما دفع الباحثين إلى تطوير أدوات قياس غير لفظية تقلل من أثر اللغة والثقافة. وفي هذا السياق، ظهرت مجموعة من الاختبارات غير اللفظية، كاختبار المتاهات، واختبار تكملة الصور، ولوحات الأشكال، إضافة إلى اختبار المصفوفات المتتابعة الذي طوره رافن، والذي يعد من أبرز هذه الأدوات.

ويتوافر اختبار رافن في عدة مستويات، ويعد من الاختبارات التي طورت لتكون صالحة للاستخدام عبر ثقافات متعددة. فقد أعده رافن لأول مرة عام (1947)، ثم خضع للتعديل عام (1956)، ويطبق على فئات عمرية متنوعة تشمل الأطفال العاديين وذوي التأخر العقلي، وكذلك كبار السن الذين تتراوح أعمارهم بين (65–85) سنة. ويتميز هذا الاختبار بكونه من الاختبارات العابرة للثقافات (Cross-Cultural Tests)، إذ يهدف إلى تقليل تأثير العوامل اللغوية والثقافية في أداء المفحوص. ومع ذلك، يشير رافن إلى أهمية استخدام اختبارات لفظية مرافقة لاختبار المصفوفات، وذلك للحصول على صورة أكثر شمولًا للنشاط العقلي للفرد. ويركز هذا الاختبار على قياس القدرة على إدراك العلاقات المكانية وتنظيمها، كما يستند في بنائه إلى نظرية العاملين لسبيرمان، حيث أظهرت الدراسات أن أداء الأفراد عليه يتشبع بدرجة كبيرة بالعامل العام للذكاء (Raven, 1977).

وتتوفر حاليًا ثلاثة مستويات رئيسية من مصفوفات رافن المتتابعة، وهي:

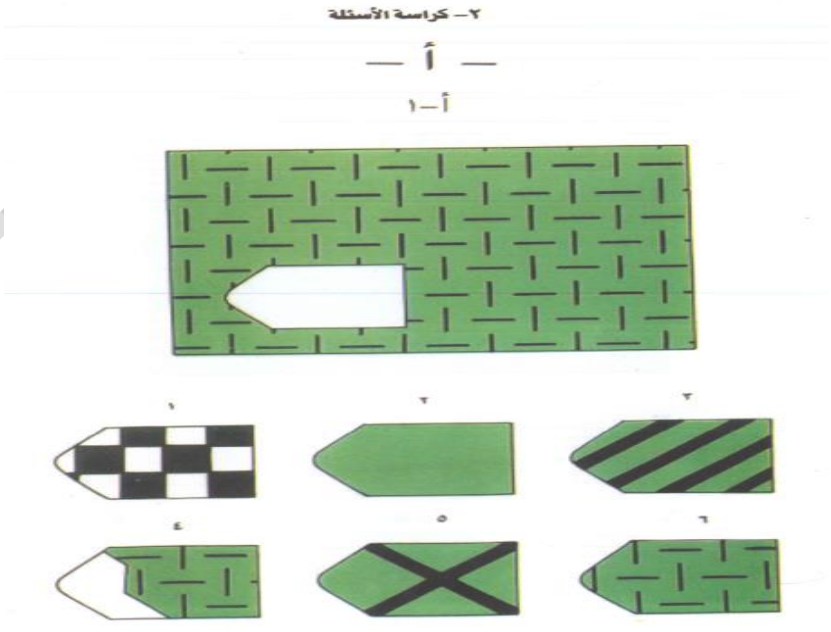
أ- المصفوفـات المتتابعـة العاديــة (Standard Progressive Matrices - SPM): تعد الصورة الأساسية للاختبار، وتتكون من (60) بندًا موزعة على خمسة أقسام (أ، ب، ج، د، هـ)، وتناسب الفئة العمرية من 6 إلى 60 سنة. وقد ظهرت هذه الصورة لأول مرة عام 1938، ثم أُجريت عليها تعديلات عام 1956.

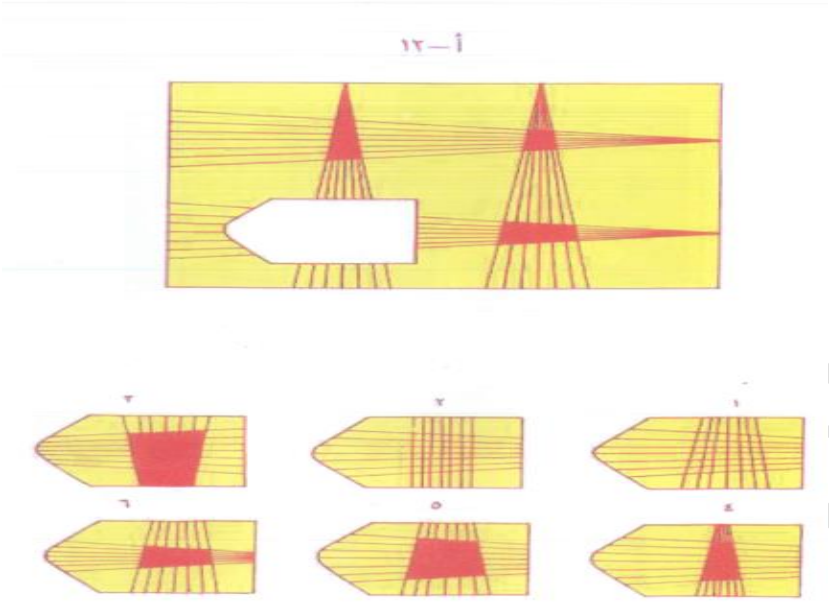
ب- المصفوفـات المتتابعـة الملونـة (Coloured Progressive Matrices - CPM): تتكون من (36) بندًا موزعة على ثلاثة أقسام (أ، أب، ب)، وتناسب الأطفال الذين تتراوح أعمارهم بين 5.5 و11 سنة، كما تُستخدم مع الأفراد ذوي التأخر العقلي وكبار السن. وقد ظهرت هذه النسخة عام 1947، وخضعت للتعديل عام 1956.

ج- المصفوفـات المتتابعـة المتقدمــة (Advanced Progressive Matrices - APM): أُعدت هذه الصورة عام 1941 لقياس القدرات العليا لدى الأفراد ذوي الذكاء المرتفع، وبدأ استخدامها بشكل محدود عام 1947. وكانت في صورتها الأولى تتكون من (48) بندًا، ثم عدلت عام 1962 ليصبح عدد بنودها (36) بندًا.

ويطلق على هذه السلسلة بمستوياتها الثلاثة اسم مصفوفات رافن المتتابعة (Raven's Progressive Matrices - RPM). ويقتصر التركيز في هذا السياق على المصفوفات المتتابعة الملونة (CPM)، نظرًا لملاءمتها لفئات عمرية وقدرات معرفية محددة.

نماذج من فقرات اختبار رافن:





4- اختبار أوتس – لينون للقدرة العامة (Otis-Lennon Mental Ability Test (OLMAT)):

يعد هذا الاختبار أحد المستويات ضمن سلسلة اختبارات أوتس–لينون للقدرة العقلية، وقد صمم لقياس القدرات العقلية لدى طلبة المرحلة المتوسطة الذين تتراوح أعمارهم بين (11–16) سنة. ويهدف إلى تقييم مجموعة من القدرات المرتبطة بالتحصيل الدراسي، من خلال قياس المهارات المعرفية التي تسهم في التعلم والنجاح الأكاديمي. ويصنَّف الاختبار ضمن اختبارات التفكير المجرد وحل المشكلات، حيث تشير نتائج الدراسات إلى وجود علاقة ارتباطية بين أداء الطلبة عليه ومستوى تحصيلهم الدراسي. كما يبين مركز الاختبارات الأمريكي أن هذا الاختبار يقيس قدرة الطلبة على اتباع التعليمات، والتفكير المنطقي، والتحليل، والتذكر، إضافة إلى توظيف الخبرات السابقة في مواقف جديدة.

ويستخدم الاختبار كذلك في الكشف عن الطلبة الموهوبين، إذ يطبق إلى جانب مجموعة من المقاييس الأخرى، مثل مقياس ستانفورد–بينيه، بهدف تصنيف الطلبة ضمن فئات الموهوبين. كما يُستفاد من نتائجه في تصنيف الطلبة وفق مستوى تمكنهم من المقررات الدراسية واجتيازهم لها (Otis & Lennon, 1996).

ويتكون الاختبار من جزأين: لفظي وغير لفظي. أما النسخة العربية التي أعدها وقننها مراد وعبد الغفار عام (1985) في البيئة المصرية، فتتضمن (80) فقرة متنوعة، صيغت جميعها في شكل اختيار من متعدد يتضمن خمسة بدائل للإجابة. وتتدرج البنود من السهولة إلى الصعوبة، ويخصص

للاختبار زمن يقارب (50) دقيقة، ويمكن تطبيقه في الحصة الدراسية العادية. كما تتوافر له أوراق إجابة منفصلة يمكن تصحيحها يدويًا أو آليًا باستخدام بطاقات خاصة. وتصنف فقرات الاختبار في أربعة مجالات رئيسية هي (Otis & Lennon, 2003):

أ-الاستيعاب اللفظـي (Verbal Comprehension): ويشمل التعاريف، والألفاظ المترادفة والمتضادة، وإكمال الجمل، وفهم الجمل الغامضة .

ب-الاستدلال اللفظـي (Verbal Reasoning): ويتضمن تكوين الكلمات، والقياسات اللفظية، والاستدلالات، والاختيار المنطقي.

ج- الاستدلال الشكلـي (Figural Reasoning): ويشمل القياسات الشكلية، وسلاسل الإكمال، ونماذج التراكيب .

د- الاستدلال الكمـي (Quantitative Reasoning): ويتضمن التسلسلات العددية والمسائل الحسابية.

4-1-الهدف من الاختبار:

تستند فكرة اختبار أوتيس–لينون للذكاء إلى كل من نظرية فيرنون (Vernon) وبيرت (Burt) الهرمية للقدرات العقلية، حيث تُعنى سلسلة اختبارات أوتيس–لينون بقياس الجوانب اللفظية للقدرات العقلية، إضافة إلى القدرات المرتبطة بالتعلم المدرسي والنجاح الأكاديمي، فضلاً عن قياس القدرة على التحصيل والتفكير المجرد.

ويتمثل الهدف الأساسي من هذا الاختبار في توفير أداة معيارية تتيح تقدير القدرة العقلية العامة أو الاستعداد الدراسي العام لدى طلبة المرحلة المتوسطة. ويركز هذا المستوى من الاختبار بشكل خاص على قياس القدرة على الاستدلال، من خلال تضمين مجموعة متنوعة من الفقرات اللفظية والرمزية والشكلية التي تمثل طيفًا واسعًا من القدرات المعرفية.

وقد صممت هذه السلسلة من الاختبارات بهدف الوصول إلى قياس دقيق وموثوق للعامل العام للقدرة العقلية، بحيث تعكس الدرجة الكلية التي يحصل عليها الطالب أداءه عبر مختلف أنواع البنود والمهام المعرفية المتنوعة (مراد وعبد الغفار، 1985).

5- مقياس الينوي للقدرات السيكولغوية (Illinois Test of Psycholinguistic Abilites):

ظهر مقياس إلينوي للقدرات السيكو-لغوية (Illinois Test of Psycholinguistic Abilities) الذي وضعه كريك ومكارثي عام (1961)، ثم جرى تطويره لاحقًا عام (1968). ويهدف هذا المقياس

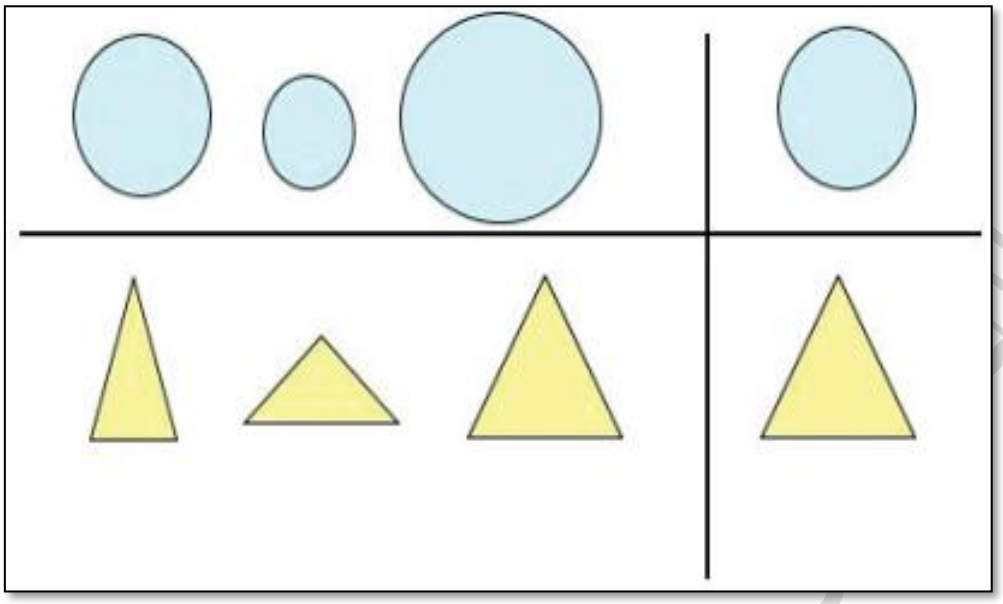
إلى قياس وتشخيص مظاهر الاستقبال والتعبير اللغوي، ولا سيما لدى الأطفال ذوي صعوبات التعلم، ويعد من المقاييس الفردية المقننة والشائعة الاستخدام في هذا المجال.

ويطبق المقياس على الفئة العمرية من (2–10) سنوات، ويستغرق تصحيحه ما يقارب (30–40) دقيقة، في حين يستغرق تطبيقه نحو ساعة ونصف تقريبًا. ويتكون من (12) اختبارًا فرعيًا تغطي أنماط الاتصال اللغوي ومستوياته المختلفة، إضافة إلى العمليات النفسية والمعرفية المرتبطة بها، وذلك على النحو الآتي:

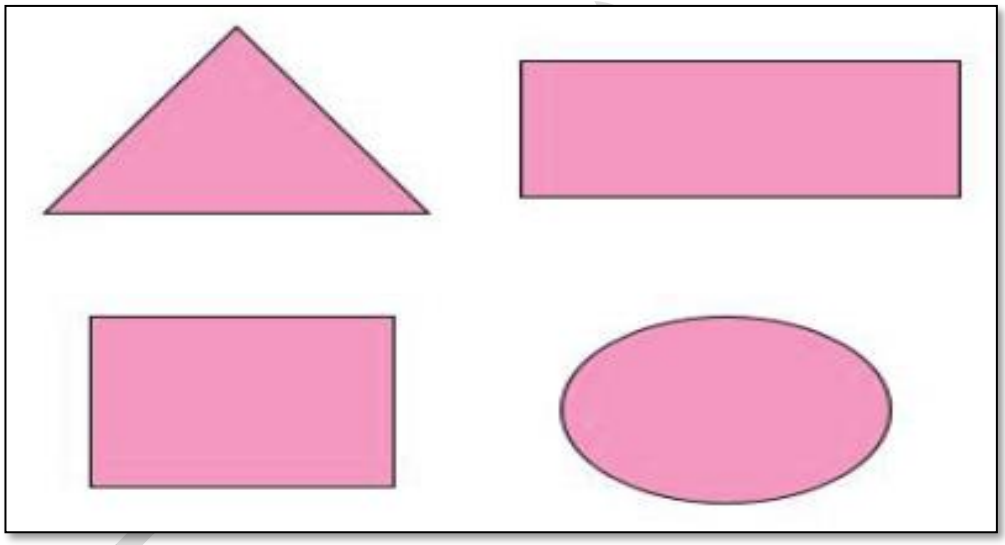
- **اختبار الاستقبال السمعي:** يقيس قدرة الطفل على فهم المثيرات السمعية والاستجابة لها بإجابات بسيطة مثل نعم أو لا.
- **اختبار الاستقبال البصري:** يقيس القدرة على مطابقة الصور ذات العلاقة بالمفاهيم المعروضة.
- **اختبار الترابط السمعي:** يقيس قدرة الطفل على إكمال جمل متجانسة من الناحية اللغوية.
- **اختبار الترابط البصري:** يقيس القدرة على الربط بين المثيرات البصرية ذات العلاقة أو المتشابهة.
- **اختبار التعبير اللفظي:** يقيس قدرة الطفل على التعبير الشفهي عن المفاهيم والأشياء المطلوبة.
- **اختبار التعبير العملي:** يقيس القدرة على التعبير عن الأفكار من خلال الأداء العملي أو الحركي.
- **اختبار الإكمال القواعدي:** يقيس القدرة على إكمال جمل وفق قواعد لغوية صحيحة ومترابطة.
- **اختبار الإكمال البصري:** يقيس القدرة على إدراك العناصر الناقصة في الصور وتمييزها.
- **اختبار التذكر السمعي:** يقيس القدرة على تذكر سلاسل رقمية قد تصل إلى (8) أرقام، تعرض بمعدل رقمين في الثانية.
- **اختبار التذكر البصري:** يقيس القدرة على تذكر أشكال غير ذات معنى تعرض لمدة (5) ثوانٍ لكل شكل، وبمدى أقصى يصل إلى ثمانية أشكال.
- **اختبار الإكمال السمعي (اختبار احتياطي):** يقيس القدرة على إكمال مفردات لغوية ناقصة ذات مستويات صعوبة متدرجة.
- **اختبار التركيب الصوتي (اختبار احتياطي):** يقيس القدرة على تركيب الأصوات لتكوين كلمات، حيث تعرض الأصوات بفاصل زمني قدره نصف ثانية، بدءًا من كلمات ذات معنى وصولًا إلى أصوات غير ذات معنى.

5-1- نماذج من فقرات الاختبار:

1- ضع إشارة (X) أمام الشكل الصحيح:



2- ضع إشارة (X) تحت المثلث، و (XX) تحت المربع، و (X) تحت الدائرة، و (XX) تحت مستطيل:



المحـــاضرة العاشرة

مدخل عـــام إلى النظريـــة الحديثـــة في القيـــاس

الأهداف:

تهدف هذه المحاضرة إلى تزويد الطلبة بفهمٍ متكامل لمدخل النظرية الحديثة في القياس من خلال التعرف على التطور التاريخي لنظرية الاستجابة للمفردة (IRT) ومراحل نشأتها وتطورها مقارنة بالنظرية الكلاسيكية، بما يسهم في إدراك التحولات التي شهدها مجال القياس النفسي والتربوي. كما تسعى المحاضرة إلى تمكين الطلبة من استيعاب الافتراضات الأساسية لنماذج نظرية الاستجابة للمفردة أحادية البعد، من خلال توضيح مفهوم أحادية البعد (Unidimensionality) وأهميته في تفسير الأداء، وفهم مبدأ الاستقلال الموضعي (Local Independence) ودوره في ضمان دقة القياس، إضافة إلى التعرف على خصائص المنحنيات المميزة للمفردات (Item Characteristic Curve - ICC) وكيفية تفسيرها، وأخيرًا إدراك مفهوم الحد الأدنى من أثر التخمين (Minimal Guessing) وأثره في تحسين جودة القياس. وبذلك تهدف المحاضرة إلى بناء أساس نظري متين يمكن الطلبة من توظيف هذه المفاهيم في تحليل وتطوير الاختبارات النفسية والتربوية.

محتوى المحاضرة:

1- التطور التاريخي لنظريـــة الاستجابـة للمفردة (IRT)

2- افتراضات نماذج نظريـة الاستجابـة للمفردة الاختباريـة أحاديـة البعد

2-1- أحاديـــة البعد Unidimensionality

2-2- الاستقـلال الموضعـي Local Independence

2-3- توازي المنحنيات المميزة للمفردات Item Characteristic Curve (ICC)

2-4- الحد الأدنـى من أثر التخمين Minimal guessing

مدة المحاضرة: ساعة ونصف

تمهيد:

تقوم نظرية الاستجابة للمفردة على مجموعة من المفاهيم والأسس التي تختلف بصورة جوهرية عن تلك التي ترتكز عليها النظرية الكلاسيكية للاختبار في القياس النفسي والتربوي. حيث تعتمد هذه النظرية على نماذج رياضية احتمالية تهدف إلى تحقيق دقة أكبر في القياس، من خلال التحكم في العوامل التي قد تشوه المقارنات بين درجات الأفراد، وذلك عبر نمذجة بارامترات العملية الاختبارية. ويعود الفضل في عرض هذا الإطار النظري بصورة متكاملة إلى Frederic Lord و Melvin Novick في عملهما المرجعي "النظريـات الإحصائيـة لدرجات الاختبارات العقلية" (1968)، حيث قدما تصورًا رياضيًا يصف كيفية استجابة الفرد للمفردات الاختبارية، بما يسمح بتفسير أدائه أو التنبؤ به اعتمادًا على خصائص محددة تُعرف بالسمات.

وتستخدم هذه النماذج لتقدير نوعين من البارامترات: بارامترات خاصة بالمفردات، وأخرى خاصة بالأفراد فيما يتعلق بهذه السمات. ومن خلال هذه التقديرات يمكن تقديم تفسير أكثر دقة لأداء الأفراد في الاختبارات. ونظرًا لكون هذه السمات غير قابلة للملاحظة المباشرة، يتم الاستدلال عليها بطريقة غير مباشرة من خلال تقديرات كمية تعرف بـ السمات الكامنة أو القدرات (Abilities)، وهي تمثل الأساس الذي تبنى عليه عملية القياس في هذا الإطار النظري (علام صلاح، 2013، ص47).

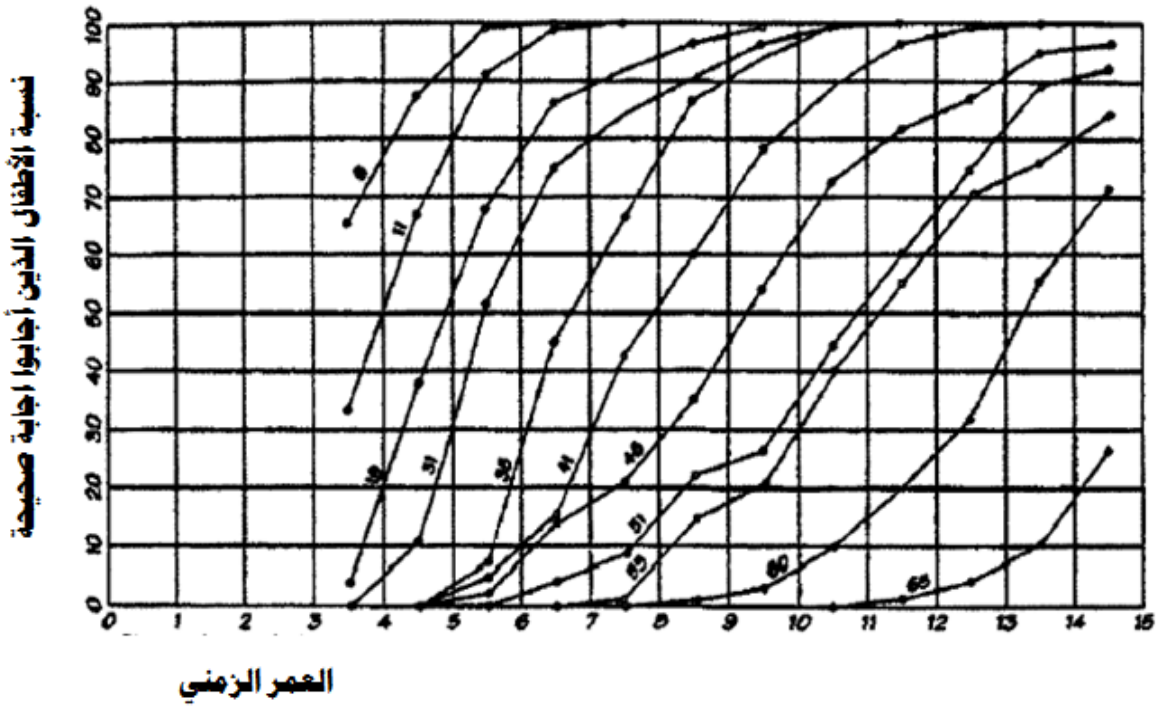
1- التطور التاريخي لنظريــة الاستجابـة للمفردة (IRT)**:**

ترجع البدايات الأولى لـ نظرية الاستجابة للمفردة إلى العمل الرائد الذي قدمه Louis Leon Thurstone عام 1925 في مقالته المعنونة "منهجية لمقايسة الاختبارات النفسية والتربوية (A Method of Scaling Psychological and Educational Tests)". وقد سعى ثرستون في هذا العمل إلى معالجة إحدى الإشكاليات الأساسية التي واجهت الباحثين آنذاك، والمتمثلة في تحديد أسلوب دقيق لمعايرة القدرة العقلية اعتمادًا على مفردات مقياس بينيه – سيمون للأطفال وفق تدرج العمر الزمني. وفي سبيل ذلك، قام ثرستون بتحليل العلاقة بين نسبة الأطفال الذين تمكنوا من الإجابة الصحيحة عن مهام المقياس (على المحور الرأسي) وأعمارهم الزمنية (على المحور الأفقي)، وذلك باستخدام عينة كبيرة من الأطفال. وقد أظهرت النتائج منحنيات ذات شكل انسيابي يشبه الحرف(S) ، تعكس تزايد احتمال الاستجابة الصحيحة مع التقدم في العمر الزمني.

وتعد هذه المنحنيات من أوائل المؤشرات على الفكرة الأساسية التي تقوم عليها النماذج الحديثة في القياس، حيث تتشابه مع المنحنيات اللوجستية المستخدمة لاحقًا في نموذج راش، والتي تعرف

عادةً باسم منحنيات خصائص الفقرة أو منحنيات القدرة، وتمثل العلاقة بين مستوى القدرة واحتمال الاستجابة الصحيحة.

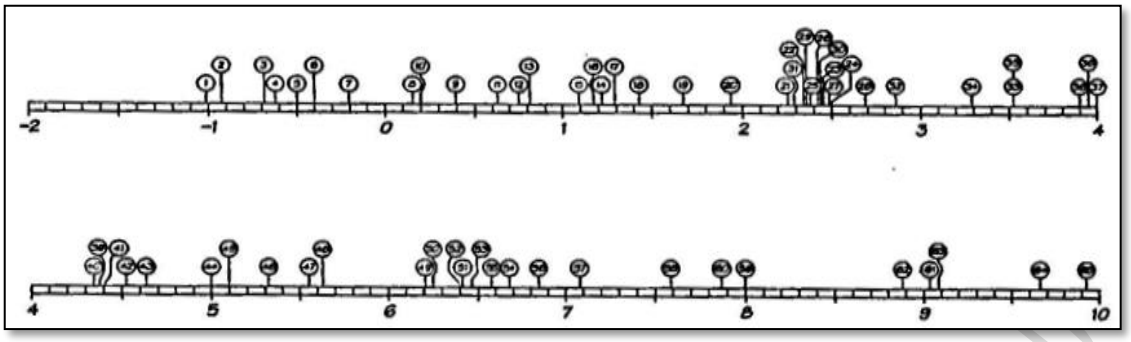
شكل رقم 11 رسم بياني لثورستون (Thurstone,1925**) يوضح نسبة الاجابة الصحيحة للأطفال على مفردات اختبار "بينيه- سيمون" للأطفال تبعاً لمتغير العمر الزمني** (Wilson, 2005, p89).



واصل Louis Leon Thurstone جهوده بعد ذلك من خلال تحديد مواقع (65) مفردة من مفردات اختبار الذكاء الخاص بـ مقياس بينيه – سيمون على متصل العمر الزمني، حيث اعتبر نقطة الأصل الاعتباطية مساوية للصفر، وحدد وحدة القياس بمقدار انحراف معياري واحد. ويبين هذا الإجراء عملية معايرة دقيقة مكنته من ترتيب المفردات على أساس صعوبتها.

ويظهر التدريج العمري الناتج عن هذه المعايرة أن موضع كل مفردة يعكس مستوى صعوبتها بالنسبة للعمر الزمني، بحيث ترتبط المفردات الأكثر صعوبة بمستويات عمرية أعلى. وبهذا الإسهام، وفر ثرستون أساسًا علميًا لمستخدمي المقياس يسمح بترتيب مفرداته وفق تدرج الصعوبة، كما أتاح إمكانية تقدير العمر العقلي للفرد اعتمادًا على عدد استجاباته الصحيحة، انسجامًا مع التصور الذي وضعه Alfred Binet (ديانا، 2010، ص36).

شكل رقم 12 المواضع المقدرة للمفردات على مقياس للعمر ذي نقطة أصل اعتباطية ووحدات قياس متساوية (Louis, 1925, p449**).**



يمثل هذا الشكل مقياسًا مطلقًا لمفردات اختبار مقياس بينيه – سيمون، حيث تعبر المسافات الفترية بين القيم عن وحدة الانحراف المعياري لمستوى ذكاء الأطفال في عمر ثلاث سنوات ونصف على هذا الاختبار، بينما تشير نقطة الأصل إلى متوسط مستوى الذكاء لهؤلاء الأطفال في العمر ذاته.

وفي ضوء ذلك، يمكن اعتبار Louis Leon Thurstone من أوائل الباحثين الذين قدموا إسهامات مبكرة تتقاطع مع الأسس المفاهيمية لـ نظرية الاستجابة للمفردة، وذلك قبل أن تتبلور هذه النظرية بمسماها وإطارها المعاصر.

2- افتراضات نماذج نظريـة الاستجابـة للمفردة الاختباريـة أحاديـة البعد:

تعتمد نماذج نظرية الاستجابة للمفردة أحادية البعد على مجموعة من الافتراضات الجوهرية التي يتعين تحققها في بيانات الاختبارات لضمان سلامة النتائج ودقتها. ويتم تقويم مدى توافق النموذج مع البيانات من خلال فحص هذه الافتراضات، حيث إن تحققها يسهم في تعزيز دقة القياس وموثوقيته. وفيما يلي عرض لأهم هذه الافتراضات الأساسية:

2-1- أحاديـــة البعد Unidimensionality:

تفترض نماذج نظرية الاستجابة للمفردة أن أداء الفرد في الاختبار يمكن تفسيره من خلال قدرة واحدة كامنة. غير أن تحقق هذا الافتراض قد يختلف باختلاف خصائص العينة؛ فقد يكون مناسبًا لمجموعة من المفحوصين، بينما لا ينطبق بالدرجة نفسها على مجموعة أخرى. فعلى سبيل المثال، في اختبارات المسائل اللفظية في الرياضيات، يتوقع أن يسهم الفهم القرائي بدور مهم في حل هذه المسائل. فإذا كان أفراد العينة يمتلكون مستوى مرتفعًا من الفهم القرائي، فإن أداءهم يعتمد أساسًا على القدرة الرياضية، وعندئذ يمكن اعتبار الاختبار أحادي البعد. أما إذا كان مستوى الفهم القرائي منخفضًا لدى المفحوصين، فإن الأداء سيتأثر بكل من القدرة الرياضية والقدرة القرائية معًا، مما يعني

أن تفسير الأداء لا يمكن إرجاعه إلى بعد واحد فقط، وبالتالي لا يتحقق شرط أحادية البعد(ميمي السيد، 2014، ص56).

-أمثلة أخرى:

- في اختبار الذكاء العام، من المتوقع أن تفسر الإجابات بواسطة البعد العام للذكاء فقط، ولكن إذا احتوى الاختبار على بنود قياس ذاكرة محددة، فقد تتداخل القدرة الذاكرية مع البعد العام، مما يخرق شرط الأحادية.

- في مقياس الاكتئاب، إذا صممت البنود لقياس الحزن فقط، فإن المقياس أحادي البعد، أما إذا ضمت البنود أعراضًا فسيولوجية أو سلوكية، فقد يظهر بعد إضافي يؤثر على الأداء، وبالتالي لا تتحقق الأحادية.

2-2- الاستقلال الموضعي Local Independence:

يفترض هذا الافتراض في نماذج نظرية الاستجابة للمفردة أن احتمال استجابة الفرد إجابة صحيحة على أي مفردة اختبارية يكون مستقلًا عن استجابته لبقية المفردات، بعد ضبط كل من مستوى القدرة التقديري للفرد وصعوبة المفردة. بمعنى آخر، تشير الفكرة إلى أن المفردات لا تتسم بأي ترابط إحصائي بمجرد التحكم في هذين العاملين. قد يبدو هذا الافتراض متناقضًا مع أحد مبادئ النظرية الكلاسيكية للاختبار الذي يؤكد على أهمية الاتساق الداخلي، حيث يفترض أن ارتفاع الترابط بين درجات المفردات يؤدي إلى زيادة معامل الثبات. لكن في النماذج الحديثة أحادية البعد، تصبح استجابات الأفراد للمفردات مشروطة بمستوى قدرتهم وصعوبة المفردات، بحيث تختفي أي ارتباطات بين المفردات بعد هذا الشرط، مما يعكس أن المفردات تقيس بعدًا واحدًا فقط.

ولتحقيق الاستقلال الموضعي، يجب أن يكون نمط استجابة الفرد (صحيح = 1، خطأ = 0) مساويًا لحاصل ضرب احتمالات حصوله على الدرجة المحددة في كل مفردة. ويتحقق ذلك عندما تقيس جميع المفردات سمة كامنة واحدة أو قدرة واحدة، أي أن جميع أزواج المفردات تحقق الاستقلال الموضعي ضمن أي مجتمع فرعي متجانس من الأفراد فيما يتعلق بالسمة الكامنة. بهذا، يرتبط الاستقلال الموضعي بعدد السمات أو القدرات التي يتضمنها الأداء على المفردات، ويعد مؤشرًا مهمًا على تحقق أحادية البعد عندما يكون النموذج يقدّر قدرة الفرد على بعد واحد.

وعلى الرغم من العلاقة الوثيقة بين الاستقلال الموضعي وأحادية البعد، إلا أنهما ليسا المفهوم نفسه؛ فقد يكون الاختبار متعدد الأبعاد إذا وجدت أكثر من سمة كامنة، حيث يمكن أن تتحقق استقلالية

المفردات بالنسبة للأفراد المتجانسين في كل سمة على حدة. بشكل عام، يمكن القول إن عدد أبعاد الاختبار يعكس عدد السمات الكامنة اللازمة لتحقيق الاستقلال الموضعي. ومن هذا المنطلق، يتضح أن تحقيق افتراض الاستقلال الموضعي قد يكون صعبًا في كثير من الاختبارات، ولا سيما الاختبارات العقلية (علام، 2013، ص63).

أمثلة نفسيـة وتربويـة:

- في اختبار الذكاء اللفظي، إذا كانت جميع البنود تقيس القدرة اللفظية فقط، فإن استجابة كل مفردة مستقلة عن الأخرى بعد ضبط مستوى القدرة، مما يحقق الاستقلال الموضعي. أما إذا تضمنت بعض البنود فهمًا رياضيًا، فقد تظهر ارتباطات بين المفردات، فينتفى الاستقلال.

- في اختبار التحصيل الدراسي في الرياضيات، يتوقع أن تكون استجابات الطلاب مستقلة إذا كانت كل مسألة تختبر قدرة رياضية واحدة، لكن إذا كانت المسائل اللفظية تعتمد على فهم القراءة، قد تتداخل استجابات البنود، فتفقد الاستقلالية.

- في مقاييس الاكتئاب، إذا كانت البنود تقيس الحزن فقط، فإنها تتحقق الاستقلالية، أما إذا ضمت البنود أعراضًا فسيولوجية أو سلوكية مرتبطة بالحزن، فقد تظهر بعض المفردات ترابطًا بعد ضبط مستوى الاكتئاب، مما يشير إلى انتفاء الاستقلال الموضعي.

2-3- توازي المنحنيات المميزة للمفردات Item Characteristic Curve (ICC):

تتفق كل من النظرية الكلاسيكية والنظرية الحديثة في القياس على فرضية أساسية مفادها وجود متصل للقدرة، بحيث يمكن عند معرفة موقع الفرد على هذا المتصل تقدير احتمال إجابته الصحيحة على أي مفردة اختبارية تقيس هذه القدرة. ورغم هذا الاتفاق، تختلف النظريتان في شكل الدالة التي تربط موقع الفرد باحتمال الإجابة الصحيحة:

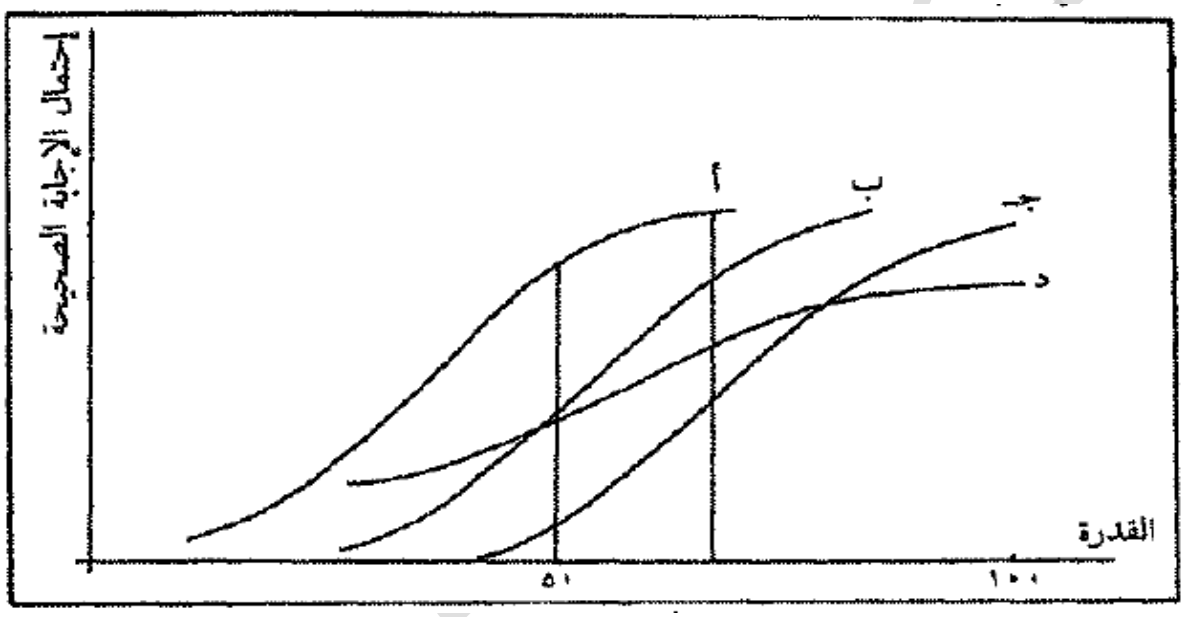
- في النظرية الكلاسيكية، يفترض أن تأخذ الدالــة شكل المنحنى الاعتدالـي في الاختبـارات المرجعية الجماعة، أو شكلًا ملتوياً ذو انحناء سلبي في الاختبـارات محكية المرجع .

- أما النظرية الحديثة، فتفترض أن تتخذ الدالـة شكل المنحنى اللوغاريتمي التراكمي .

وبناءً على ذلك، تفترض نماذج نظرية الاستجابـة للمفردة وجود دالـة مميزة لكل مفردة في الاختبار تعتمد أساسًا على موقع كل من الفرد والمفردة على متصل القدرة، وتتخذ شكل المنحنى اللوغاريتمي التراكمي (وليد، 2014، ص17).

وترى أمينة كاظم (1988) أن تساوي قوة التمييز بين مفردات الاختبار يرتبط بتحقيق شرط استقلالية القياس، حيث تتقارب الصعوبات النسبية للمفردات عبر مستويات القدرة المختلفة. وتعرَف قوة تمييز المفردة بأنها معدل التغير، أي الميل النسبي للمنحنى المميز للمفردة على محور القدرة. وتتأثر قدرة المفردة على التمييز بين الأفراد عند مستويات القدرة المختلفة بمدى انحناء منحناها المميز، فقد تزداد أو تنقص تبعًا لشكل هذا الانحناء. يوضح الشكل التالي أربع منحنيات مميزة لأربع مفردات (أ، ب، ج، د)، حيث يمكن ملاحظة الفروق في صعوبة وميزة كل مفردة.

شكل رقم 13 يوضح المنحنيات المميزة لأربع مفردات (أمينة كاظم، 1988، ص72**).**



أ- تتوازي تقريبا المنحنيات (أ)، (ب)، (جـ)، أي أن لها شكلا وانحناء عاما واحدا. ومعنى هذا أن الزيادة المتساوية في مستوى القدرة (مستوى الدرجة الكلية) يقترن بزيادة متساوية تقريباً في احتمال الاجابة الصحيحة على أي من هذه المفردات. وبعبارة أخرى أن قوة تمييز المفردات بين قدرات الأفراد متساوية وهذا ما يختلف بالنسبة للمنحنى (د)؛

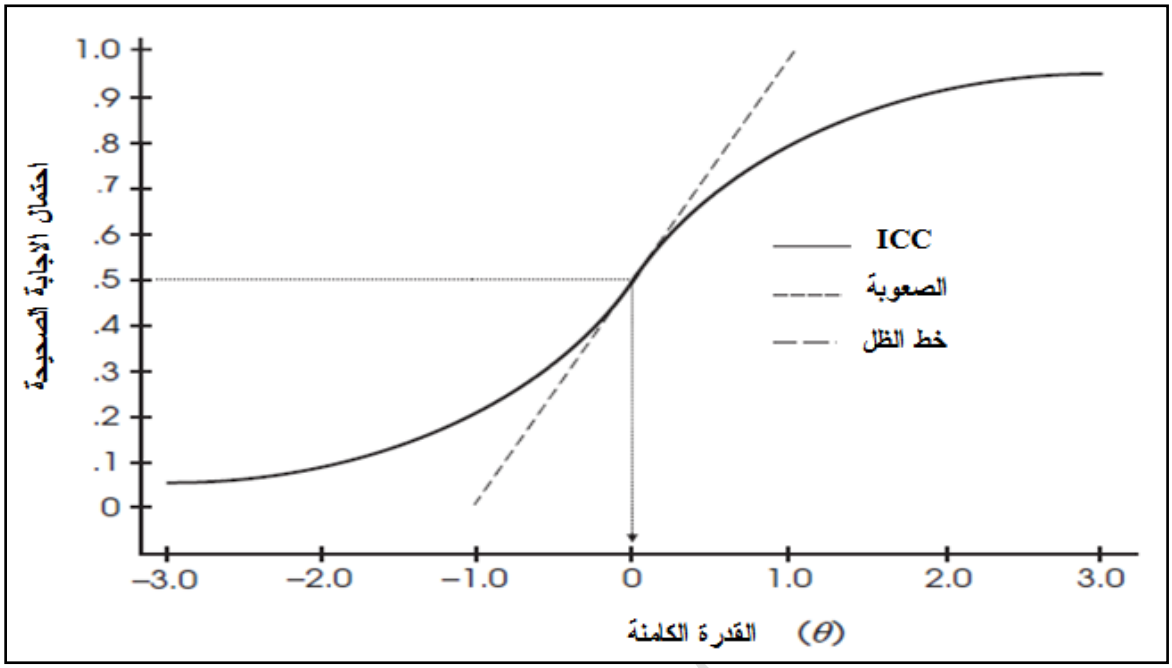
ب- إن فاعلية المفردة (أ) تمتد من المستوى المنخفض من القدرة وحتى المستوى فوق المتوسط مباشرة وليس لهذه المفردة فائدة تذكر للأفراد ذوي المستوى العالي من القدرة ولا تستطيع أن تميز بينهما على هذه القدرة حيث يمكن لهؤلاء الأفراد جميعا الاجابة الصواب على هذه المفردة. أما المفردة (جـ) فتمتد فاعليتها من المستوى العالي من القدرة وليس لهذه المفردة فائدة تذكر للأفراد منخفضي القدرة. حيث يخفق هؤلاء جميعا في الاجابة على هذه المفردة ومن ثم فهو لا يستطيع أن يميز بينهما. أما المفردة (ب) فتمتد فاعليتها عبر المستويات المتوسطة من القدرة؛

ج- في المدى المشترك لفاعلية هذه المفردات الثلاثة (أ، ب، جـ) يكون احتمال الاجابة الصواب على المفردة (أ) أكبر دائما من احتمال الاجابة صواب على المفردة (ب) وهذا يكون دائما أكبر من احتمال الاجابة الصواب على المفردة (جـ) وذلك عند أي مستوى من مستويات هذه القدرة المحدود بالمدى المشترك لفاعلية هذه المفردات. أي أن المفردة (أ) تكون دائما أسهل من المفردة (ب) وهذا يكون دائما أسهل من المفردة (جـ) وذلك عند المستويات المختلفة من القدرة في هذا المدى المشترك لفاعلية هذه المفردات أي أن صعوبة هذه المفردات تكون مستقرة عبر مستويات القدرة مما يعني ملاءمة هذه المفردات للنموذج؛

د- إن احتمال الاجابة الصواب على المفردة (د) يكون أكبر من احتمال الاجابة الصواب على المفردة (ب) عند المستوى المنخفض من القدرة في حين تكون أقل منها عند المستوى العالي من القدرة. معنى هذا أن المفردة (د) تكون أسهل من المفردة (ب) في المستويات المنخفضة من القدرة بينما يكون أصعب منه في المستويات العالية من القدرة. وهذا ما يمكن استنتاجه أيضا عند مقارنته بالمفردة (جـ) حيث تكون المفردة (د) أسهل من المفردة (جـ) في المستويات المتوسطة من القدرة في حين يكون أصعب منه عند المستويات العالية من القدرة. ومعنى هذا أن صعوبة هذه المفردة (د) غير مستقرة عبر مستويات القدرة المختلفة. وهو اخلال بفرض النموذج عن استقلال صعوبة المفردة عند قدرات العينة مما يعني عدم ملاءمة هذه المفردة للنموذج (أمينة كاظم، 1988، ص72).

وبالتالي فإن درجة تمييز المفردة تتناسب طرديًا مع ميل منحناه المميز عند نقطة الانقلاب. ويمكن تمثيل منحنى خصائص المفردة (ICC) والتنبؤ باحتمال الإجابة الصحيحة في حالة الاستجابة الثنائية وفق نموذج الاستجابة ثنائي المعلم، كما هو موضح في الشكل التالي:

شكل رقم 14 يوضح منحنى خصائص المفردة (ICC) والتنبأ باحتمال الاجابة الصحيحة في الاستجابة الثنائية، في نموذج الاستجابة للمفردة ثنائية المعلم. صعوبة المفردة عند القدرة صفر، وتمييز المفردة وهو ميل خط الظل عند القدرة صفر (Rex .B. Kline, 2011, p95**).**



يتضح من ذلك أن المنحنى المميز للمفردة الاختبارية يمثل العمود الفقري لنظرية القياس الحديثة وهو من الافتراضات الأساسية في نماذج الاستجابة للمفردة أحادية البعد. أما في حالة النماذج متعددة الأبعاد، فإن السطح المميز للمفردة يعد من الافتراضات الجوهرية لها (علام، 2013، ص 62).

2-4- الحد الأدنى من أثر التخمين Minimal guessing:

يفترض نموذج راش اللوغاريتمي انعدام التخمين أو أن يكون في حدوده الدنيا، ويقاس ذلك من خلال ارتفاع منحنى خصائص المفردة على متصل القدرة عند أدنى مستويات القدرة. نظريًا، تتراوح قيم معامل التخمين بين 0 و1، بينما عمليًا يعتبر مقبولًا أن تقع بين 0 و0.35، وأي قيمة تتجاوز هذا الحد تعد غير مقبولة .(Baker, 2001, p.28) يمثل هذا المعامل احتمال الإجابة الصحيحة بمحض الصدفة، ويستمد من الخط المقارب الأدنى (Lower Asymptote) في منحنى خصائص المفردة (أبو حطب وآخرون، 1993، ص.285).

2-5- السرعة Speediness:

ميز "لورد ونوفيك" بين نوعين رئيسيين من الاختبارات: اختبارات السرعة واختبارات القوة. ففي اختبارات السرعة (Speed Tests)، تعتمد الدرجة التي يحصل عليها الممتحَن على قدرته على

الإجابة عن أكبر عدد ممكن من المفردات ضمن زمن محدد وثابت، حيث تعكس الدرجة بشكل رئيسي معدل الاستجابة. أما في اختبارات القوة (Power Tests)، فتتركز على قدرة الممتحن على الإجابة عن مفردات الاختبار دون التقيد بالزمن، بحيث تكون الدرجات انعكاسًا لمستوى القدرة فقط. ويعرف فاروق عبد الفتاح موسى سرعة الأداء في الاختبار بأنها مدى اعتماد درجة الفرد على سرعة إنجازه، وغالبًا ما تقاس بنسبة الأفراد الذين لم يتمكنوا من الوصول إلى أو الإجابة عن آخر مفردة في الاختبار. تفترض نماذج نظرية الاستجابة للمفردة أن عامل السرعة لا يؤثر بشكل مباشر على استجابات المفردات؛ أي أن عدم تمكن بعض المفحوصين من الإجابة يعزى إلى انخفاض مستوى قدرتهم بدلاً من تأثير عامل الزمن.

مع ذلك، يمكن أن يكون لعامل السرعة تأثير غير مباشر، إذ قد يدفع بعض الأفراد إلى التخمين لتجاوز زمن الاختبار، مما يجعل سرعة الأداء في بعض الحالات أحد العوامل التي يقيسها الاختبار بشكل غير مقصود (ميمي، 2014، ص59).

المحـــاضرة الحادية عشر

أسس ومبادئ وتوجهات النظرية الحديثة في القياس

الأهداف:

تهدف هذه المحاضرة إلى إكساب الطلبة فهمًا عميقًا لأسس ومبادئ وتوجهات النظرية الحديثة في القياس من خلال التعرف على معنى الموضوعية في نموذج راش ودورها في تحقيق استقلالية القياس عن عينة الأفراد أو المفردات، إلى جانب استيعاب البعد السيكولوجي الكامن في هذا النموذج من حيث تفسير تفاعل الفرد مع مفردات الاختبار. كما تسعى المحاضرة إلى تمكين الطلبة من فهم طبيعة الاختبارات التحصيلية المبنية على نظرية الاستجابة للمفردة وكيفية بنائها وتطويرها، مع توضيح مفهومي الصدق والثبات في ضوء نموذج راش اللوغاريتمي الاحتمالي وأوجه تميزهما عن المفاهيم التقليدية. إضافة إلى ذلك، تهدف المحاضرة إلى تعريف الطلبة ببنوك الأسئلة (Item Banking) وأهميتها في تطوير الاختبارات، وكذلك الاختبارات التكيفية المحوسبة (Computerized Adaptive Testing - CAT) باعتبارها من أبرز التطبيقات الحديثة في القياس، بما يسهم في إعدادهم لتوظيف هذه الاتجاهات المعاصرة في الممارسات التقويمية بكفاءة وفاعلية.

محتوى المحاضرة:

1- معنى الموضوعيـة في نموذج راش

2- الجانب السيكولوجي في نموذج راش

3- الاختبارات التحصيلية المبنية على نظرية الاستجابة للمفردة

4- صدق الاختبار وثباته في ضوء نموذج راش اللوغاريتمي الاحتمالي

5- بنوك الأسئلة (Item Banking)

6- الاختبارات التكيفية المحوسبة (Computerized Adaptive Testing - CAT)

مدة المحاضرة: ساعة ونصف

تمهيد:

يشهد مجال القيـاس والتقويـم النفسـي والتربـوي تطورًا ملحوظًا في ضوء اعتماد نماذج حديثة تسعى إلى تجاوز محدودية الأساليب التقليدية وتحقيق قدر أكبر من الدقة والموضوعية. ويعد نموذج راش أحد أبرز هذه النماذج، حيث يقدم تصورًا متكاملًا يجمع بين البعد الرياضي والسيكولوجي في تفسير استجابات الأفراد. كما أسهمت النظرية الحديثة في القياس في إحداث نقلة نوعية في بناء الاختبارات التحصيلية، وإعادة صياغة مفاهيم الصدق والثبات، إلى جانب تطوير تطبيقات متقدمة مثل بنوك الأسئلة والاختبارات التكيفية المحوسبة، التي تعكس توجهات معاصرة نحو قياس أكثر كفاءة ومرونة.

1- معنى الموضوعيـة في نموذج راش:

في نموذج راش، تعرف الموضوعية بأنها القدرة على إجراء مقارنات بين ناتج تفاعل قدرات الأفراد مع صعوبة المفردات المناسبة، كما تمتد لتشمل المقارنة بين صعوبات مفردات أجاب عنها أفراد ملائمون. ومع ذلك، فإن هذه المقارنات لا تكون دقيقة تمامًا، بل تقع ضمن نطاق حول القيمة المقدرة، قد يزيد عنها أو ينقص قليلًا. وتتجلى هذه الموضوعية في جانبين رئيسيين:

1-1- استقلال تقدير بارامتر القدرة الفرد عن المفردة المستخدمة:

رغم أن مقارنة قدرات الأفراد تتطلب استخدام مفردات ملائمة، فإن نتيجة هذه المقارنة لا تتأثر بنوع المفردة المختارة من بينها. ويقصد بالمفردات الملائمة تلك التي تتسم بالصدق في تمثيل المتغير المقاس، والقدرة على قياسه بدقة، والتدرج على متصل واضح، فضلًا عن تماثل قدرتها على التمييز بين مستويات الأفراد. وبناءً على ذلك، فإن المقارنة بين قدرات الأفراد تكون مستقلة عن المفردة المستخدمة، بحيث تعطي استجاباتهم لأي مفردة ملائمة التقدير الكمي نفسه للفارق في القدرات، مع مراعاة الخطأ المعياري. وعليه، فإن ما يتم تقديره ليس القدرة المطلقة لفرد معين، بل المسافة بين قدرته وقدرة فرد آخر تتخذ كمرجع.

1-2- استقلال تقدير بارامتر صعوبة المفردة عن الفرد المجيب:

وبالمثل، فإن مقارنة صعوبات المفردات تعتمد على استجابات أفراد ملائمين، لكنها لا تتأثر بهوية الفرد الذي يجيب. ويعد الفرد ملائمًا إذا كانت إجابته متسقة مع ما هو متوقع من الأفراد الذين يمتلكون مستوى القدرة نفسه. وبالتالي، فإن المقارنة بين صعوبات المفردات تكون مستقلة عن الشخص المجيب، إذ تؤدي استجابة أي فرد ملائم إلى نفس التقدير الكمي للفارق في الصعوبة، مع الأخذ في

الاعتبار الخطأ المعياري. ومن ثم، فإن ما يقدر ليس الصعوبة المطلقة لمفردة معينة، بل المسافة بينها وبين مفردة أخرى تتخذ كنقطة مرجعية. وفي المجمل، كلما ازداد التوافق بين مستوى صعوبة المفردات وقدرات الأفراد، تعززت درجة الموضوعية في القياس.

مثال: لنفترض أن لدينا مفردتين:

- المفردة (A) سهلة.
- المفردة (B) صعبة.

ولدينا طالبان:

- الطالب (1) ذو قدرة عالية.
- الطالب (2) ذو قدرة منخفضة.

أ- استقلال قدرة الفرد عن المفردة:

إذا قارنا بين الطالبين باستخدام المفردة (A)، سنجد أن الطالب (1) يتفوق على الطالب (2). ولو استخدمنا المفردة (B)، سيظل الطالب (1) أفضل من الطالب (2).

- النتيجة: الفرق في القدرة بين الطالبين يبقى نفسه تقريبًا، بغض النظر عن المفردة المختارة (طالما كانت مناسبة).

ب- استقلال صعوبة المفردة عن الفرد:

إذا أجاب الطالب (1) عن المفردتين، سيجد أن (B) أصعب من (A). ولو أجاب الطالب (2)، سيصل إلى نفس النتيجة (B) أصعب من (A).

- النتيجة: الفرق في صعوبة المفردتين يبقى ثابتًا تقريبًا، بغض النظر عن الشخص الذي أجاب عليهم.

الخلاصة: نموذج راش يسمح بمقارنة عادلة وثابتة نسبيًا، حيث لا تعتمد النتائج على اختيار المفردات أو الأفراد، بل تعكس فروقًا حقيقية في القدرة والصعوبة.

2- الجانب السيكولوجي في نموذج راش:

يتجلى نموذج راش في صيغة رياضية تعبر عن واقع سيكولوجي، إذ يصف بدقة تفاعل الأفراد مع مفردات الاختبار. وتعتمد معادلته على تفسير احتمال الإجابة الصحيحة من خلال مبدأين أساسيين:

- كلما ارتفعت قدرة الفرد زادت احتمالية إجابته الصحيحة.
- تزداد احتمالية الإجابة الصحيحة عندما تكون المفردة سهلة بالنسبة للفرد.

وعند تحقق هذين المبدأين، يعد نموذج راش تمثيلًا واقعيًا لهذا التفاعل، بل والنموذج الوحيد القادر على تفسير أداء الأفراد في الاختبارات والمقاييس. لذلك، فإن توافق بيانات الاختبار مع متطلبات النموذج يعني صحة هذين الافتراضين، بينما يشير عدم التوافق إلى خلل في الأساس الذي يقوم عليه النموذج. فعلى سبيل المثال، إذا لم تتطابق مفردة ما مع النموذج، فإن افتراض تفوق الفرد الأعلى قدرة في احتمال الإجابة الصحيحة يصبح موضع شك، مما يثير تساؤلات حول ما تقيسه هذه المفردة. وبالمثل، إذا لم يتوافق أداء فرد معين مع النموذج، فإن افتراض زيادة احتمال إجابته الصحيحة على المفردات السهلة مقارنة بالصعبة لا يكون قائمًا، وهو ما يستدعي البحث عن عوامل مؤثرة في أدائه، كضعف قدرة المفردة على التمييز بين مستويات القدرة (علام، 2000، ص 211).

ومن هنا، فإن أي نموذج لا يسهم في التفسير السيكولوجي للسمات المستهدفة لا يصلح كأداة دقيقة في القياس النفسي والتربوي. ومع ذلك، يتميز نموذج راش بتطبيقات عملية مهمة، إذ يعتمد على عدد محدود من المعلمات، ويستند في الوقت ذاته إلى أساس سيكولوجي قوي، وهو ما يفسر الاهتمام الواسع به في مجال القياس النفسي والتربوي (علام، 2013، ص 58).

3- الاختبارات التحصيلية المبنية على نظرية الاستجابة للمفردة:

يعد هذا النوع من الاختبارات التحصيلية نموذجًا للقياس الموضوعي، حيث تتحدد درجات الطلبة فيه وفق خصائص مفردات الاختبار، وذلك لاعتماده في بنائه على نظرية السمات (Trait Theory). وتتيح هذه الاختبارات إمكانية قياس القدرات الحقيقية للطلبة بدرجة أعلى من الدقة.

وتفترض نظرية السمات، أو ما يعرف بنظرية الخصائص الكامنة، أنه يمكن تفسير الأداء التحصيلي للطالب أو التنبؤ به من خلال مجموعة من الخصائص المميزة التي يطلق عليها السمات. وتهدف هذه النظرية إلى تقدير مستويات الأفراد على هذه السمات بالاعتماد على أدائهم في مجموعة من مفردات الاختبار (أمين، 2009، ص 236).

4- صدق الاختبار وثباته في ضوء نموذج راش اللوغاريتمي الاحتمالي:

4-1- صدق الاختبار (Test Validity):

في إطار نموذج راش، يقصد بصدق الاختبار أن مفرداته تقيس بعدًا واحدًا، أي أنها تتدرج في مستوى الصعوبة بشكل يعكس قياس متغير كامن مشترك. كما يدل ذلك على إمكانية ترتيب قدرات الأفراد على هذا المتغير استنادًا إلى تقديرات أدائهم على المقياس ذاته، بحيث تنتظم كل من صعوبة المفردات وقدرات الأفراد على متصل واحد يمثل هذا المتغير (كاظم، 1988أ، ص 98).

2-4- ثبـات الاختبـار (Test Reliability):

في إطار نموذج راش، يتحقق ثبات الاختبار من خلال استيفاء شروط الموضوعية في القياس، إذ ينبغي أن تتسم الأداة المبنية وفق هذا النموذج بما يلي:

أ- استقلال القياس عن الأداة أو المقياس المستخدم.

ب- استقلال القياس عن عينة الأفراد التي طبق عليها المقياس.

ويعني ذلك أن خاصية الاستقلال التي يوفرها نموذج راش تضمن ثبات القياس، بحيث لا تتأثر تقديرات قدرة الفرد أو صعوبة المفردة بتغير عينة التدريج أو باختلاف الأداة المستخدمة لقياس القدرة ذاتها (الكحلوت، 1991، ص 131).

كما يتميز نموذج راش بعدم اقتصاره على تقديم معامل ثبات للمقياس ككل، بل يتيح أيضًا حساب معاملات ثبات على مستوى كل فرد وكل مفردة بصورة مستقلة (El-Korashy, 1995, p. 760).

5- بنوك الأسئلة (Item Banking):

حظي مفهوم بنوك الأسئلة (Item Banking) باهتمام متزايد من قبل خبراء القياس المعاصر ومراكز القياس والتقويم التربوي في العديد من دول العالم خلال العقدين الماضيين، وقد تعاظم هذا الاهتمام في الوقت الراهن بعد أن أظهرت نماذج نظرية الاستجابة للمفردة قدرتها على معالجة كثير من الصعوبات المنهجية والفنية والتقنية التي كانت تعوق بناء بنوك الأسئلة في إطار نماذج القياس الكلاسيكية (Masters & Evans, 1986; Rudner, 1998; Wainer, 2000).

وعلى الرغم من استخدام النماذج الكلاسيكية سابقًا في تصميم وبناء أنظمة بنوك الأسئلة، فإن الاستفادة منها ظلت محدودة، ويرجع ذلك إلى أن تحليل مفردات البنك وفق هذه النماذج يعتمد بدرجة كبيرة على خصائص العينة المستخدمة، مما ينعكس سلبًا على اتساق البنك وفاعليته ومرونته. ومن هنا، اتجه خبراء القياس إلى تبني منهجيات نظرية الاستجابة للمفردة (IRT) لمعالجة أوجه القصور في النماذج التقليدية، حيث تتميز معلمات هذه النظرية كصعوبة المفردة وقدرتها على التمييز بين مستويات القدرة أو السمة المقاسة بخاصية الثبات (Invariance) بالمعنى الإحصائي. وبفضل هذه الميزة، أصبحت بنوك الأسئلة بمفهومها الحديث مكوّنًا أساسيًا في الأنظمة التعليمية في العديد من الدول.

5-1- مفهوم بنـوك الأسئلـة ومستويـاته:

تشير أدبيات القياس المعاصر إلى تعدد التعريفات الخاصة بمفهوم بنوك الأسئلة، كما تتنوع التسميات المستخدمة للدلالة عليه، مثل: بنوك المفردات (Item Banks)، ومخزون الأسئلة (Question Reservoirs)، ومجموعات المفردات (Item Collections)، وقوائم أو ملفات الأسئلة (Item Pools)، ومكتبات المفردات الاختبارية (Test Items Library)، وبنوك المهام (Tasks Bank)، وغيرها من المسميات ذات الدلالة المتقاربة (علام، 2013، ص248).

ومن بين التعريفات التوضيحية لهذا المفهوم، تعريف شوبن (Choppin, 1978)، الذي يرى أن بنك الأسئلة هو مجموعة من المفردات الاختبارية يتم تنظيمها وفهرستها بطريقة مشابهة لتنظيم الكتب داخل المكتبات، بحيث يراعى في هذا التنظيم محتوى المفردة وخصائصها السيكومترية مثل الصعوبة والثبات والصدق وغيرها من الخصائص. ويمكن استخدام هذه المفردات لتكوين اختبارات تُخضع لاحقًا لعملية التعيير (Calibration) بهدف تحويلها إلى أدوات قياس معيارية. ويشبه شوبن بنك الأسئلة بمجموعة أدوات قياس الطول المختلفة، مثل المساطر والأشرطة المترية والميكرو مترات، حيث إن لكل أداة استخدامات مناسبة لمواقف قياس مختلفة؛ فقياس برج مرتفع يختلف عن قياس شجرة أو قطر أنبوب نحاسي. وعلى الرغم من هذا التنوع، فإن جميع أدوات قياس الطول تعاير على مقاييس مشتركة مثل البوصات أو السنتيمترات أو المليمترات، بما يسمح بالمقارنة بين نتائج القياسات المختلفة.

وبالطريقة نفسها، يوفر بنك الأسئلة مرونة في القياس التربوي، حيث يمكن بناء اختبارات من عينات فرعية مختلفة من المفردات المستخرجة من البنك، مع إمكانية استخدام معاملات التعيير (Item Calibration) في تحديد الخصائص السيكومترية للاختبار. كما يمكن تفسير درجات اختبارات مختلفة تحتوي على مفردات متباينة من البنك في إطار مشترك، شريطة أن يكون أساس تعيير المفردات موحدًا ومتسقًا لجميع مفردات البنك.

تعريف وود (Wood, 1978): يعرف بنك الأسئلة بأنه نظام قياس يتيح إجراء قياسات متسقة عبر مواقف وظروف متعددة. أما إذا كان الهدف هو إجراء قياس واحد في ظروف ثابتة، فلا تكون هناك حاجة إلى بنك أسئلة بما يتضمنه من تعقيد، إذ يكفي حينها استخدام اختبار تقليدي. كما يميز بين مفهوم بنك الأسئلة (Item Banks) ومفهوم قائمة أو ملف الأسئلة (Item Pool)، حيث إن مفردات بنك الأسئلة تكون مجمعة ضمن اختبارات معروفة الخصائص السيكومترية، مما يسهل تفسير الدرجات الناتجة

عن اختبارات تحتوي على مجموعات مختلفة من المفردات المستخرجة من البنك. في المقابل، يفتقر ملف أو قائمة الأسئلة إلى عملية التعيير التي تحدد الخصائص السيكومترية للمفردات، الأمر الذي يجعل المقارنة بين الدرجات على مقياس مشترك غير ممكنة.

تعريف رايت وبل (Wright & Bell, 1984): يرى "رايت وبل" أن بنك الأسئلة يتضمن مجموعة من المفردات التي خضعت لعملية تعيير، بحيث تشكل بنية متناسقة من البنود ترتبط بفكرة أو موضوع رئيسي مشترك (Common Theme)، وبذلك يقدّم البنك تمثيلاً إجرائيًا لمتغير معين يتم قياسه (علام، 2013، ص249).

تعريف ميلمان وأرتر (Millman & Arter, 1984): يعرف بنك الأسئلة بأنه مجموعة كبيرة نسبيًا من المفردات أو الأسئلة التي يسهل الوصول إليها، حيث يقصد بكبر الحجم أن عدد المفردات يفوق بكثير العدد المستخدم في أي اختبار واحد. كما يقصد بسهولة الوصول أن تكون المفردات مفهرسة ومنظمة وفق بنية واضحة، أو أن تتوفر عنها معلومات تساعد على اختيارها لبناء الاختبارات، سواء كانت هذه المعلومات متعلقة بمحتواها أو نوعها أو مصدرها أو مستواها المعرفي، أو بخصائصها السيكومترية كالصعوبة والتمييز.

تعريف درلندن وإيجن (Van der Linden & Eggen, 1986): يعد بنك الأسئلة المعاير (Calibrated Item Bank) مجموعة من المفردات التي تقيس السمة نفسها أو نطاقًا معرفيًا محددًا، ويتم تخزينها حاسوبيًا مع تقديرات تجريبية لمعلمات البنود. وتقدر هذه المعلمات بالاعتماد على نموذج مناسب من نماذج الاستجابة للمفردة، حيث يتم ملاءمة النموذج مع استجابات المفحوصين الفعلية.

تشير العديد من التعريفات إلى أن بنك الأسئلة لا يقتصر على مفردات الاختيار من متعدد أو الصواب والخطأ فقط، بل يمكن أن يتخذ طابعًا أكثر شمولًا واتساعًا، بحيث يشمل أيضًا الأسئلة المقالية المفتوحة، والمهام الأدائية، والتجارب العملية، والموضوعات البحثية، والأسئلة الشفوية، والأنشطة الإملائية، وغيرها من المواقف التقويمية التي تتطلب استجابات متعددة أو متدرجة (Graded Responses).

ومن خلال استقراء هذه التعريفات وغيرها، يتضح عدم وجود اتفاق كامل حول تحديد مفهوم موحد لبنوك الأسئلة، ويعود ذلك إلى تباين أهداف استخدامها، واختلاف المعايير المعتمدة في اختيار المفردات التي يتم إدراجها في البنك، وكذلك تباين محكات الحكم على جودتها. ومع ذلك، يمكن تصنيف هذه التعريفات ضمن ثلاثة مستويات رئيسة كما يلي:

المستوى الأول: يتضمن تخزين مفردات أو أسئلة يتم بناؤها بطريقة علمية مناسبة، مع التحقق من صدق محتواها. وعلى الرغم من أن هذا النوع من البنوك لا يعتمد على وجود مقياس أو تدريج (Scale)، إلا أنه يعد مفيدًا في إعداد اختبارات صفية، خاصة في الحالات التي لا تكون فيها الدرجة الكلية أو المقياس التفسيري عنصرًا أساسيًا. كما يسهم في تنمية مهارات المعلمين في بناء اختبارات مشابهة وذات جودة أفضل.

المستوى الثاني: يشمل تخزين مفردات أو أسئلة منظمة ومفهرسة وفق محتواها وخصائصها السيكومترية، والتي تحدد بالاعتماد على نماذج القياس الكلاسيكية، كمعامل ارتباط درجة المفردة بالدرجة الكلية (كمؤشر للتمييز)، وتوزيع استجابات المفحوصين على بدائل الاختيار من متعدد، وصعوبة المفردة، وصدق محتواها. وقد استخدم هذا النوع من البنوك سابقًا في بناء صيغ متكافئة من الاختبارات المقننة (Standardized Tests)، من خلال اختيار مجموعات مختلفة من المفردات بهدف مقارنة مستويات التحصيل بين المؤسسات التعليمية المختلفة.

المستوى الثالث: يتضمن تخزين المفردات أو الأسئلة بعد تعييرها وتدريجها باستخدام نماذج نظرية الاستجابة للمفردة (Calibrated Item Bank)، حيث يتم تجريب البنود وجمع بياناتها ومطابقتها مع النموذج المناسب لتقدير معلمات الصعوبة وقدرات الأفراد على مقياس مشترك يحدده النموذج المستخدم. ويسمح هذا النوع من البنوك بتصميم اختبارات قادرة على تقديم أقصى قدر من المعلومات حول قدرة المفحوص عند مستوى معين من الثقة.

ويعد هذا المستوى مرتبطًا ارتباطًا وثيقًا بنظرية القياس الحديثة (IRT)، ويمكن اعتباره نظام قياس متكامل (Measurement System)، حيث يمكن تقدير صلاحية مفردات جديدة لقياس السمة نفسها، وتعييرها على نفس أساس مفردات البنك. وبفضل هذا التعيير، يصبح من الممكن مقارنة نتائج اختبارات مختلفة تتضمن مجموعات متنوعة من المفردات المستخرجة من البنك (Visalberghi, 1990; Wright). كما يستخدم هذا النوع من بنوك الأسئلة في تصميم اختبارات قائمة على معارف ومهارات محددة، وفي تقدير قدرة المتعلمين في مجالات معرفية دقيقة من المنهاج. ويتطلب ذلك بالضرورة تخزين المفردات إلكترونيًا داخل قاعدة بيانات محوسبة (Computerized Data Base) لتسهيل عمليات التعيير والاسترجاع بكفاءة عالية (علام، 2013، ص250).

ويرى هامبلتون (Hambleton, 1986) أن أحد أسباب تعثر المحاولات المبكرة لإنشاء بنوك الأسئلة في أواخر الستينيات وبدايات السبعينيات في كل من الولايات المتحدة الأمريكية وإنجلترا يعود إلى الاعتماد الكبير على الإجراءات الورقية وما رافقها من أعباء تنظيمية وإدارية معقدة.

وبناءً على ما سبق، يتضح أن لبنوك الأسئلة مستويات متعددة تتدرج من الأنظمة اليدوية أو شبه المحوسبة التي تعتمد على بطاقات المفردات وتحليل البيانات إحصائيًا، وصولًا إلى البنوك المحوسبة المتكاملة التي تعتمد على أنظمة إدارة قواعد البيانات والتقنيات الحديثة في القياس والتقويم (علام، 2013، ص251).

6- الاختبارات التكيفية المحوسبة (Computerized Adaptive Testing - CAT):

أدت التطورات الحديثة في تقنيات الحاسوب وبرمجياته إلى إحداث نقلة نوعية في الأساليب والممارسات التعليمية، حيث تم توظيف هذه التقنيات في تطوير العملية التعليمية ومراجعتها بما يسهم في تحقيق أهدافها بكفاءة وفاعلية أكبر. وقد انعكس هذا التقدم بشكل إيجابي على تطور الاختبارات التربوية والنفسية وتطبيقاتها، إذ أسهمت الثورة في تكنولوجيا الحاسوب وصناعة البرمجيات في ظهور نظريات قياس متقدمة، من أبرزها نظرية الاستجابة للمفردة (Item Response Theory - IRT). كما نتج عن ذلك بروز أنماط حديثة من الاختبارات المعتمدة على الحاسوب، التي بدأت تحل محل اختبـارات الورقـة والقلـم (Paper-and-Pencil)، ومن أهمها الاختبـار التكيفـي المحوسب (Computerized Adaptive Testing - CAT)، الذي يتميز بتباين عدد مفرداته ومستوى صعوبتها وفقًا لمستوى قدرة المفحوص، بحيث يصمم الاختبار بصورة تفصيلية (Tailored) لكل فرد، فلا يواجه مفردات شديدة السهولة أو الصعوبة، بل تقدم له مفردات تتلاءم مع مستوى أدائه. وقد استدعى ذلك الاعتماد على نظرية قياس حديثة تمكن من تقدير قدرة الفرد بدقة واختيار المفردات المناسبة له.

وتعد نظرية الاستجابة للمفردة من أبرز هذه النظريات الحديثة، إذ تمثل أحد أهم التطورات الحاسمة في مجال القياس، حيث تقوم على تقدير احتمال الإجابة الصحيحة لكل مفردة على حدة من خلال نمذجة العلاقة بين المتغيرات الكامنة (قدرات الأفراد) واحتمال استجابتهم الصحيحة للمفردات الاختبارية. كما تعد إطارًا إحصائيًا عامًا يهتم بتحليل مفردات الاختبار، والأداء عليه، والعلاقة بينهما وبين القدرة المقاسة (Baker and Kim, 2004).

وفي إطار هذه النظرية، لا يعد خطأ القياس ثابتًا بين جميع الأفراد، بل يختلف باختلاف مستويات القدرة، كما أنها لا تتطلب استخدام اختبارات متوازية لتقدير الثبات. إضافة إلى ذلك، تتيح النظرية

قياس دقة كل مستوى من مستويات القدرة بشكل مستقل من خلال دالــة معلومــات المفردة (Information Function)، الأمر الذي يسمح بتصميم اختبارات توفر أعلى قدر من المعلومات عند مستويات معينة من القدرة بحسب الحاجة. وتركز نظرية الاستجابة للمفردة على جودة المفردة ودقتها أكثر من تركيزها على الشكل الكلي للاختبار، وقد أسهمت هذه الخصائص في تقديم حلول فعالة للعديد من مشكلات القياس، مثل تعميم الاختبارات، والكشف عن المفردات المتحيزة، ومعادلة الدرجات، وتطوير الاختبارات التكيفية المحوسبة (الصبحي والأحمدي، 2011، ص87).

يتميز نموذج نظرية الاستجابة للمفردة (IRT) بأن خطأ القياس فيه ليس ثابتًا بين جميع الأفراد، بل يختلف تبعًا لمستوى القدرة، كما أنه لا يتطلب استخدام اختبارات متوازية لتقدير الثبات. إضافة إلى ذلك، يتيح هذا النموذج تقدير دقة كل مستوى من مستويات القدرة بصورة مستقلة من خلال دالة معلومات المفردة (Information Function)، وهو ما يسمح بتصميم اختبارات توفر أعلى قدر من المعلومات عند مستويات معينة من القدرة وفقًا للحاجة. ومن هذا المنطلق، تركز نظرية الاستجابة للمفردة على جودة المفردة ودقتها أكثر من تركيزها على الشكل العام للاختبار، الأمر الذي أسهم في تقديم حلول فعالة للعديد من مشكلات القياس، مثل تعميم الاختبارات، والكشف عن المفردات المتحيزة، ومعادلة الدرجات، وتطوير الاختبارات التكيفية المحوسبة.

وقد شهدت الاختبارات التكيفية المحوسبة تطورًا ملحوظًا على المستويين النظري والتطبيقي، مع دخولها حيز التنفيذ الفعلي في الدول المتقدمة منذ أوائل تسعينيات القرن الماضي. وقد وصف علام (2005، ص 324) هذا التحول بأنه تحول جوهري، مشيرًا إلى أن عددًا من المؤسسات والهيئات العالمية تبنت أنظمة الاختبار التكيفي المحوسب (CAT)، وقامت بتحويل اختباراتها التقليدية إلى اختبارات تكيفية. ومن أبرز هذه المؤسسات وزارة الدفاع الأمريكية التي اعتمدت النسخة التكيفية المحوسبة من بطارية الاستعداد المهني للقوات المسلحة الأمريكية (Armed Services Vocationnel Aptitude Battery - ASVAB). كما قامت مؤسسة خدمات الاختبارات التربوية Educational) (Testing Service - ETSبتحويل عدد من الاختبارات المعروفة إلى صيغ تكيفية محوسبة، مثل اختبار اللغة الإنجليزية كلغة أجنبية (TOEFL)، واختبار القبول للدراسات العليا(GRE) ، واختبار الاستعداد الدراسي (SAT)، بالإضافة إلى اختبار ترخيص التمريض (NCLEX) (الصبحي والأحمدي، 2011، ص87).

المحـــاضرة الثانية عشر

نمـــاذج من النظريـــة الحديثـــة في القيـــاس

الأهداف:

تهدف هذه المحاضرة إلى تمكين الطلبة من التعرف على أنواع نماذج الاستجابة للمفردة أحادية البعد، مع التركيز على النماذج الاستاتيكية (Static Models) بوصفها من الأسس الرئيسة في النظرية الحديثة في القياس. كما تسعى إلى توضيح طبيعة النماذج المتعلقة بالمفردات ثنائية الدرجة، من خلال استيعاب خصائص كل من النموذج أحادي المعلم (1PL) وفكرته القائمة على معلمة الصعوبة، والنموذج ثنائي المعلم (2PL) الذي يضيف معلمة التمييز، إضافة إلى النموذج ثلاثي المعلم (3PL) الذي يتضمن معلمة التخمين. وتهدف كذلك إلى تنمية قدرة الطلبة على المقارنة بين هذه النماذج وفهم افتراضاتها واستخداماتها في تحليل الاختبارات النفسية والتربوية.

محتوى المحاضرة:

أولاً- أنواع نمـاذج الاستجابـة للمفردة أحاديـة البعد

1- النماذج الاستاتيكية (Static Models)

1-1- نمـاذج تتعلق بالمفردة ثنائية الدرجـة

1-1-1- النموذج آحادي المعلم (One-Parameter Logistic Model (1PL))

1-1-2- النموذج ثنائـي المعلـم (Two-Parameter Logistic Model (2PL))

1-1-3- النموذج ثلاثـي المعلم (Three-Parameter Logistic Model (3PL))

مدة المحاضرة: ساعة ونصف

تمهيد:

تندرج نماذج الاستجابة للمفردة أحادية البعد ضمن الإطار المتقدم للنظرية الحديثة في القياس، حيث تسهم في تقديم تفسير دقيق لطبيعة العلاقة بين قدرة الفرد وخصائص مفردات الاختبار. ومن بين هذه النماذج، تبرز النماذج الاستاتيكية التي تهتم بتحليل استجابات الأفراد في ضوء خصائص المفردة، خاصة في الاختبارات ذات الدرجتين (صحيح/خطأ). وقد تطورت هذه النماذج لتشمل عدة صيغ، تبدأ بالنموذج أحادي المعلم الذي يركز على صعوبة المفردة، ثم النموذج ثنائي المعلم الذي يضيف خاصية التمييز، وصولًا إلى النموذج ثلاثي المعلم الذي يأخذ في الاعتبار أثر التخمين. ويعكس هذا التنوع في النماذج تطورًا في فهم القياس ودقته، بما يخدم بناء اختبارات أكثر كفاءة وموضوعية.

أولاً- أنواع نمـاذج الاستجابـة للمفردة أحاديـة البعد:

تتنوع النماذج ضمن إطار نظرية الاستجابة للمفردة تبعًا لاختلاف الافتراضات التي تقوم عليها البيانات الاختبارية. فقد يفترض أن الأداء في الاختبار يعكس سمة واحدة فقط (أحادية البعد)، أو عدة سمات في الوقت نفسه (متعددة الأبعاد). كما قد تفترض بعض النماذج اختلاف المفردات في قدرتها على التمييز بين مستويات القدرة، في حين تفترض نماذج أخرى تأثر الإجابات الصحيحة في مفردات الاختيار من متعدد بعامل التخمين العشوائي أو عدم تأثرها به، كما هو الحال في الأسئلة المقالية المفتوحة. كذلك تختلف هذه النماذج باختلاف طبيعة الدرجات، سواء كانت ثنائية التصحيح أو متعددة الفئات، إضافة إلى افتراضات أخرى متعددة.

وبناءً على ذلك، يمكن تصنيف نماذج نظرية الاستجابة للمفردة إلى قسمين رئيسيين:

1- نماذج أحادية البعد(Unidimensional Models).

2- نماذج متعددة الأبعاد (Multidimensional Models).

ويندرج تحت كل من هذين القسمين نوعان فرعيان هما: النماذج الاستاتيكية (Static Models) والنماذج الديناميكية (Dynamic Models) (علام، 2013، ص67).

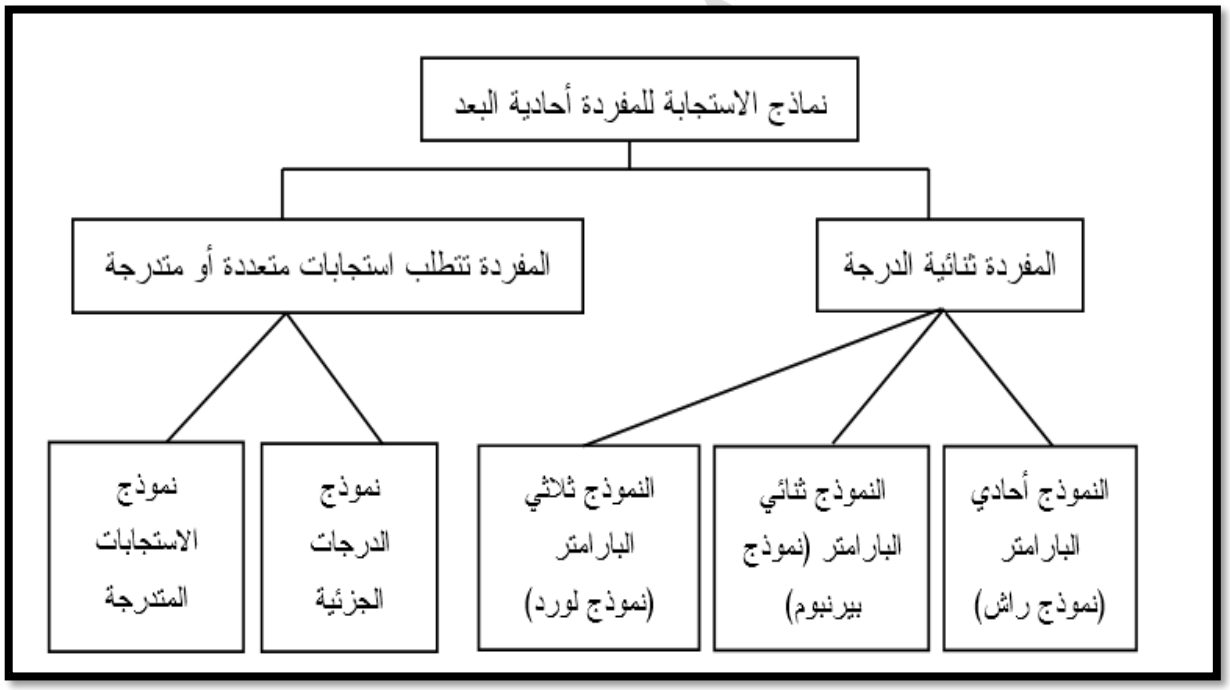
1- النماذج الاستاتيكية (Static Models):

يعنى هذا النوع من النماذج بقياس الأداء في لحظة زمنية محددة، مع التركيز على الكشف عن العمليات الكامنة التي تفسر أداء الأفراد في الاختبارات النفسية والتربوية. وتختلف هذه النماذج من حيث الصيغ الرياضية التي تصف العلاقة بين استجابة الفرد في الاختبار ومستوى قدرته، وكذلك

من حيث عدد المعلمات المستخدمة في توصيف كل مفردة من مفردات الاختبار (ميمي، 2014، ص 61). وفي الوقت الحالي، توجد ثلاثة نماذج رئيسة شائعة الاستخدام، يتميز كل منها بصيغة رياضية خاصة تنعكس في منحنيات خصائص المفردات. وتطبق هذه النماذج عادة على المفردات ثنائية التدريج التي لا تتطلب عامل السرعة في الإجابة، والتي تقيس بعدًا واحدًا. كما توجد نماذج أخرى مخصصة للتعامل مع الاستجابات الاسمية (Nominal Responses) والاستجابات متعددة الفئات (Polytomous Responses)، ومن أبرزها نموذج الدرجات الجزئية (Partial Credit Model) ونموذج الاستجابات المتدرجة (Graded Response Model)، إلى جانب نماذج أخرى مشابهة.

وتستخدم هذه النماذج في تحليل البيانات المستمدة من مقاييس الاتجاهات والشخصية، ومقاييس التقدير، إضافة إلى المفردات الاختبارية التي تتطلب خطوات متعددة، حيث يتم منح درجات جزئية لكل خطوة يتم إنجازها في حل المشكلة، كما هو الحال في المسائل الرياضية. ويوضح الشكل التالي هذه النماذج بصورة تخطيطية (علام، 2013، ص67).

شكل رقم 15 يوضح أنواع نماذج الاستجابة للمفردة أحادية البعد.



1-1- نمـاذج تتعلق بالمفردة ثنائية الدرجـة:

1-1-1- النموذج آحادي المعلم ((One-Parameter Logistic Model (1PL)):

يعد النموذج أحادي المعلم من أبسط نماذج نظرية الاستجابة للمفردة أحادية البعد، ويعرف شائعًا باسم نموذج راش (Rasch Model)، نسبةً إلى عالم الرياضيات الدنماركي جورج راش (,Rasch

1960). وعلى الرغم من ارتباط اسمه بهذا النموذج، فإن راش (1901–1980) كان يعارض تسمية النموذج باسمه أثناء حياته، مؤكدًا أن ما قدّمه لا يقتصر على نموذج واحد، بل يتضمن مجموعة من النماذج الإحصائية التي أطلق عليها " نماذج القياس (Measurement Models) "، وليس "نموذج راش" بالمفهوم المتداول.

وقد تم تطوير هذا النموذج بشكل مستقل عن بقية نماذج الاستجابة للمفردة، حيث يفترض أن المنحنى المميز للمفردة يعبر عنه من خلال دالة لوجستية أحادية المعلم. وفي هذا الإطار، يمثل المتغير التابع احتمال قيام الفرد (**j**) بالإجابة الصحيحة عن المفردة (**i**)، في حين تتمثل المتغيرات الكامنة المستقلة في كل من قدرة الفرد (**θ**) وصعوبة المفردة ($\mathbf{b_i}$). ويتم دمج هذين المتغيرين في صيغة رياضية تعتمد على الفرق بين قدرة الفرد وصعوبة المفردة.

ويرتبط هذا الفرق باحتمال الاستجابة الصحيحة من خلال دالـــة لوجستية غير خطية، تستخدم لنمذجة العلاقة بين المتغيرات الكامنة والاستجابة الملاحظة (علام، 2013، ص 69).

كان "راش" قد نشر نموذجه عام 1960 متناولًا تحليل بيانات الاختبارات من منظور نظرية الاحتمالات وعلى الرغم من انطلاقه من إطار مرجعي مختلف إلا أن المنحنى المميز للمفردة الذي توصل إليه كان لوجستي الشكل. ويعرف معلم الصعوبة (b_j) بأنه النقطة على مقياس القدرة التي تكون عندها احتمالية الإجابة الصحيحة على المفردة مساوية لـ .0.5 ورغم أن المدى النظري لهذا المعلم يمتد من ∞+ إلى∞− فإن القيم العملية المستخدمة في الأمثلة عادة ما تقع بين 3+ و−3. وبما أن المنحنى المميز للمفردة يأخذ شكل الحرف اللاتيني S فإن ميله (Slope) يتغير تبعًا لمستوى القدرة ويبلغ قيمته العظمى عندما تتساوى القدرة مع صعوبة المفردة. وفي إطار هذا النموذج فإن معلم التمييز ثابت ويساوي a = 1 لجميع المفردات بينما يُسمح فقط لمعلم الصعوبة بتغيير قيمه. ولهذا يعرف نموذج راش بأنه النموذج اللوجستي أحادي المعلمة والذي يمثل بالمعادلة الرياضية التالية (بيكر، 2001، ص22):

$$P\left(x_{ji}=1 \mid \theta_i, b_j\right)=\frac{e^{(\theta_i-b_j)}}{1+e^{(\theta_i-b_j)}}$$

حيث أن:

x_{ji}: الاستجابة على المفردة j من خلال الشخص i.

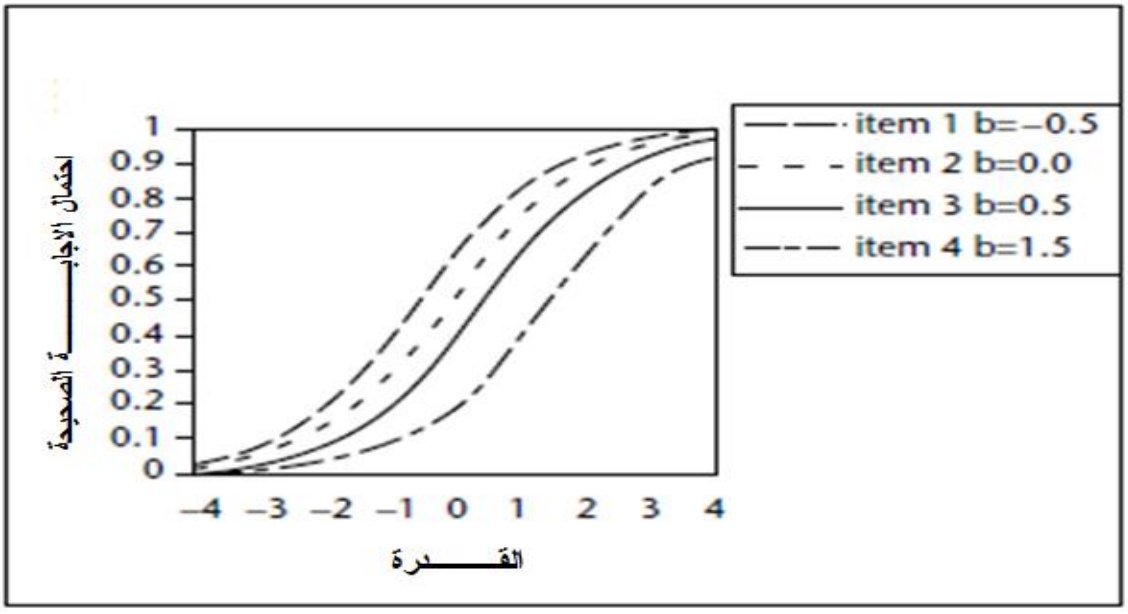
bj: وضع (صعوبة) المفردة j.

θ_i: مستوى السمة الكامنة (القدرة) لدى الشخص i.

e: هي ثابت قيمته 2.718.

وهنا لابد وأن نلاحظ أن معلم التمييز المستخدم في معادلة النموذج الثنائي المعلم ولكنه نظرًا لأنه دائمًا ما يساوي القيمة (1) فإنه عادة لا يظهر في المعادلة. وتعرف هذه الصيغة باسم "نموذج الترجيح اللوغاريتمي أحادي البارامتر(1PL) " وذلك لأنها تعتمد على الدالـة الأسيـة في التنبؤ بالاحتمالات وتشتمل على بارامتر (معلم) واحد فقط هو معلم صعوبة المفردة (Item Difficulty Parameter). والذي يستخدم لتمثيل الفروق بين المفردات (Mislevy, 1990, p.195). ويمكن توضيح هذه الصيغة الرياضية من خلال المنحنى المميز لأربع مفردات تقيس سمة كامنة واحدة.

شكل رقم 16 التمثيل البياني لصيغة نموذج راش لاربعة مفردات مختلفة الصعوبـة.



يتضح من الشكل السابق أن المنحنيات الاربعة متوازية تختلف فقط في ازاحة الموقع أي في (الصعوبة) حيث أن المفردة (4) تبدو أكثر صعوبة من بقية المفردات.

مثال حسابـي: لنفترض أن لدينا مستوى قدرة شخص ما θ_i تساوي (3-)، وقيمة معلم الصعوبة bj للمفردة تساوي (1) (بيكر، 2001، ص23).

- بداية نقوم بحساب الحد الأول في البسط من المعادلة السابقة $e^{(\theta-b)}$

$$\mathbf{EXP} = 1\ (-3-1) = 0.0183$$

- نقوم بتعوض الناتج في البسط وفي المقام مع إضافة الواحد، كالآتي:

$$P(x_{ji}) = 0.0183 / 1 + 0.0183 = 0.0179 \approx 0.02$$

إذن عند مستوى قدرة شخص ما (-3) فإن احتمال الاستجابة الصحيحة على هذه المفردة هو القيمة 0.02.

التفسير: يعني أن الشخص ضعيف جدًا في السمة المقاسة لذلك تكاد فرصته في الإجابة الصحيحة على هذا السؤال تساوي الصفر (2%) فقط.

1-1-2- النموذج ثنائـي المعلم (Two-Parameter Logistic Model (2PL)):

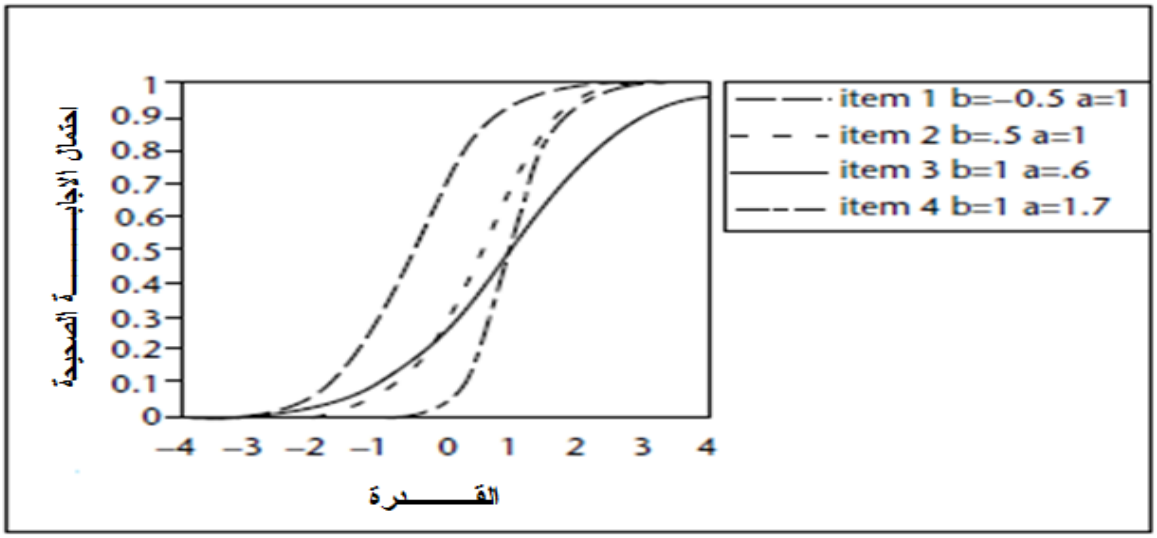
اقترح عالم الإحصاء بيرنبوم (Birnbaum, 1968) بالتعاون مع مجموعة من زملائه في جامعة كولومبيا الأمريكية نموذج الترجيح اللوغاريتمي ثنائي البارامتر (Two-Parameter Logistic Model – 2PL). ويختلف هذا النموذج عن نموذج راش في كونه يسمح باختلاف المفردات في كل من بارامتر الصعوبة وبارامتر التمييز في حين يقتصر نموذج راش على الصعوبة فقط. إلا أن هذا النموذج يفتقر إلى بعض الخصائص الإحصائية التي يتميز بها نموذج راش مما يجعل عملياته الحسابية أكثر تعقيدًا. وقد جرى عرض هذا النموذج ونشره في كتاب لورد ونوفيك (Lord & Novick, 1968) بعنوان النظريات الإحصائية لدرجات الاختبارات العقلية. وبناءً على ذلك أصبحت الصيغة الرياضية لهذا النموذج تتضمن بارامتر التمييز إلى جانب بارامتر الصعوبة بحيث يشتمل النموذج على معلمين أساسيين يمثلان خصائص المفردة كما يلي:

$$P\left(x_j = 1 \mid \theta, b_j\right) = \frac{e^{a_j\left(\theta, -b_j\right)}}{1 + e^{a_j\left(\theta, -b_j\right)}}$$

يتضح من هذه الصيغة أن تمييز المفردة (a_j) يضرب في الفرق بين مستوى القدرة وصعوبة المفردة كما هو موضح في المعادلة السابقة. ويكون لهذا الفرق تأثير أكبر في احتمالات الإجابة على المفردات ذات التمييز العالي. إن إدراج بارامتر التمييز في هذا النموذج يجعل تقدير القدرة معتمدًا على النمط العام لإجابات الفرد الصحيحة والخاطئة وليس فقط على الدرجة الكلية. وبذلك يمكن أن نحصل على تقديرات مختلفة لمستوى السمة لدى فردين حصلا على الدرجة الكلية نفسها إذا كان

نمط الإجابات مختلفًا. فالحصول على إجابات صحيحة في مفردات ذات تمييز مرتفع مع الإخفاق في مفردات ذات تمييز منخفض يؤدي إلى تقدير أعلى لمستوى السمة. ولهذا السبب لا يمنح هذا النموذج أوزانًا متساوية للمفردات عند تقدير مستويات السمة بخلاف نموذج راش (بيكر، 2001، ص24). وينتج عن ذلك إمكانية تقاطع المنحنيات المميزة للمفردات في الاختبار المعد وفق هذا النموذج، كما يوضحه الشكل التالي:

شكل رقم 17 التمثيل البياني لاربعة مفردات في النموذج ثنائي البارامتر.



يظهر الشكل السابق أن المنحنيات المميزة لأربع مفردات تختلف في مواضع تقاطعها مع المحور الأفقي مما يشير إلى اختلاف مستويات الصعوبة بين هذه المفردات. كما أن هذه المنحنيات تختلف في مقدار ميلها (Slope) حيث يتضح أن ميل المنحنى المميز للمفردة (3) أكبر من ميل منحنيات المفردات الأخرى (1، 2، 4) وهو ما يدل على أن المفردة (3) تمتلك قدرة أعلى على التمييز بين الأفراد مقارنة ببقية المفردات.

مثال حسابي: لنفترض أن لدينا مستوى قدرة شخص ما θi تساوي (3-)، وقيمة معلم الصعوبة للمفردة bj تساوي (1)، و معلم التمييز a_j يساوي (0.5) (بيكر، 2001، ص20).

- بداية نقوم بحساب الحد الأول في البسط من المعادلة السابقة $e^{a(\theta-b)}$

$$\mathbf{EXP} = 0.5\ (-3-1)$$

$$\mathbf{EXP} = -2 = 0.135$$

- نعوض الناتج في البسط وفي المقام مع إضافة الواحد، كالآتي:

$$P(x_{ji}) = 0.135 / 1 + 0.135 = 0.118 \approx 0.12$$

إذن عند مستوى قدرة شخص ما (-3) فإن احتمال الاستجابة الصحيحة على هذه المفردة هو القيمة 0.12.

التفسير: يعني هذا أن الشخص قدرته منخفضة جدًا (-3 انحراف معياري) ما يزال يملك احتمالًا مقداره 12% للإجابة الصحيحة على المفردة، بسبب سهولة نسبية للمفردة أو هناك أثر للتخمين.

3-1-1- النموذج ثلاثـي المعلم ((Three-Parameter Logistic Model (3PL)):

لا يأخذ كل من نموذج راش والنموذج ثنائي البارامتر في الاعتبار أن احتمال الإجابة الصحيحة على مفردة اختيار من متعدد قد يكون أعلى لدى الأفراد ذوي القدرة المنخفضة نتيجة لجوئهم أحيانًا إلى التخمين العشوائي. ففي نموذج راش تميل قيمة $P_i(\theta)$ إلى الاقتراب من الصفر كلما انخفضت قيمة القدرة θ. ومع ذلك يمكن توقع أن نسبة من ذوي القدرة المنخفضة ستتمكن من الإجابة الصحيحة على بعض المفردات ولو بالصدفة بسبب التخمين. انطلاقًا من هذه الملاحظة أُضيف في النموذج ثلاثي البارامتر (Three-Parameter Logistic Model) بارامتر ثالث يمثل احتمال توصل هؤلاء الأفراد إلى الإجابة الصحيحة من خلال التخمين بحيث يُؤخذ هذا الاحتمال في الاعتبار عند مطابقة بيانات الاختبار مع النموذج.

وقد صاغ لورد (Lord) الصيغة الرياضية لهذا النموذج انطلاقًا من حقيقة ثابتة في الاختبارات وهي أن بعض المفحوصين سيجيبون إجابة صحيحة على بعض المفردات نتيجة التخمين. وبالتالي فإن احتمالية الإجابة الصحيحة تشتمل على مكون صغير يُعزى إلى التخمين. ولم يضمن أي من النموذجين السابقين للمنحنى المميز للمفردة أثر التخمين في حساباته. لذلك قام آلان بيرنبوم (Allan Birnbaum, 1968) بتعديل النموذج اللوجستي ثنائي البارامتر ليضيف إليه معلمًا ثالثًا يوضح إسهام التخمين في احتمالية الإجابة الصحيحة على المفردة (بيكر، 2001، ص25).

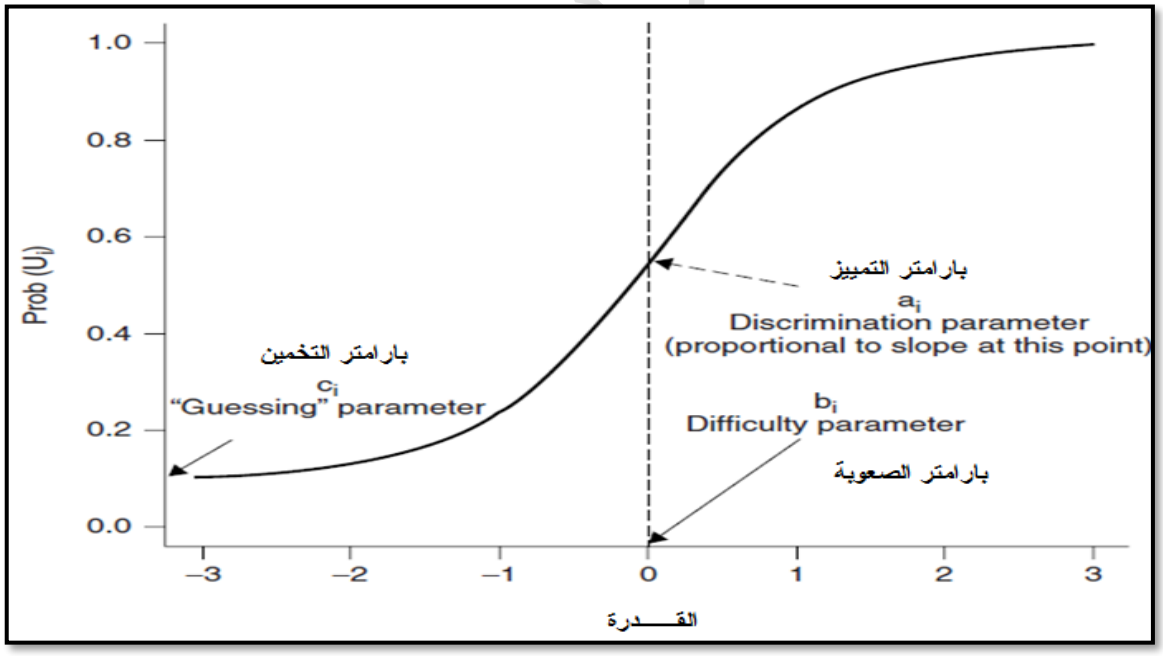
ورغم أن هذا التعديل أدى إلى فقدان بعض خصائص الدالـة اللوجستيـة إلا أن النموذج لا يزال يُعرف باسم النموذج اللوجستي ثلاثي البارامتر حتى وإن كان من الناحية الفنية لا يعد لوجستيًا بالكامل، والذي يعطى بالمعادلة التالية:

$$P_i(\theta) = c_i + (1 - c_i)\frac{e^{a_i(\theta_i - b_i)}}{1 + e^{a_i(\theta_i - b_i)}}$$

يطلق على البارامتر C_i اسم الخط التقاربي الأدنى (Lower Asymptote) للمنحنى المميز للمفردة وهو يمثل احتمال الإجابة الصحيحة لدى الأفراد ذوي القدرة المنخفظة. ويأخذ هذا المعلم مدى نظريًا بين $0 \leq C_j \leq 1$ إلا أن القيم التي تتجاوز 0.35 تعد غير مقبولة من الناحية العملية.

يهدف إدراج هذا البارامتر في النموذج إلى معالجة مشكلة عدم تطابق المنحنيات المميزة للمفردة عند الطرف الأدنى لمتصل السمة حيث يعد التخمين العشوائي أحد العوامل المؤثرة في أداء الأفراد على الاختبار(Charles et al., 2012, p246). ويعتبر تغيّر تعريف معلم الصعوبة من أبرز الآثار الجانبية لإضافة معلم التخمين C_j. ففي النموذجين السابقين كان معلم الصعوبة b_j يحدد النقطة على مقياس القدرة التي يكون عندها احتمال الإجابة الصحيحة مساويًا لـ 0.5. أما مع إدخال معلم التخمين فإن الحد الأدنى للمنحنى المميز للمفردة يصبح مساويًا لـ C_j بدلًا من الصفر مما يجعل C_j يمثل أدنى احتمال للإجابة الصحيحة على المفردة. وبناءً على ذلك يعرف معلم الصعوبة حاليًا على أنه النقطة على مقياس القدرة التي تقع عندها احتمالية الإجابة الصحيحة في المنتصف بين هذا الحد الأدنى (C_j) والواحد الصحيح. ويمكن توضيح هذه الصيغة الرياضية من خلال الشكل التالي:

الشكل رقم 18 التمثيل البياني للنموذج ثلاثي البارامتر.



يتضح من الشكل السابق منحنى لمفردة اختبارية يوضح ثلاثة بارامترات أساسية هي: الصعوبة، والتمييز، والتخمين. ويمثل الخط الأفقي التقاربي في الجزء السفلي من المنحنى عند اللانهاية بارامتر التخمين وهو القيمة التي تعبر عن احتمال حصول الممتحَن منخفض القدرة على

الإجابة الصحيحة بمحض الصدفة. إلا أن لورد (Lord) لا يفضل تسميته "بارامتر التخمين" بل يقترح تسميته بارامتر مستوى شبه الصدفة (Pseudo-Chance Level Parameter) لما تحمله هذه التسمية من دقة أكبر في التعبير عن معناه (علام، 2013، ص74).

مثال حسابي: لنفترض أن لدينا مستوى قدرة شخص ما θ_i تساوي (3-)، وقيمة معلم الصعوبة للمفردة b_j تساوي (1.5-)، و قيمة معلم التمييز a_j يساوي (1.3)، و قيمة معلم التخمين C_j يساوي (0.2) (بيكر، 2001، ص.26).

- بداية نقوم بحساب الحد الأول في البسط من المعادلة السابقة $e^{a(\theta-b)}$

$$\mathbf{EXP} = 1.3\ (-3 - 1.5)$$

$$\mathbf{EXP} = -5.85 = 0.0029$$

- نقوم بتعوض الناتج في البسط وفي المقام مع إضافة الواحد، ثم نضرب الناتج في ناتج قيمة التخمين ناقص الواحد، وأخيرا الناتج منها نضيف إليه قيمة معلم التخمين، كالآتي:

$$P(x_{ji}) = 0.2 + (1 - 0.2)\ 0.0029 / 1 + 0.0029$$

$$P(x_{ji}) = 0.2 + (0.8)\ (0.0029) = 0.2023$$

إذن عند مستوى قدرة شخص ما (3-) فإن احتمال الاستجابة الصحيحة لهذه المفردة هو القيمة 0.2023 مع مراعاة عامل التخمين.

التفسير: يعني أن الشخص ضعيف جدًا في السمة المقاسة ولكن بسبب عامل التخمين تبقى لديه فرصة طفيفة في الإجابة الصحيحة على هذا السؤال تبلغ حوالي 20% فقط.

المحــاضرة الثالثة عشر

المسؤولية الأخلاقيـة في ممارسة القيـاس والتقييم النفسـي

الأهداف:

تهدف هذه المحاضرة إلى تعريف الطلبة بمفهوم المسؤولية الأخلاقية في ممارسة القياس والتقييم النفسي وأهميتها في ضمان جودة ودقة الممارسات التربوية والنفسية، مع التركيز على الاعتبارات الأخلاقية في القياس والتقويم التربوي من خلال توضيح المعايير الأخلاقية الواجب مراعاتها عند تصميم وتطبيق أدوات القياس. كما تسعى إلى إبراز قضايا العدالة والمساواة وأثرها في تحقيق تكافؤ الفرص وتجنب التحيز في النتائج، إضافة إلى توضيح أهمية السرية وحماية البيانات في الحفاظ على خصوصية المفحوصين وضمان أمن المعلومات. وتهدف أيضًا إلى التعرف على أخلاقيات البحث في مجال القياس والتقويم، بما يشمل المبادئ المنظمة لإجراء البحوث النفسية والتربوية بطريقة علمية وأخلاقية، بما يسهم في تنمية وعي الطلبة بضرورة الالتزام بهذه المبادئ في الممارسة المهنية والبحثية.

محتوى المحاضرة:

أولاً- الاعتبارات الأخلاقية في القياس والتقويم.

1- المعايير الأخلاقية في التقويم التربوي.

2- قضايا العدالــة والمساواة.

3- السريــة وحماية البيانات.

ثانياً- أخلاقيات البحث في القياس والتقويم.

مدة المحاضرة: ثلاث ساعات

تمهيد:

في ظل التطور المتسارع في مجالي القياس والتقييم النفسي والتربوي، أصبحت المسؤولية الأخلاقية عنصرًا أساسيًا لا يمكن فصله عن الممارسة العلمية والمهنية. فعملية القياس لا تقتصر على جمع البيانات وتفسير النتائج فقط، بل تتجاوز ذلك لتشمل احترام حقوق المفحوصين، وضمان العدالة، وحماية الخصوصية، والالتزام بالمعايير العلمية والأخلاقية في جميع مراحل العمل. ومن هنا تبرز أهمية هذه المحاضرة التي تتناول موضوع المسؤولية الأخلاقية في ممارسة القياس والتقييم النفسي، باعتبارها إطارًا مرجعيًا يوجه المختصين نحو الاستخدام السليم للأدوات والإجراءات، ويحد من الممارسات غير العادلة أو غير العلمية. كما تسلط الضوء على أبرز القضايا المرتبطة بهذا المجال مثل العدالة والمساواة، السرية وحماية البيانات، وأخلاقيات البحث في القياس والتقويم، بما يسهم في بناء وعي مهني يعزز الثقة في نتائج القياس ويخدم أهداف العملية التعليمية والنفسية.

أولاً- الاعتبارات الأخلاقية في القياس والتقويم:

تعد الجوانب الأخلاقية في القياس والتقويم من الركائز الأساسية التي تؤثر في دقة النتائج وموثوقيتها في المجال التعليمي. فالتقيد بالمبادئ الأخلاقية يضمن تجنب أي شكل من أشكال التحيز أو التمييز أثناء عملية القياس، ويستلزم الالتزام بالشفافية والنزاهة عند بناء أدوات التقويم. كما أن الاعتماد على معايير موضوعية وعادلة يسهم في الحصول على نتائج دقيقة تعكس مستوى تحصيل الطلبة بصورة حقيقية. ويعد الحفاظ على سرية البيانات الأكاديمية وخصوصيتها أمرًا ضروريًا لحماية حقوق المتعلمين وتفادي أي آثار سلبية قد تمس حالتهم النفسية أو فرصهم التعليمية. إضافة إلى ذلك، ينبغي تصميم أدوات القياس بما يراعي تنوع المتعلمين، بما في ذلك ذوو الاحتياجات الخاصة، نظرًا لما قد يترتب على نتائج التقويم من تأثيرات في المسارات الأكاديمية والمهنية، الأمر الذي يجعل من أي تقويم غير عادل مصدرًا لعواقب سلبية (سليمان، 2024).

ومن المهم كذلك أن يراعي المعلمون والمقوّمون الفروق الفردية عند إعداد الاختبارات، بما يضمن أن تعكس نتائج التقويم القدرات الحقيقية للطلبة. فمراعاة هذه الاعتبارات الأخلاقية تسهم في بناء بيئة تعليمية منصفة وشاملة تتماشى مع الأهداف التربوية. ومع التطور المتسارع في التكنولوجيا، تبرز الحاجة إلى توظيف أدوات القياس بشكل أخلاقي، خاصة مع استخدام تقنيات مثل الذكاء الاصطناعي في تطوير أساليب التقويم. لذا، يصبح من الضروري وضع أطر تنظيمية واضحة

تضمن الاستخدام العادل والمسؤول لهذه التقنيات، بما يعزز الثقة في نظم التقويم الحديثة وتطبيقاتها(Bruce, 2018) (المندلاوي، 2025، ص103).

1- المعـايير الأخلاقيـة في التقويـم التربـوي:

تعد المعايير الأخلاقيـة في التقويم التربـوي ركيزة أساسية لضمان النزاهة والموضوعية في العملية التعليمية. ويستلزم مراعاتها عند بناء أدوات القياس، إذ لا تقتصر أهميتها على دقة النتائج فحسب، بل تمتد لتشمل مسؤوليات المعلمين تجاه الطلبة وأولياء أمورهم. كما يتطلب التقويم العادل وضوح الإجراءات وإتاحة الفرصة للمتعلمين لفهم المعايير المعتمدة في الحكم على أدائهم. وفي حال إهمال هذه الجوانب، قد تنشأ أشكال من عدم المساواة في التقويم، مما يؤثر سلبًا في مخرجات التعليم ويُعمّق الفجوات بين الطلبة (Bruce, 2018).

وفي هذا الإطار، تبرز أهمية الحفاظ على سرية المعلومات واحترام خصوصية الطلبة، حيث ينبغي التعامل مع بياناتهم الشخصية بحذر وعدم الإفصاح عنها إلا للجهات المخولة. ويسهم الالتزام الأخلاقي في معالجة البيانات في تعزيز الثقة بين المعلم والمتعلم، وهو ما ينعكس إيجابًا على فعالية العملية التعليمية. كذلك، يجب تصميم أدوات التقويم بطريقة تتجنب أي تحيز قائم على العوامل الاجتماعية أو الاقتصادية، بما يضمن تحقيق العدالة وتكافؤ الفرص. ويتطلب ذلك توفير بيئة تعليمية داعمة تحفز الإبداع والتميز دون خوف من التمييز. ومن جهة أخرى، تؤكد المعايير الأخلاقية ضرورة تجنب أساليب التقويم الضارة والممارسات غير المنصفة، مع إدراك أثر التقويمات التعسفية أو النقد القاسي في إضعاف دافعية الطلبة للتعلم. لذا، يصبح من الضروري تبني استراتيجيات تقويمية تُسهم في تنمية مهارات المتعلمين بدلًا من الاقتصار على قياس مستوياتهم، بما يعزز من جودة التجربة التعليمية ويحقق الرفاهية الأكاديمية والنفسية للطلبة (المندلاوي، 2025، ص103).

2- قضايـا العدالــة والمساواة:

تعد قضايـا العدالـة والمساواة من المحاور الأساسية في مناقشة القيـاس والتقويـم التربوي، إذ يرتكز أي نظام تعليمي فعال على مبادئ تضمن تكافؤ الفرص لجميع الطلبة بغض النظر عن خلفياتهم الاجتماعية أو الاقتصادية. ويتطلب تحقيق ذلك تطوير أدوات قياس وتقويم تتسم بالحياد والدقة، بما يحد من تكريس الفجوات التعليمية. وتشير الأدبيات إلى أن اعتماد معايير منصفة يسهم في إتاحة الفرصة لكل متعلم للتعبير عن قدراته، مما يعزز بيئة تعليمية إيجابية تدعم الإبداع والتميّز الفردي.

وفي هذا السياق، تتجلى أبعاد المساواة في جانبين رئيسين: أولهما ضمان تقويم عادل للقدرات الأكاديمية، وثانيهما توظيف نتائج التقويم لدعم الطلبة الذين يواجهون صعوبات في التعلم. كما أن تقويم فاعلية الأساليب التعليمية يستلزم مراعاة الفروق الفردية بين المتعلمين، الأمر الذي يفرض تطوير أدوات تقويم مرنة تستجيب لاحتياجاتهم المتنوعة. ويُعد توظيف البيانات المستندة إلى الأساليب الإحصائية الحديثة ضروريًا لتحديد مواطن الضعف والحاجة إلى الدعم، بما يسهم في تحسين جودة المخرجات التعليمية. ويتطلب ذلك تعاونًا مستمرًا بين المعلمين والممارسين التربويين لضمان حصول كل طالب على فرصته الكاملة في تحقيق إمكاناته.

وفي الختام، يتضح أن الارتباط بين تحقيق العدالة والمساواة في القياس والتقويم وبين نجاح الأنظمة التعليمية في تحسين نتائجها هو ارتباط وثيق. لذا، ينبغي على صانعي القرار تبني استراتيجيات شاملة وفعالة لتقويم الأداء التعليمي تستجيب لاحتياجات جميع الطلبة، بما يعزز الشفافية ويرسخ الثقة في النظام التعليمي، ويسهم في بناء مجتمع قائم على مبادئ العدالة والمساواة (سليمان، 2024). كما أن تعزيز هذه القيم يستدعي مراجعة دورية للمنهجيات المعتمدة وإعادة النظر في أدوات القياس بما يتلاءم مع التحولات المتسارعة في متطلبات المجتمع (Bruce, 2018).

3- السريـة وحمايـة البيانات:

تعد سرية البيانات وحمايتها من أبرز التحديات التي تواجه المؤسسات التعليمية في ظل التطور التكنولوجي المتسارع. ولضمان حماية المعلومات الحساسة، يصبح من الضروري تبني استراتيجيات شاملة تعنى بصون بيانات الطلبة والعاملين، كاستخدام تقنيات التشفير، وتطبيق أنظمة أمنية متعددة المستويات، إلى جانب تدريب الكوادر على مبادئ أمن المعلومات. كما يتطلب الأمر الالتزام بالتشريعات المحلية والدولية المرتبطة بخصوصية البيانات. ويعد التخطيط المبني على تقويم المخاطر وتحديد مواطن الضعف خطوة أساسية للحد من احتمالات تسرب البيانات وتقليل آثارها المحتملة.

وتكتسب السرية أهمية خاصة في تعزيز الثقة بين الطلبة والمؤسسات التعليمية، إذ إن فقدان هذه الثقة قد ينعكس سلبًا على التحصيل الدراسي ومستوى المشاركة. ولتحقيق أمن معلوماتي فعّال، ينبغي إدماج معايير واضحة لحماية البيانات في جميع مراحل العملية التعليمية، بدءًا من جمع المعلومات وصولًا إلى تفسير النتائج. كما يتعين على الإدارات التعليمية اعتماد إجراءات استباقية للتعامل مع أي خروقات محتملة، كتوفير آليات للإبلاغ عن الانتهاكات وتقديم الدعم للمتضررين.

ويسهم هذا التوجه المتكامل في ترسيخ ثقافة احترام الخصوصية والالتزام بها داخل البيئة التعليمية (Bruce B. Frey, 2018).

وفي السياق ذاته، يوفر توظيف التكنولوجيا الحديثة في القياس والتقويم فرصًا كبيرة لتحسين جودة التعليم، غير أنه يفرض في الوقت نفسه تحديات متعلقة بأمن البيانات. لذا، يصبح من الضروري الاستثمار في أنظمة آمنة تضمن حماية المعلومات الشخصية، خاصة في ظل التوسع في استخدام التطبيقات الذكية والحوسبة السحابية. ومن خلال الموازنة بين الابتكار التعليمي ومتطلبات الحماية، يمكن تعزيز فعالية القياس والتقويم في بيئة آمنة، بما يدعم تطوير المنظومة التعليمية ويعزز الالتزام بمعايير الجودة (سليمان، 2024) (المندلاوي، 2025، ص103).

ثانياً- أخلاقيات البحث في القياس والتقويم:

يشير حامد زهران إلى ضرورة إجراء البحوث على الإنسان في إطار من الضوابط الأخلاقية التي تحكم البحث في علم النفس، موضحًا أن مؤسسات علمية كرابطة علماء النفس الأمريكيين (American Psychological Association - APA) (1973) قد وضعت مجموعة من الأخلاقيات التي يجب الالتزام بها، ومن أهمها ما يأتي:

1- الحصول على موافقة الأفراد الذين سيجرى عليهم البحث (المفحوصين).

2- تعريف المفحوص بطبيعة البحث وأهدافه بما يتناسب مع مستوى فهمه، لضمان أن تكون مشاركته طوعية.

3- في حال تطلبت إجراءات البحث إخفاء الهدف الحقيقي، ينبغي التأكد من إدراك المفحوص لأهمية ذلك.

4- احترام حق المفحوص في الانسحاب في أي وقت دون التعرض لأي ضغوط أو إغراءات.

5- حماية المفحوص من أي أذى أو ضرر محتمل، مع إبلاغه بذلك مسبقًا والحصول على موافقته، واتخاذ الاحتياطات اللازمة لتقليل المخاطر إلى أدنى حد.

6- إطلاع المفحوص على نتائج البحث، مع تصحيح أي سوء فهم قد يحدث.

7- في حالة الأطفال، يجب الحصول على موافقة أولياء الأمور، مع توضيح أهداف البحث ومخاطره لهم، وشرح الإجراءات للطفل بطريقة مبسطة تناسب عمره، وقد يقدم البحث له في صورة نشاط أو لعبة.

8- عند التعامل مع فئات خاصة كالمرضى أو المسنين أو المودعين في المؤسسات، ينبغي مراعاة

ظروفهم، وضمان حقهم في الانسحاب متى أرادوا في الحالات التي تكون فيها حرية المفحوص محدودة، مثل (السجناء أو المرضى أو الطلبة)، يجب أخذ ذلك بعين الاعتبار عند تنفيذ إجراءات البحث.

9- إذا ثبت أن الإجراءات التجريبية تحقق فائدة للمجموعة التجريبية، فلا ينبغي حرمان المجموعة الضابطة من هذه الفائدة.

10- عند إجراء أي تغييرات تجريبية في البيئة، يجب أن يتم ذلك بحذر شديد لتجنب إحداث أي ضرر محتمل (حامد، 2005).

الدكتور زيـاد رشيـد

خاتمـــة:

في ختام هذه المطبوعـة البيداغوجية، يتضح أن القيـاس النفســي يعد مجالًا علميًا محوريًا في علم النفس، يقوم على أسس نظرية ومنهجية دقيقة تهدف إلى فهم السلوك الإنساني وقياس الفروق الفردية بصورة موضوعية وموثوقة. وقد سعت المحاضرات المقدمة إلى بناء تصور تدريجي وشامل لدى الطلبة حول تطور القيـاس النفسـي، ومسلماته الأساسية، ومجالاته المختلفة، سواء في القياس العقلـي أو قيـاس الشخصيـة، إضافة إلى التعرف على أسس بناء الاختبارات النفسية والتربوية وخصائصها السيكومترية، وطرق التحقق من صدقها وثباتها.

كما تناولت المطبوعة أهم النماذج الكلاسيكية والحديثة في القيـاس، بما في ذلك نظرية الاستجابة للمفردة ونماذجها المختلفة، وما ارتبط بها من تطبيقات متقدمة كنموذج راش، وبنوك الأسئلة، والاختبارات التكيفية المحوسبة، بما يعكس التحول نحو قياس أكثر دقة وموضوعية ومرونة. ولم تغفل المطبوعة البعد الأخلاقي في ممارسة القيـاس والتقويـم النفسـي، باعتباره إطارًا ضابطًا يضمن العدالة والمهنية في التعامل مع الأفراد وبياناتهم.

وبذلك، تمثل هذه المطبوعـــة خطوة أساسية في تكوين الطالب معرفيًا ومنهجيًا، وتمكينه من فهم مبادئ القيــاس النفســي وتطبيقاته، بما يؤهله للتعامل العلمي الرشيد مع أدوات القيـاس في مساره الأكاديمي والمهني المستقبلي.

قائمة المصادر والمراجـــــع:

1- أبو حطب، ف .(1993) .التقويم النفسي. القاهرة: مكتبة الأنجلو المصرية.

2- أمين محمد، علي سليمان .(2009) .القياس والتقويم في العلوم الإنسانية: أسسه وأدواته وتطبيقاته (مراجعة رجاء محمود أبو علام). القاهرة: دار الكتاب الحديث

3- بركات، زياد. (2018). القياس والتقويم النفسي والتربوي بين النظرية والتطبيق. جامعة القدس المفتوحة، عمادة البحث العلمي.

4- بشير، معمرية (2007). المرجع في القياس النفسي وتصميم أدواته. للطلاب والباحثين في علم النفس والتربية. الطبعة الثانية.

5- بشير، معمرية (2012). أساسيات القياس النفسي وتصميم أدواته. للطلاب والباحثين في علم النفس والتربية.

6- بشير، معمرية (2022). المرجع في القياس النفسي وتصميم أدواته. الطبعة الرابعة.

7- بوسالم، ع .(2020/2021) .محاضرات في القياس النفسي موجهة لطلبة السنة الثانية علم النفس . جامعة البليدة 2 – لونيسي علي، كلية العلوم الاجتماعية والإنسانية، قسم العلوم الاجتماعية.

8- بوسالم، عبد العزيز (2014). القياس في علم النفس والتربية الأسس النظرية والمبادئ التطبيقية. منشورات مخبر القياس والدراسات النفسية. دار قرطبة للنشر والتوزيع. الجزائر.

9- بوسنة، م .(2007) .*علم النفس القياسي: المبادئ الأساسية* .الجزائر: ديوان المطبوعات الجامعية.

10- بيكر، ف. ب .(2001) .أسس نظرية الاستجابة للمفردة (عبد الله هاشم، سوسن أبو لهب شلبي، أحمد الطوري، عبد الرحمن بن سليمان أبو هاشم، مترجمون). الرياض.

11- حامد، زهران.(2005). *علم نفس النمو: الطفولة والمراهقة* (الطبعة السادسة). القاهرة: عالم الكتب.

12- حنورة، مصري عبد الحميد .(1998) .القياس والتقويم في التربية وعلم النفس .القاهرة: دار الفكر العربي.

13- ديانا فهمي، علي حماد .(2010) .علاقة الحكمة الاختبارية بالأداء الناتج من اختبار تحصيلي ذي اختيار من متعدد مبني وفق نموذج راش لدى طالبات كلية التربية للبنات الأقسام الأدبية بجامعة أم القرى(رسالة دكتوراه غير منشورة، كلية التربية، قسم علم النفس، جامعة أم القرى).

14- ديفيليس، ر. ف.(2021). *تطوير المقاييس: النظرية والتطبيق*. ترجمة: د. سعد بن سعيد القحطاني وأ.د. طارق عطية عبد الرحمن. مراجعة الترجمة: أ.د. عبد الرحمن بن سليمان العنقري. 1442هـ.

15- ربيع، محمد شحاتة .(2014). قياس الشخصية.ط5. الناشر: دار المسيرة. عمان. الأردن

16- رمزية، الغريب (1985). التقويم والقياس النفسي والتربوي. ط 11، القاهرة: مكتبة الأنجلو المصرية.

17- زيـاد، رشيد .(2023) .التقويم السيكومتري للاختبارات والمقاييس النفسية والتربوية .المسيلة، الجزائر: دار المتنبي للطباعة والنشر.

18- سامي، محمد ملحم .(2011) .القياس والتقويم في التربية وعلم النفس .عمّان: دار المسيرة للنشر والتوزيع.

19- سليمان، شاهر. (2024). النموذج البنائي العاملي لمعايير تقويم برامج الدراسات العليا التربوية: دراسة تقويمية لبرنامج ماجستير التربية في القياس والتقويم بجامعة تبوك. *مجلة كلية التربية (أسيوط)*.

20- الصبيحي، علي بن أحمد، والأحمدي، شرف بنت حامد. (2011). الاختبار التكيفي المحوسب (Computerized Adaptive Testing (CAT)): كتطبيق لتطوير أساليب التقويم والاختبارات الإلكترونية. *مجلة كلية التربية - جامعة دمنهور، 3*(1).

21- الطريري، عبد الرحمن بن سليمان .(1997) .القياس النفسي والتربوي: نظريته، أسسه، تطبيقاته .الرياض: مكتبة الرشد.

22- عبد الخالق، أحمد و الأنصاري، بدر (1996). العوامل الخمسة الكبرى في مجال الشخصية، الهيئة المصرية العامة للكتاب، مجلة علم النفس، العدد (38).

23- عداي، ن. ع .(2019) .القياس النفسي: نظريات القياس النفسي والتربوي [مذكرة ماجستير غير منشورة]. الجامعة المستنصرية، كلية التربية الأساسية، قسم الإرشاد النفسي والتوجيه التربوي.

24- علام، صلاح الدين. (2005). *نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي* .القاهرة: دار الفكر العربي.

25- علام، صلاح الدين. (2013). *نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي* .القاهرة: دار الفكر العربي

26- على، عماد أحمد حسن .(2016) .اختبار المصفوفات المتتابعة الملونة لـ .Raven القاهرة: مكتبة الأنجلو المصرية.

27- علي أحمد، سيد مصطفى؛ محمد محمود، محمد علي (2012). مقاييس الذكاء والقدرات العقلية، الناشر: دار الزهراء، الرياض.

28- علي، عبد الحميد محمد (2010). القياس النفسي والاختبارات النفسية. القاهرة: مؤسسة طيبة للنشر والتوزيع.

29- علي، محمد محمود، ومصطفى، علي أحمد .(2012) .مقاييس الذكاء والقدرات العقلية .دار الزهراء للنشر والتوزيع.

30- العنكبي ، حيدر جليل عباس (2009). المفاضلة في تحديد درجات القطع لاختبار محكي المرجع ،اطروحة ، كلية التربية ، جامعة بغداد ابن الرشد.

31- غباري، ثائر أحمد، وأبو شعيرة، خالد محمد .(2009) .سيكولوجية الشخصية: علم نفس الأفرد. مكتبة المجتمع العربي.

32- فالك، س .(2024) .*سيكولوجية الذكاء* (أ. س. سعد، مترجم؛ د. فرج، مُحرر؛ م. شرف، مراجعة لغوية؛ ر. عبدالله، إخراج). مركز المحروسة. القاهرة. مصر

33- كاظم، أمينة محمد. (1988أ) .دراسة نظرية نقدية حول القياس الموضوعي للسلوك: نموذج راش .مؤسسة الكويت للتقدم العلمي.

34- كاظم، أمينة محمد. (1988ب) .استخدام نموذج راش في بناء اختبار تحصيلي في علم النفس وتحقيق التفسير الموضوعي للنتائج .مطبوعات جامعة الكويت.

35- الكحلوت، أحمد إسماعيل .(1991) .الخصائص السيكومترية لمقياس (مايرز–بريجز) لأنماط الشخصية وقدرة المقياس على التنبؤ بالتحصيل الدراسي (أطروحة دكتوراه، كلية الدراسات العليا، الجامعة الأردنية).

36- محمود عبد الحليم منسي، د. عفاف محمد عبد المنعم (2006) علم النفس والقدرات العقلية، الإسكندرية، دار المعرفة الجامعية

37- محمود، حسين بشير .(2001) .القياس والتقويم في التربية وعلم النفس .القاهرة: دار الفكر العربي.

38- محمود، علام (2000). القياس والتقويم النفسي والتربوي أساسايته وتطبيقاته وتوجهاته المعاصرة: القاهرة: دار الفكر العربي.

39- مراد، صلاح أحمد، وعبد الغفار، محمد حسن .(1985) .الاختبارات والمقاييس في العلوم النفسية والتربوية .القاهرة: دار النهضة العربية.

40- المندلاوي، علاء عبد الخالق حسين .(2025) .*القياس والتقويم التربوي: الأسس النظرية والتطبيقات العملية* (الطبعة الأولى). مؤسسة دار الصادق الثقافية.

41- المندلاوي، علاء عبد الخالق .(2025) .القياس والتقويم التربوي: الأسس والنظرية والتطبيقات العملية .بغداد، العراق: مؤسسة دار الصادق الثقافية.

42- ميمي السيد، أحمد إسماعيل .(2014) .التوجهات الحديثة في القياس النفسي والتربوي .القاهرة: دار الكتاب الحديث.

43- وليد، مسعود .(2014) .محاضرات في القياس: نظرية الاستجابة للمفردة مع تطبيقات عملية . كلية التربية، قسم علم النفس، جامعة أم القرى.

44- Frey, B. B. (2018). *The SAGE encyclopedia of educational research, measurement, and evaluation*. SAGE Publications.

45- Baker, F. B., & Kim, S. (2004). *Item response theory: Parameter simulation techniques*. New York, NY: Marcel Dekker, Inc.

46- Guilford, J. P. (1936). *Psychometric methods* (1st ed.). McGraw-Hill.

47- McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of giants. Intelligence, 37(1), 1–10.

48- Sternberg, R. J. (2020). Cognitive psychology (7th ed.). Cengage Learning.

Roid, G. H. (n.d.). Stanford-Binet Intelligence Scales, Fifth Edition (SB-5). WPS Publish. https://www.wpspublish.com/sb-5-stanford-binet-intelligence-scales-fifth-edition

49- Jacobsen, J., & George, E. L. (2024, May 31). Wechsler Adult Intelligence Scale (WAIS) – Psychological testing. MentalHealth.com. https://www.mentalhealth.com/library/wechsler-adult-intelligence-scale-psychological-testing

50- U.S. Department of Defense Education Activity. (n.d.). Scholastic Assessment Test (SAT). https://www.dodea.edu/education/assessment/resources/scholastic-assessment-test-sat

51- Tippett, J. (2025, June 11). What is the GRE test? – A guide to the Graduate Record Examination. FindAMasters. https://www.findamasters.com/guides/masters-gre-test

52- Anastasi, A., & Urbina, S. (1997). Psychological testing (7th ed.). Prentice Hall.

53- Groth-Marnat, G., & Wright, A. J. (2016). Handbook of psychological assessment (6th ed.). Wiley.

54- Sattler, J. M. (2018). Assessment of children: Cognitive foundations and applications (6th ed.). Sattler Publisher.

55- Holland, J. L., & Messer, M. A. (2017). Self-Directed Search®: 5th Edition [Assessment kit]. PAR, Inc. Retrieved from https://www.parinc.com/products/SDS-5-Edition

56- Yazak-1, D. L. (2024). Strong Interest Inventory (SII). In EBSCO Research Starters – Health & Medicine. EBSCO. Retrieved from https://www.ebsco.com/research-starters/health-and-medicine/strong-interest-inventory-sii

57- Yazak-2, D. L. (2024). Kuder Occupational Interest Survey (KOIS). In EBSCO Research Starters – Health & Medicine. Retrieved from https://www.ebsco.com/research-starters/health-and-medicine/kuder-occupational-interest-survey-kois

58-Ajzen, I., & Fishbein, M. (1980). Understanding attitudes and predicting behavior. Englewood Cliffs, NJ: Prentice Hall.

59- Blalock, S. J., DeVellis, R. F., Brown, G. K., & Wallston, K. A. (1989). Validity of the Center for Epidemiological Studies Depression Scale in arthritis populations. Arthritis and Rheumatism, 32, 991–997.

60- DeVellis, R. F., & Callahan, L. F. (1993). A brief measure of helplessness: The helplessness subscale of the Rheumatology Attitudes Index. Journal of Rheumatology, 20, 866–869.

61- DeVellis, R. F., DeVellis, B. M., Revicki, D. A., Lurie, S. J., Runyan, D. K., & Bristol, M. M. (1985). Development and validation of the child improvement locus of control (CILC) scales. Journal of Social and Clinical Psychology, 3, 307–324.

62- DeVellis,R.F.,Holt,K.,Renner,B.R.,Blalock,S.J.,Blanchard,L.W.,Cook, H. L., et al. (1990). The relationship of social comparison to rheumatoid arthritis symptoms and affect. Basic and Applied Social Psychology, 11, 1–18.

63-Fry,E.(1977).Fry'sreadabilitygraph:Clarifications,validity,andextension to level 17. Journal of Reading, 21, 249.

64- Kelly, J. R., & McGrath, J. B. (1988). On time and method. Newbury Park, CA: Sage

65-Mayer, J. M. (1978). Assessment of depression. In M. P. McReynolds (Ed.), Advances in psychological assessment (Vol. 4, pp. 358–425). San Francisco: Jossey-Bass.

66- Radloff, L. (1977). The CES-D scale: A self-report depression scale for research in the eneral population. Applied Psychological Measurement, 1, 385–401.

67- Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). Statetrait anxiety inventory (STAI) test manual for form X. Palo Alto, CA: Consulting Psychologists Press.

68- Strahan, R., & Gerbasi, K. (1972). Short, homogenous version of the Marlowe-Crowne Social Desirability Scale. Journal of Clinical Psychology, 28, 191–193.

69- Wallston, K. A., Wallston, B. S., & DeVellis, R. (1978). Development and validation of the multidimensional health locus of control (MHLC) scales. Health Education Monographs, 6, 161–170.

70- Wallston, K. A., Stein, M. J., & Smith, C. A. (1994). Form C of the MHLC Scales: A condition-specific measure of locus of control. Journal of Personality Assessment, 63, 534–553.

71- Zorzi, M., Priftis, K., & Umilita, C. (2002, May). Brain damage: Neglect disrupts the mental number line. Nature, 417, 138–139.

72-Zuckerman, M. (1983). The distinction between trait and state scales is not arbitrary: Comment on Allen and Potkay's "On the arbitrary distinction between traits and states." Journal of Personality and Social Psychology, 44, 1083–1086.

73- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York, NY: Macmillan.

74- Anastasi, A., & Urbina, S. (1997). Psychological Testing (7th ed.). Prentice Hall.

75- Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric Theory (3rd ed.). McGraw-Hill.

76- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (1999). Evidence-centered assessment design. Princeton, NJ: Educational Testing Service

77- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (1999). Evidence-centered assessment design. Princeton, NJ: Educational Testing Service.

78- Charles, S ; Brian, D.(2012). Handbook on measurement, assessment, and evaluation in higher education. First published by Routledge 711 Third Avenue, New York , NY 10017.

79- EL – Korashy , AF . (1995) : Applying the Rasch model to the selection of items for a mental ability test. Educational and psychological Measurement ,Vol .55 , No . 5 , pp 753 -763

80- Kline, Rex B .(2011). Principles and practice of structural equation modeling. 3rd ed. A Division of Guilford Publications, Inc. 72 Spring Street, New York, NY 10012

81- Louis L. Thurstone. (1925). A Method of Scaling Psychological and Educational Tests. Journal of Educational Psychology. 16: 433- 451

82- Wilson, Mark .(2005). Constructing measures: an item response modeling approach. Lawrence Erlbaum Associates, Inc., Publishers 10 Industrial Avenue Mahwah, New Jersey, 07430, www.erlbaum.com

83- Tes and Rabu. (2011, June 8). Reliabilitas instrument tes (Bag. 1). Serpihan Media dan untuk Berbagi Pengetahuan. Ruang https://gerbangbuku.blogspot.com/2011/06/reliabilitasinstrumen-tes-bag-1.html

84- Portney, L. G., & Watkins, M. (2008). Foundations of Clinical Research: Applications to Practice (3rd ed.). Prentice-Hall. Practical approach. Educational Measurement Solutions, Melbourne. pp01-87.

85- Rust, J., Kosinski, M., & Stillwell, D. (2021). Modern psychometrics: The science of psychological assessment (4th ed.). Routledge.

86- Garson, G. D. (2013). Validity and reliability. Statistical Associates Publishing. https://www.statisticalassociates.com

87- Sivakumar, A., & Thirumoorthy, G. (2019). Measurement and evaluation in education. New Delhi, India: A.P.H. Publishing Corporation.

88- L'Année psychologique. (1905). *L'Année psychologique*, 12. Paris: Félix Alcan.

89- Raven, J. C. (1977). *Standard progressive matrices: Sets A, B, C, D & E*. Oxford, England: Oxford Psychologists Press.

90- Otis, A. S., & Lennon, R. T. (1996). *Otis–Lennon School Ability Test (8th ed.)*. San Antonio, TX: The Psychological Corporation.

91- Kelly, J. R., & McGrath, J. E. (1988). *On time and method: The interplay of research design and statistical analysis*. Newbury Park, CA: Sage Publications.

92- Nitko, A. J. (1983). *Educational tests and measurements: An introduction*. New York, NY: Harcourt Brace Jovanovich.

93- Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Belmont, CA: Wadsworth/Cengage Learning.

94- Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied, 80*(1), 1–28. https://doi.org/10.1037/h0092976

95- Levenson, H. (1973). Distinctions within the concept of internal–external control: Development of a new scale. *Proceedings of the 81st Annual Convention of the American Psychological Association, 8*, 261–262.

96- Radloff, L. S. (1974). A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*(3), 385–401.

97- Cattell, J. M. (1978). Mental tests and measurements. *Mind, 15*, 373–381.

98- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.

99- Bernstein, I. H., Jaremko, M. E., & Hinkley, B. S. (1994). On the utility of the SCL-90-R with low-back pain patients. *Spine*, *19*(1), 42-48.

100-DeVellis, R. F. (1991). *Scale development: Theory and applications*. Newbury Park, CA: Sage.

101-Thayn, S. (2011). *An evaluation of multiple choice test questions deliberately designed to include multiple correct answers*. Brigham Young University.

102-Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Measurement and evaluation in psychology and education* (4th ed.). New York, NY: Macmillan.

الدكتور زيـاد رشيـد