

القياس التربوي وبناء الاختبارات المدرسية

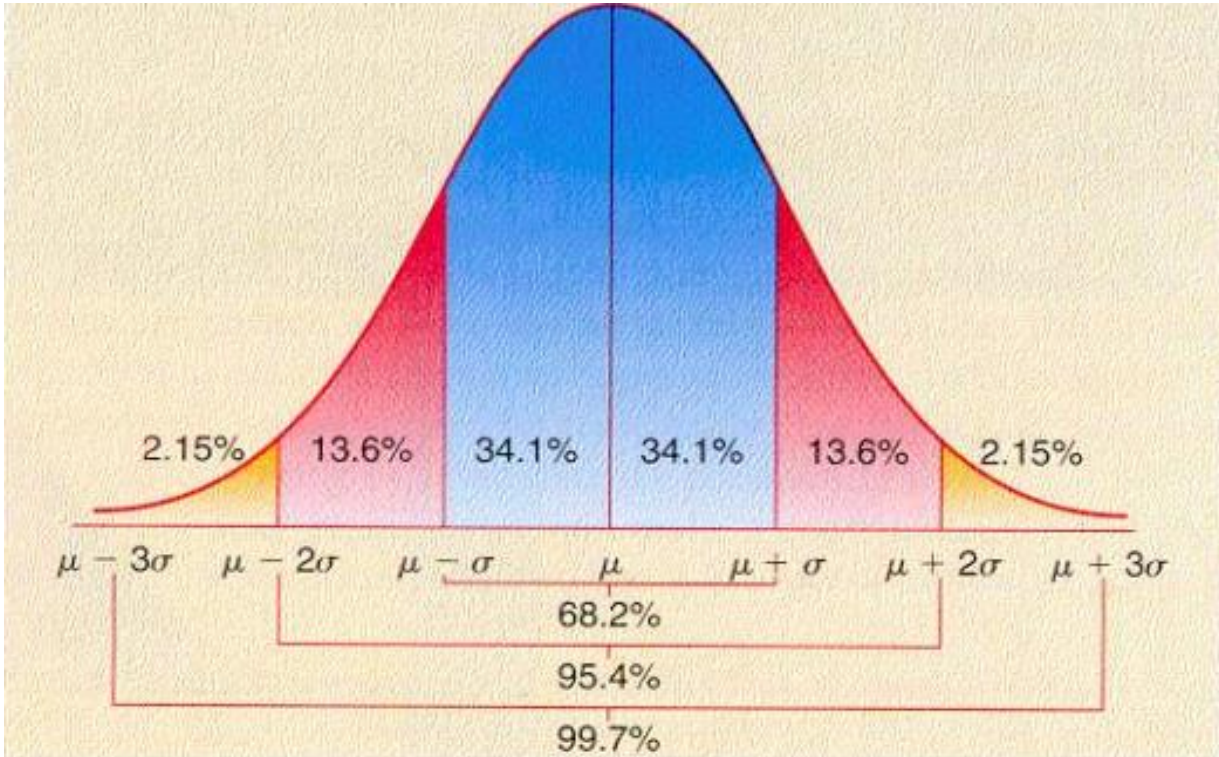
Educational Measurement and Construction of School Tests

كتاب بيداغوجي موجه لطلبة السنة الثانية ليسانس علوم التربية

د/ زياد رشيد

أستاذ القياس والتقويم والإحصاء النفسي والتربوي (MCA)

جامعة الشهيد الشيخ العربي التبسي - تبسة



الإهداء

إلى روح والدي الغالي، رحمه الله وأسكنه فسيح جناته،
إلى والدتي الحبيبة، أطل الله في عمرها وحفظها وجعلها ذخراً لنا،
إلى زوجتي العزيزة وأبنائي الأحبة، نبض قلبي وسند دربي،
إلى إخوتي الأعزاء، شركاء الحياة،
وإلى طلبتي الأعزاء، الذين بهم يزدان العمل ويثمر الجهد،
أهدي هذا العمل المتواضع بكل الحب والوفاء.

فهرس المحتويات

الفصل الأول

الأسس النظرية للقياس والتقويم التربوي والنفسي

- 1- مفهوم القياس 09
- 2- التطور التاريخي للقياس والتقويم التربوي والنفسي 10
- 3- القياس التربوي (Educational Measurement) 11
- 4- التقويم التربوي (Educational Evaluation) 12
- 5- أسس القياس التربوي (Principles of Educational Measurement) 12
- 6- دور القياس والتقويم في عملية التقييم 13
- 7- طبيعة القياس التربوي 14
- 8- طبيعة التقويم التربوي 17
- 9- أشكال أو أنواع القياس والتقويم التربوي 18
- 10- خصائص التقويم التربوي 22

الفصل الثاني

المفاهيم الأساسية في الاختبارات النفسية والتربوية: التصنيف، التقنين، المعايير، ومستويات القياس

- 1- تصنيف الاختبارات والمقاييس النفسية والتربوية 25
- 2- تقنين الاختبارات التربوية والنفسية 29
- 3- معنى المعايير وأهميتها 30
- 4- المعايير وعينة التقنين 31
- 5- أنواع المعايير 31
- 6- أنواع المعايير أو الدرجات المحولة (Norms) 36
- 7- المعايير والصفحة النفسية أو التربوية (Psychological or Educational Profile) 42
- 8- تصنيف الاختبارات التحصيلية 45
- 9- مستويات القياس التربوي والنفسي 47
- 10- خصائص مستويات القياس 53
- 11- مستويات القياس والإحصاء 54

الفصل الثالث

الأهداف التعليمية أنواعها ومستوياتها

- 1- الأهداف التعليمية: أنواعها ومستوياتها..... 56
- 2- مستويات الأهداف التعليمية..... 60
- 1-2- المجال المعرفي أو الإدراكي (العقلي) (The Cognitive Domain) 60
- 2-2- المجال الانفعالي أو الوجداني (العاطفي) (The Affective Domain) 67
- 3-2- المجال النفس حركي أو المهاري (The Psychomotor Domain) 69
- 3- شروط صياغة الأهداف السلوكية..... 75
- 4- تحليل محتوى المادة الدراسية..... 76
- 5- إعداد جدول المواصفات..... 77

الفصل الرابع

النظرية الكلاسيكية في القياس (CTT)

- 1- نظريات القياس التربوي والنفسي..... 85
- 2- النظرية الكلاسيكية للقياس Classical Test Theory (CTT) 86
- 1-2- مميزات النظرية الكلاسيكية للقياس في بناء الاختبارات..... 88
- 2-2- أوجه قصور وعيوب النظرية الكلاسيكية للقياس..... 88
- 3- الخطأ المعياري للقياس (Standard Error of Measurement) 89

الفصل الخامس

نظرية الاستجابة للمفردة الاختبارية (IRT)

- 1- التطور التاريخي لنظرية الاستجابة للمفردة (IRT)..... 95
- 2- افتراضات نماذج نظرية الاستجابة للمفردة الاختبارية أحادية البعد..... 97
- 3- أنواع نماذج الاستجابة للمفردة أحادية البعد..... 105
- 4- معنى الموضوعية في نموذج راش..... 113
- 5- الجانب السيكلوجي في نموذج راش..... 114
- 6- وحدات القياس المستخدمة في نموذج راش..... 115
- 7- تقدير بارامترات النماذج أحادية البعد..... 116
- 8- الاختبارات التحصيلية المعتمدة في بنائها على نظرية الاستجابة للمفردة..... 119

- 9- تدريج الاختبار وفق نموذج راش 120
- 10- كيفية التحقق من حسن مطابقة البيانات الاختبارية الافتراضات نموذج راش 121
- 11- دالة المعلومات (Information Function) 125

الفصل السادس

الخصائص السيكمترية للاختبارات والمقاييس النفسية والتربوية

- 1- خصائص الاختبارات والمقاييس الجيدة 130
- 2- الشروط العامة للاختبار النفسي والتربوي 131
- 3- الشروط التجريبية أو السيكمترية (Psychometric Conditions) 131
- 3-1- الصدق (Validity) 132
- 3-2- العوامل المؤثرة على الصدق 141
- 3-2- الثبات (Reliability) 142
- 3-3- العوامل المؤثرة على ثبات الاختبار 158

الفصل السابع

الخصائص السيكمترية لفقرات الاختبار

- 1- معامل صعوبة الفقرة Item Difficulty Index 160
- 2- معامل سهولة الفقرة Item Easiness Index 163
- 3- معامل تصحيح الصعوبة من أثر التخمين Guessing 164
- 4- قوة تمييز فقرات الاختبار Item Discrimination Index 166
- 5- طرق أخرى لتقييم تمييز فقرات الاختبار 171
- 6- المموهات (المشتتات) Distractions 176
- 7- معامل ثبات المفردة 175

الفصل الثامن

الاختبارات المدرسية أنواعها وأغراضها

- 1- مفهوم الاختبار (Test) 178
- 2- الاختبار التحصيلي (Achievement tests) 179
- 3- بناء الاختبارات التحصيلية بناءً على النظرية الكلاسيكية في القياس 179
- 3-1- الاختبارات التحصيلية الموضوعية مرجعية المعيار (Norm-Referenced Tests: NRT) 180

- 3-2- الاختبار التحصيلية مرجعية المحك (CRT) Criterion-Referenced Tests191
- 4- العدالة في التقييم التربوي.....195
- 5- العدالة الاختبارية (Test Fairness)195
- 6- العدالة والإنصاف (Fairness and Equity)196
- 7- معاني العدالة في الاختبارات.....197
- 8- الأداء التفاضلي للمفردة والعدالة الاختبارية.....206

الفصل التاسع

ملفات أعمال الطالب بورتفوليو (Portfolio)

- 1- التقييم التوثيقي الشامل والكفايات Portfolio Assessment209
- 2- مفهوم واستخدامات ملف الأعمال (Portfolio).....211
- 3- مكونات الصحائف الوثائقية أو ملفات الإنجاز (Portfolios).....215
- 4- أنواع ملفات أعمال الطالب (ملفات الإنجاز)215
- 5- أهداف ملف الإنجاز.....217
- 6- المكونات الأساسية لملف الإنجاز.....217
- 7- خصائص ملف الإنجاز.....219
- 8- خطوات تصميم واستخدام ملف الإنجاز.....221
- 9- تصحيح ملف الإنجاز.....222
- 10- تقييم ملف الإنجاز.....222
- 11- ملف الإنجاز الإلكتروني (E-Portfolio)223
- 12- بعض مشكلات تقييم التحصيل باستخدام الصحائف الوثائقية (ملفات الأعمال)223

الفصل العاشر

القياس التربوي في العصر الرقمي

- 1- الاختبارات الإلكترونية (E-tests)227
- 2- الاختبارات التكيفية المحوسبة (CAT).....239
- 3- القياس التربوي والنفسي في عصر الآلة الذكية.....248

المقدمة

يُعَدّ القياس التربوي من الركائز الأساسية في ميدان العلوم التربوية، إذ يوفّر أدوات علمية دقيقة تساعد على فهم قدرات المتعلمين وتشخيص مواطن القوة والضعف لديهم واتخاذ قرارات تربوية قائمة على أسس موضوعية. وقد أصبح الاهتمام بالاختبارات المدرسية مطلبًا ملحًا في ظل التطورات المتسارعة التي يشهدها التعليم، حيث أضحت أدوات تقويم صادقة وثابتة وعادلة جزءًا لا يتجزأ من ضمان جودة العملية التعليمية وتحقيق أهدافها.

ومن هذا المنطلق جاء هذا الكتاب ليكون مرجعًا تعليميًا موجّهًا لطلبة السنة الثانية علوم التربية ليضع بين أيديهم إطارًا نظريًا متينًا ومهارات عملية أساسية في مجال القياس التربوي وبناء الاختبارات المدرسية. وقد صُمّم محتواه ليوازن بين الجانب النظري المتمثل في المفاهيم والمبادئ السيكمترية، وبين الجانب التطبيقي المتمثل في طرق بناء الاختبارات تحليلها وتوظيفها في الميدان التربوي.

يهدف الكتاب إلى تمكين الطالب من إدراك الأسس التي يقوم عليها القياس والتقويم التربوي، وفهم تطور هذا المجال ودوره في تحسين التعلم. فهو يتيح له التعرف على الفروق الجوهرية بين مفاهيم القياس والتقويم والتقييم والتعمق في قضايا التصنيف والتقنين والمعايير، بما في ذلك مستويات القياس المختلفة. كما يتيح للطالب الوقوف على الأهداف التعليمية بمستوياتها الثلاثة: المعرفية والانفعالية والمهارية مع التدريب على صياغة الأهداف السلوكية وتحليل المحتوى وبناء جداول المواصفات باعتبارها مداخل أساسية لتصميم اختبارات مدرسية دقيقة.

ويمتد الهدف أيضًا إلى تعريف الطالب بالنظريات التي تشكل الأساس العلمي للقياس، بدءًا بالنظرية الكلاسيكية وما تقدمه من مفاهيم مثل الخطأ المعياري للقياس، وصولاً إلى نظرية الاستجابة للمفردة (IRT) بمختلف نماذجها وخاصة نموذج راش وما يرتبط به من مفاهيم حديثة كدالة المعلومات وموضوعية القياس. كما يكتسب الطالب معرفة معمقة بالخصائص السيكمترية للاختبارات كالصدق والثبات والعوامل المؤثرة فيهما والتمييز بين الشروط العامة والسيكمترية للاختبار الجيد.

ولا يغفل الكتاب الجانب التطبيقي الميداني، إذ يدرب الطالب على تحليل البنود من حيث الصعوبة والتمييز والكشف عن أثر التخمين وجودة المموهات وفحص ثبات المفردة. كما يتناول أنواع الاختبارات المدرسية وأغراضها موضحًا الفرق بين الاختبارات معيارية المرجع ومحكية المرجع ومبرزًا قضايا العدالة والإنصاف والحياد في التقييم التربوي. ويخصص الكتاب حيزًا مهمًا

لملفات أعمال الطالب (Portfolio) باعتبارها أحد البدائل الحديثة في التقييم، حيث يوضح أهدافها وخصائصها وخطوات تصميمها وكيفية توظيفها في التقييم الإلكتروني.

وفي ختام هذه الرحلة يفتح الكتاب آفاقًا جديدة أمام الطالب للتعرف على التطورات الراهنة في مجال القياس كالاختبارات الإلكترونية والاختبارات التكيفية المحوسبة، إلى جانب استشراف دور الذكاء الاصطناعي في صياغة مستقبل القياس النفسي والتربوي. وبذلك يكون الطالب قد اكتسب معارف نظرية ومهارات عملية وقدرات نقدية تمكنه من بناء اختبارات مقننة وتحليل نتائجها على نحو يساهم في تحسين جودة التعليم ويؤهله للممارسة التربوية الرصينة.

ينطلق السداسي الثالث لتحقيق جملة من الأهداف الأساسية للمادة التي تتمثل في:

- 1- إطلاع الطالب على مبادئ القياس التربوي وأسس بنائه.
- 2- تعريف الطالب بكيفية بناء الاختبارات المدرسية وفق أسس علمية دقيقة.
- 3- تمكين الطالب من الربط بين القياس والإحصاء والمنهجية في المجال التربوي.
- 4- تدريب الطالب على التمييز بين أنواع الاختبارات التربوية وأغراضها.
- 5- إكساب الطالب مهارات تحليل وتقويم أدوات القياس المختلفة.

- المعارف السابقة المطلوبة:

- الإلمام بأساسيات الإحصاء الوصفي والاستدلالي.
- معرفة أولية بمناهج البحث التربوي وأسس المنهجية.
- خبرة عامة في مفاهيم التربية والتعليم.
- القدرات المكتسبة بناءً على مفردات المادة للسداسي الثالث:
- 1- التمكن من أسس القياس التربوي.

- 2- التعرف على النظرية الكلاسيكية للقياس ونظرية الاستجابة للمفردة.
- 3- الإلمام بالاختبارات المعيارية والاختبارات محكية المرجع.
- 4- القدرة على بناء وتصميم الاختبارات التربوية.
- 5- التمييز بين أنواع الاختبارات (الأدائية، الموضوعية، المقننة...).
- 6- اكتساب مهارات التقييم باستخدام أدوات متنوعة مثل ملف التقييم (Portfolio).
- 7- التعرف على أخطاء القياس وكيفية معالجتها.
- 8- تنمية مهارات تحليل البنود واختبار الثبات بمختلف أنواعه.

ثم يتدرج السداسي الرابع ليحقق أهدافاً أكثر تعمقاً في المادة، وهي:

- 1- تعميق معارف الطالب في مجال القياس التربوي وبناء الاختبارات المدرسية.
- 2- إكساب الطالب القدرة على التأكد من صلاحية الاختبارات التربوية وجودتها.
- 3- تمكين الطالب من اكتشاف تقنيات متقدمة في التقييم التربوي.
- 4- ربط المفاهيم النظرية بتطبيقات عملية من خلال متابعة ما تعلمه في مقياس القياس التربوي وبناء الاختبارات 1.

- المعارف السابقة المطلوبة:

- 1- استيعاب المفاهيم الأساسية في القياس التربوي وبناء الاختبارات المدرسية 1.
 - 2- معرفة أولية بالثبات، الصدق، وأسس بناء الاختبارات.
 - 4- إلمام بالمبادئ الإحصائية والمنهجية التي تدعم التحليل والتفسير في القياس.
- القدرات المكتسبة بناءً على مفردات المادة للسداسي الرابع:

- 1- التمكن من مفاهيم الصدق بأنواعه.
- 2- التمييز بين الصدق وصدق التقييم التربوي.
- 3- فهم نقد الاختبارات التربوية بما في ذلك مفهوم العدالة (Fairness).
- 4- التعرف على مؤشرات صدق الاختبار.
- 5- اكتساب القدرة على استخدام تقارير ونتائج التقييم التربوي.
- 6- إعداد وتحليل نماذج تقارير التقييم التربوي.
- 7- التمكن من تصنيف الاختبارات ونماذج التقييم التربوي.
- 8- الإلمام بأنواع متعددة من التقييم (التشخيصي، البنائي، الختامي).
- 9- التعرف على تطورات حديثة في مجال قياس التحصيل.
- 10- الإحاطة بمشكلات الاختبارات المدرسية وسبل معالجتها.
- 11- القدرة على التعامل مع الاختبارات القائمة على الحاسوب.

تبسة في: 01 أكتوبر 2025

الدكتور/ زياد رشيد

أستاذ القياس والتقويم والاحصاء النفسي والتربوي
بجامعة الشهيد الشيخ العربي التبسي-تبسة- الجزائر

الفصل الأول:

الأسس النظرية للقياس والتقويم التربوي والنفسي

الأهداف (Objectives):

يهدف هذا الفصل إلى تمكين الطالب من التعرف على مفهوم القياس والتقويم التربوي والنفسي وتتبع تطورهما التاريخي، والتمييز بين القياس التربوي والتقويم التربوي من حيث الطبيعة والأهداف، وفهم الأسس التي يقوم عليها القياس التربوي ودوره في عملية التقييم، إضافة إلى استيعاب طبيعة كل من القياس والتقويم، والتعرف على أشكالهما وأنواعهما، مع إبراز خصائص التقويم التربوي الجيد.

المحتوى (Contents):

- 1- مفهوم القياس.
- 2- التطور التاريخي للقياس والتقويم التربوي والنفسي.
- 3- القياس التربوي (Educational Measurement).
- 4- التقويم التربوي (Educational Evaluation).
- 5- أسس القياس التربوي (Principles of Educational Measurement).
- 6- دور القياس والتقويم في عملية التقييم.
- 7- طبيعة القياس التربوي.
- 8- طبيعة التقويم التربوي.
- 9- أشكال أو أنواع القياس والتقويم التربوي.
- 10- خصائص التقويم التربوي.

تمهيد:

يُعدّ القياس والتقويم من الركائز الأساسية في العملية التربوية والنفسية، إذ لا يمكن تصور أي عملية تعليمية أو علاجية أو بحثية دون الاعتماد عليهما في جمع البيانات وفهمها وتفسيرها. فالقياس يوفر أدوات كمية تساعد في تحديد مستوى المتعلم أو المفحوص، بينما يقدّم التقويم رؤية شمولية للحكم على مدى تحقق الأهداف واتخاذ القرارات المناسبة. وقد مرّ كل من القياس والتقويم بمراحل تاريخية مهمة أسهمت في بلورة مفاهيمهما وأهدافهما وأساليبهما، حتى أصبحا اليوم علماً قائماً على أسس ومبادئ علمية واضحة. ومن هنا تأتي أهمية هذا الفصل، حيث يتناول المفاهيم الأساسية للقياس والتقويم، تطورهما التاريخي، أسسهما وطبيعتهما، بالإضافة إلى أشكالهما وأنواعهما وخصائص التقويم التربوي الجيد، مما يمكّن الطالب من الإلمام بالجانب النظري والعملية لهذا المجال الحيوي.

1- مفهوم القياس:

يشير القياس إلى العملية التي يتم من خلالها تحديد صفات أو أبعاد جسم مادي معين. وعندما يُستخدم في سياق التعلم، فإنه يعني تطبيق مقياس معياري أو أداة قياس على جسم أو مجموعة من الأجسام، أو على أحداث وظروف وفقاً للممارسات المتعارف عليها بين المتخصصين في استخدام تلك الأداة أو المقياس. يقول ثورندايك (E.L. Thorndike)، بأن "أي شيء موجود بأي شكل من الأشكال موجود بكمية معينة؛ وأي شيء موجود بكمية يمكن قياسه". وبمعنى آخر، القياس يتعلق بتحديد مقدار شيء ما، سواء كان كبيراً أو صغيراً، كثيراً أو قليلاً، أكثر أو أقل، حيث توضح موسوعة البحوث التربوية هذا المفهوم بشكل أدق، فتقول: "القياس يعني ملاحظة أو تحديد مقدار متغير معين". يقوم القياس بالجواب على السؤال: "كم؟". ففي حياتنا اليومية، نقوم بقياس الطول، الوزن، المسافة المقطوعة، وغيرها. فعلى سبيل المثال، يقوم الخياط بقياس أبعاد جسم الشخص لتحضير الملابس بالمقاس المناسب، كما يقوم البائعون بوزن السلع كالأرز والقمح والسكر والفواكه والخضروات. وغالباً ما نعتقد خطأً أن القياس يقتصر على استخدام الأشرطة والموازين فقط، بينما في الواقع يمكن اعتبار تصنيف المتسابقين في مسابقة مناظرة نوعاً من القياس، وكذلك تقييم السلوك البشري. لذلك، يمكن تعريف القياس بأنه "تخصيص رقم من مجموعة أرقام لكل شخص أو جسم بناءً على قواعد محددة".

إن عملية القياس كما هو واضح تتضمن القيام بعملية قياس فعلية من أجل إسناد معنى كمي إلى خاصية ما؛ أي مثلاً: ما هو طول السبورة؟ تحديد ذلك يجب أن يتم عملياً. وعليه فالقياس هو عملية

إسناد أعداد إلى الأشياء أو الكميات أو الأحداث من أجل إعطائها معنى كمياً (Sivakumar, et al, 2019, p.6).

إذن القياس هو عملية وضع سلسلة من الإجراءات التي تُحوّل الملاحظات إلى بيانات كمية تعبر عن مقدار أو درجة أو مدى خاصية معينة، فالقياس نوعان، هما:

- **القياس الفيزيائي:** هو قياس يتعلق بالأشياء التي تمتلك قيمة مطلقة تبدأ دائماً من الصفر، ويتميز بالدقة والكمية الواضحة، مثل: قياس الطول والوزن.

- **القياس العقلي (النفسي/التربوي):** هو قياس نسبي لا يتوفر فيه صفر مطلق، ويُستخدم لقياس خصائص غير ملموسة، كالذكاء أو القدرات العقلية في الفصل الدراسي لتحديد أداء المتعلم، من خلال الحصول على مقاييس كمية من درجاته الفردية. فإذا حصل المتعلم على الدرجة 80 في الرياضيات مثلاً، فلا يمكن إعطاؤه تفسيراً واضحاً، ولا يمكن القول إنه قد نجح أو رسب. يتميز هذا القياس بكونه ذاتياً وقد يختلف باختلاف الأفراد. هذا النوع من القياس قد يكون نوعياً أو كمياً، وأدواته غير ثابتة دائماً وقد لا تناسب جميع الأشخاص بشكل متساوٍ (Tripura University, Directorate of Distance Education, 2016, p.4).

2- التطور التاريخي للقياس والتقويم التربوي والنفسي:

بدأ الاهتمام بقياس الفروق الفردية سنة 1776 عندما لاحظ العالم ببسال (Bessel) اختلاف زمن الاستجابة بين الأفراد فيما عُرف لاحقاً بـ "المعادلة الشخصية". وفي أواخر القرن التاسع عشر، أسس فونت (Wundt) أول معمل لعلم النفس سنة 1879، مؤسساً لمرحلة ركزت على قياس الحساسية للمثيرات وضبط ظروف الاختبار. برز تلميذه فرانسيس غالتون (Galton) بإسهامات كبيرة في القياس ودراسة الوراثة، وتطوير مقاييس إحصائية مع كارل بيرسون (Pearson)، مؤسسين المنهج الإحصائي في البحث النفسي. كما أدخل جيمس ماكين كاتل (Cattell) مصطلح "الاختبار العقلي" عام 1890، وأسهم في نشر حركة القياس النفسي. في أوائل القرن العشرين، تُرجمت اختبارات "بينيه - سيمون" (Binet-Simon) للذكاء إلى الإنجليزية على يد جودارد (Goddard)، ثم طُوّر لويس تيرمان (Lewis Terman) نسخة أمريكية موسعة عُرفت باختبار "ستانفورد-بينيه" (Stanford-Binet) عام 1916. لاحقاً، صمّم ديفيد وكسلر (Wechsler) عام 1939 اختبار "وكسلر-بلفيو" (Wechsler-Bellevue) لقياس ذكاء البالغين، جامعاً بين اختبارات لفظية وأدائية. شهدت الفترة بعد الحرب العالمية الثانية تطوراً كبيراً في القياس النفسي نتيجة نجاح الاختبارات في المجالين العسكري والبحثي، مما

عزز مكانتها وأسهم في انتشار مؤسسات متخصصة لتطوير وتقنين أدوات القياس النفسي والتربوي (معمرية، 2007، ص46).

3- القياس التربوي (Educational Measurement):

يُعدّ القياس عنصرًا محوريًا في أي مجال علمي، إذ تسعى جميع العلوم إلى تطوير أساليب دقيقة وموضوعية لقياس الظواهر التي تدرسها، بهدف فهمها وتفسيرها، والتنبؤ بالعلاقات التي تربط بين متغيراتها، والعمل على ضبطها والتحكم فيها. ويرتبط تقدم أي علم بمدى دقته في قياس السمات التي يتناولها. وقد أشار كارل بيرسون (Karl Pearson) إلى أن الأساس الذي تقوم عليه العلوم لا يكمن في طبيعة مادتها أو موضوعها، وإنما في الأسلوب العلمي أو المنهجية المتبعة في دراسة تلك السمات. وانطلاقًا من هذا المبدأ، يسعى علماء النفس والتربية إلى دراسة السمات النفسية والتربوية بمنهجية علمية تعتمد على الملاحظة الدقيقة واستخدام الوسائل المناسبة لجمع البيانات وتحليلها. ويستند هذا التوجه إلى مسلمة أساسية مفادها أنه رغم أوجه التشابه بين الأفراد، فإن بينهم فروقًا جوهرية ومتعددة. بعض هذه الفروق يمكن ملاحظته بسهولة، كالاختلاف في الطول أو الوزن أو لون البشرة، بينما تظل صفات أخرى غير ظاهرة، وتتطلب تدقيقًا ودراسة متعمقة، مع الاستعانة بأدوات قياس خاصة للكشف عنها وتمييزها. إن التعرف على الفروق في أي سمة يتطلب أولاً تحديد طبيعتها وفهمها، ثم تطوير أو اختيار وسائل مناسبة لقياسها، بهدف الوصول إلى تفسير علمي دقيق لها. وقد مرّت هذه العملية بمراحل تطور تاريخي يمكن تتبع أبرز معالمه لفهم الأسس التي يقوم عليها القياس والتقويم التربوي المعاصر.

يشير القياس التربوي إلى تطبيق مبادئ القياس ضمن المجال التعليمي في النظام التربوي. ويُقصد بالقياس التقييم الكمي لأداء الطلاب في اختبار معين، حيث يُستخدم لمقارنة مستويات الأداء بين الطلاب المختلفين، ولتحديد نقاط القوة والضعف لديهم. كما يساعد القياس في تصنيف الطلاب إلى مجموعات متجانسة، مما يسهل تخصيص التوجيه التربوي والمهني، واتخاذ الإجراءات العلاجية المناسبة للطلاب ذوي التحصيل المنخفض. يمثل القياس أداة هامة في يد علماء النفس التربوي لدراسة السلوك البشري، حيث يعتمد هؤلاء على استخدام اختبارات نفسية ذات صدق وثبات عالي لقياس مختلف السمات الفردية. ومن بين هذه الاختبارات نجد اختبارات الذكاء، والتحصيل، والميول، والقدرات، والاهتمامات، والشخصية، وغيرها، حيث تتعدد الأساليب المستخدمة في تطبيق هذه الاختبارات، كالملاحظة، والمقابلات، وقوائم المراجعة، ومقاييس التقدير، والامتحانات، وسجلات

الأداء التراكمي، بالإضافة إلى السجلات القصصية (Tripura University, Directorate of Distance Education, 2016, p.5).

4- التقويم التربوي (Educational Evaluation):

يُعرّف التقويم بأنه عملية توفير معلومات موضوعية تهدف إلى إصدار حكم، بينما يُعتبر القياس تعبيرًا كميًا عن هذه المعلومات بلغة رقمية أو حسابية، حيث يُستخدم مقدار معياري (بوحدّة قياس معينة) للتعبير عن الظاهرة المقاسة. بعبارة أخرى يُمكن اعتبار القياس عملية "تكميم" أو تعبير كمي عن صفات أو عوامل أو ظواهر تتعلق بموضوعات نوعية أو معنوية أو سلوكية، مما يستدعي إصدار حكم أو تقويم بشأنها. ففي مجال القياس التربوي أو النفسي لا يتم القياس بشكل مباشر كما هو الحال في العلوم الطبيعية، بل يعتمد على الاستدلال وهو ما ينطبق على معظم الظواهر السلوكية. يتطلب ذلك استخدام أساليب متطورة ودقيقة للقياس، تتسم بخصائص الصدق والثبات والقدرة على التمييز بين المستويات الافتراضية. تُقاس سمات السلوك كالذكاء والقدرات العقلية والميول والاتجاهات، أو أي سمة من سمات الشخصية، من خلال قياسات استدلالية غير مباشرة، تعتمد على افتراضات لوحدات سلوكية تُعبر عنها غالبًا من خلال اختبارات نفسية تتعلق بمظاهر السلوك. مع زيادة تعقيد عملية القياس، تزداد احتمالات الخطأ، مما يعني أن هناك عوامل قد تؤدي إلى حدوث أخطاء في القياس، والتي تكون أكثر شيوعًا في العلوم التربوية والنفسية مقارنة بالعلوم الطبيعية (علام، 2000، ص31).

فالتقويم إذن يضيف بُعد الحكم القيمي إلى التقدير أو التقييم (Assessment). فهو يهتم بتطبيق نتائجه، ويتضمن إصدار حكم حول مدى الفعالية أو الجدوى الاجتماعية أو الرغبة في منتج أو عملية أو تقدم، وذلك في ضوء أهداف أو قيم محددة ومتفق عليها بعناية. وغالبًا ما يتضمن التقويم توصيات لإجراءات بناءة. ومن ثم يُعدّ التقويم مقياسًا نوعيًا للوضع السائد يتطلب أدلة على الفعالية أو الملاءمة أو جودة البرنامج. وهو عملية تقدير لقيمة شيء أو عملية أو برنامج بهدف الوصول إلى قرارات ذات معنى بشأن ذلك الشيء أو العملية أو البرنامج (Sivakumar, et al, 2019, p.6).

5- أسس القياس التربوي (Principles of Educational Measurement):

أسس القياس التربوي وهي المبادئ والأسس النظرية والعملية التي يقوم عليها قياس الأداء أو السلوك التعليمي لدى الطلاب، والتي يمكن تلخيصها كما يلي:

أ- **الوضوح في الهدف:** يجب تحديد الهدف التعليمي أو السلوك المراد قياسه بدقة قبل إعداد أي أداة قياس.

ب- **الصدق (Validity):** ويشير إلى مدى قدرة أداة القياس على قياس ما صممت لقياسه فعليًا.

مثال: اختبار الرياضيات يجب أن يقيس المهارات الرياضية فقط وليس المهارات اللغوية.

ج- **الثبات (Reliability):** وهو مدى اتساق نتائج الاختبار عند إعادة تطبيقه على نفس المجموعة في ظروف مشابهة. فكلما زاد الثبات كانت نتائج القياس أكثر موثوقية.

د- **الموضوعية (Objectivity):** يجب أن تكون نتائج القياس غير متأثرة بآراء المصحح أو عوامل ذاتية.

هـ **القابلية للتطبيق العملي (Practicality / Usability):** أن تكون أداة القياس قابلة للتنفيذ بسهولة وبتكلفة ووقت مناسبين.

و- **القياس الكمي (Quantifiability):** ترجمة السلوكيات والمهارات إلى أرقام أو درجات لتسهيل التحليل والمقارنة.

ز- **الشمولية والتغطية (Comprehensiveness / Content Coverage):** أن يغطي الاختبار جميع الأهداف والمحتوى التعليمي المرتبط بالموضوع.

ح- **العدالة (Fairness):** أن تكون أداة القياس متساوية لجميع الطلاب بغض النظر عن خلفياتهم الاجتماعية أو الثقافية.

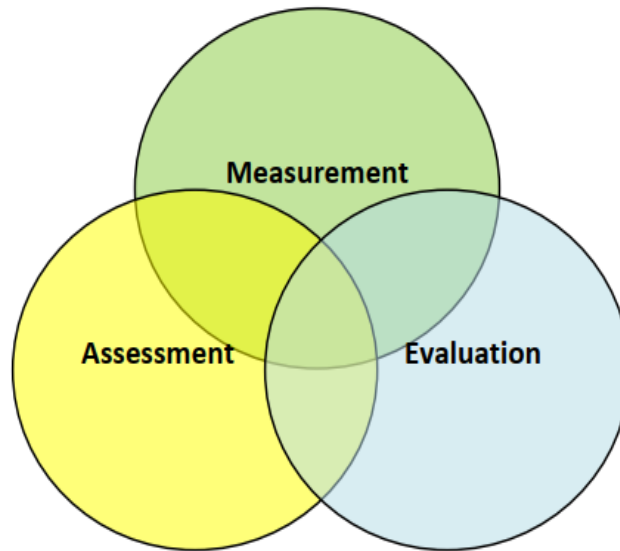
تجسد الأسس السابقة الدعائم النظرية والعملية التي يقوم عليها القياس التربوي، فهي تمثل الإطار الذي يضمن دقة وموثوقية النتائج وعدالتها وموضوعيتها. وبالرغم من أننا قدمنا هنا لمحة موجزة عن هذه الأسس، فإننا سنتناول كل أساس منها بشيء من التفصيل والتحليل في الفصول القادمة من هذا الكتاب، مع تقديم أمثلة تطبيقية تساعد القارئ على فهم كيفية توظيفها عمليًا في إعداد أدوات القياس والتقويم التربوي.

6- دور القياس والتقويم في عملية التقييم:

إن مفهوم التقييم (Assessment) يعد من المفاهيم الواسعة والشاملة في العملية التربوية، إذ لا يقتصر على مجرد إصدار أحكام عامة، بل يتضمن في جوهره عمليات مترابطة من القياس والتقويم (Measurement and Evaluation). فالتقييم يعتمد على إجراءات قياسية دقيقة تهدف إلى جمع بيانات كمية وموضوعية حول أداء المتعلم أو فعالية العملية التعليمية. غير أن دور التقييم لا يتوقف عند حدود الأرقام والمؤشرات، بل يتجاوزها إلى مرحلة أعمق تتمثل في التقويم (Evaluation)، حيث تُوظف نتائج التقييم لاتخاذ القرارات التربوية السليمة، وصياغة الأحكام الموضوعية، ورسم

استراتيجيات التطوير والتحسين المستمر للعملية التعليمية. والشكل الآتي يوضح العلاقة والتداخل بين القياس والتقويم والتقييم:

شكل 01 يوضح العلاقة بين القياس والتقويم والتقييم



(Magno, et al, 2010, p.04)

7- طبيعة القياس التربوي:

يُعد القياس عنصراً أساسياً في عملية التقييم، إذ يتميز بخصائص الكم والتجريد والتحليل، وهي من السمات الجوهرية للبحث العلمي. فكثير من نتائج التقييم تأخذ صورة قيم كمية تسمح بإجراء تحليلات أعمق وأكثر دقة. إن الحصول على الأدلة المتعلقة بمختلف الظواهر في العالم يعتمد في جوهره على القياس؛ فالتعبيرات لا تُقبل على أنها صحيحة أو خاطئة إلا إذا كان الحدث قابلاً للملاحظة المباشرة. وفي المجال التربوي لا يمكن الحكم على طالب بأنه "ذكي جداً" إلا إذا وُجدت دلائل واضحة يمكن ملاحظتها تثبت هذا الحكم. ومن ثم فإن من يتولى مهمة تحديد ما إذا كان الطالب "موهوباً بدرجة عالية" ينبغي عليه جمع بيانات دقيقة وأدلة موضوعية تدعم هذا الوصف. وعندما يُظهر الأفراد سمات معينة مثل "الذكاء" من خلال أحكام صادرة عن الآخرين، أو من خلال درجات مرتفعة في الاختبارات، أو أداء متميز في مهام معرفية، أو معدلات دراسية عالية، فإن ذلك يدل على أن عملية القياس قد تمت ونُفذت بشكل دقيق ومنهجي، واستوفت شروط البحث العلمي.

ومن المهم إدراك أن القياس لا يطل الأشياء في ذاتها، وإنما يطل خصائصها أو سماتها، وهذه الخصائص أو السمات يُطلق عليها "المتغيرات"، ومن أمثلة المتغيرات التي يتم تناولها في المجال التربوي: الذكاء، التحصيل، الاستعدادات، الاهتمامات، الاتجاهات، الشخصية، وغيرها.

وقد عرّف نونالي (1970) Nunnally القياس بأنه "مجموعة من القواعد التي تُسند بموجبها أرقام إلى الأشياء بطريقة تعكس كميات من السمات أو الخصائص". وبذلك فإن القياس يهدف إلى تحويل السمات إلى مقادير كمية، وهو ما يتيح العديد من المزايا كالمقارنة الموضوعية، وإجراء التحليلات الإحصائية، والوصول إلى تفسيرات أكثر دقة للظواهر المدروسة. إن تكميم الخصائص أو السمات يحقق عدة مزايا أساسية، من أبرزها (Magno, et al, 2010, p.04):

أ- **الكمية تحدد مقدار السمة أو الخاصية الموجودة:** فعلى سبيل المثال إذا حصل طالب على الرتبة المئينية 10 على اختبار في مادة الرياضيات، فإن ذلك يعني أن إنجازه أقل من غالبية زملائه. أما إذا حصل طالب آخر على الدرجة الكاملة في اختبار قصير حول قوانين نيوتن للحركة، فإن ذلك يشير إلى أنه قد تذكر وأتقن المعلومات المرتبطة بالموضوع بشكل كامل.

ب- **الكمية تسهّل الحصول على معلومات دقيقة:** فإذا حصل طالب على درجة معيارية قدرها 2- في اختبار مقنن، حيث تتراوح الدرجات المعيارية بين 3- و 3+ ويكون المتوسط هو الصفر، فهذا يدل على أن أدائه أقل من المتوسط. وبالمثل إذا حقق الطالب درجة معيارية تساعية (stannine) قدرها 8، حيث يتراوح المدى بين 1 و 9 ويكون 5 هو المتوسط، فهذا يشير إلى أن الطالب أعلى من المتوسط ويتميز بقدرات عالية في السمة التي يقيسها الاختبار.

ج- **الكمية تتيح المقارنة الموضوعية بين المجموعات:** فلو اختبر كلٌّ من الذكور والإناث في قدرتهم على حل مسائل الرياضيات باستخدام نفس الاختبار، وكانت النتيجة أن متوسط درجات الذكور هو 92.3 مقابل 81.4 للإناث، فإن ذلك يسمح لنا القول وطبعاً بعد اختبار دلالة الفروق بينهم إن الذكور أدّأهم أفضل من الإناث في هذا الاختبار.

د- **الكمية تمكّن من تصنيف المجموعات:** إن التقسيم الشائع للأقسام أو الصفوف غالباً ما يعتمد على المعدل العام للطلاب في السنة الدراسية السابقة، خصوصاً في حالة وجود أقسام مميزة ضمن المستوى. وفي هذه العملية تُرتب درجات الطلاب من الأعلى إلى الأدنى، وتُحدد حدود فاصلة وفقاً لعدد المقاعد المتاحة في كل قسم.

هـ- **الكمية تجعل البيانات قابلة لمزيد من التحليل:** عندما تُحوّل البيانات إلى مقادير كمية، يستطيع المعلمون، والمستشارون التربويون، والباحثون، والإداريون، وغيرهم من المهتمين استخراج نتائج

متنوعة لتلخيص البيانات واستنتاج دلالاتها. كما يمكن عرض هذه النتائج في جداول ورسوم بيانية ومخططات تُظهر المتوسطات والنسب المئوية. وإضافة إلى ذلك يمكن توظيف الإحصاء الاستدلالي لمقارنة المجموعات، أو وضع معايير مرجعية، أو تقييم فاعلية البرامج التعليمية (Magno, et al, 2010, p.05).

إن عملية القياس في العلوم الطبيعية كالفيزياء والكيمياء والأحياء تشبه إلى حد كبير ما يتم في ميدان التربية والعلوم الاجتماعية، حيث يعتمد كلاهما على أدوات أو وسائل للحصول على نتائج القياس. غير أن الاختلاف الجوهرى يكمن في طبيعة المتغيرات المقاسة؛ ففي العلوم الطبيعية تكون عملية القياس أكثر دقة وموضوعية بسبب إمكانية ملاحظة البيانات الفيزيائية بشكل مباشر ولموس. أما في ميادين التربية وعلم النفس والعلوم السلوكية، فإن البيانات أكثر عرضة لأخطاء القياس وتقلبات أكبر نتيجة الفروق الفردية وصعوبة التحكم في الظروف المحيطة بالقياس. ومع ذلك توجد في هذه المجالات إجراءات إحصائية خاصة للكشف عن أخطاء القياس، كحساب الانحراف المعياري والخطأ المعياري للقياس والتباين.

كما أن القياس يعزز موضوعية الملاحظة، حيث يساهم في الحد من الفروق الكبيرة في النتائج شريطة أن تتسم الظروف بالثبات والسيطرة على الفروق الفردية. وهذا يعني أنه إذا قاس شخصان متغيراً واحداً وفق نفس الشروط، فإن نتائجهما يجب أن تكون متسقة. وقد تظهر بعض الفروق الطفيفة، خصوصاً عند قياس متغيرات نفسية، لكن التكرار والحصول على نتائج متقاربة يعكس موضوعية الإجراء.

تتطوي عملية القياس أيضاً على قدر من التجريد؛ فقبل قياس أي متغير باستخدام أداة معينة، لا بد من دراسته بدقة وتحديد طبيعته من خلال تعريفه مفهوماً وإجراءً. فالتعريف المفهومي يوضح الأساس النظري أو الإطار المرجعي للمتغير، بينما يبرز التعريف الإجرائي الخطوات والإجراءات العملية لقياسه. وغالباً ما تكون بعض المتغيرات معقدة وتتألف من عدة عوامل أو مكونات والتي يُطلق عليها المتغيرات الكامنة (Latent Variables). ومن أمثلتها التحصيل الدراسي الذي يتضمن جوانب متعددة كالرياضيات، العلوم، اللغة الإنجليزية، والدراسات الاجتماعية. بعد تحديد هذه المكونات يمكن اختيار الأداة المناسبة التي تقيسها بدقة، كاختبار (GRE) الذي يقيس جوانب متعددة كالقدرات اللفظية والكمية والتحليلية وغيرها.

ويُطلق على المتغير الذي يتكون من عدة عوامل متغير متعدد الأبعاد (Multidimensional Variable)، والذي يحتاج عادة إلى أداة قياس تحتوي على عدة اختبارات فرعية، بينما يُسمى المتغير

البسيط الذي يقيس سمة واحدة فقط متغيراً أحادي البعد (Unidimensional Variable). ومن أمثلة أدوات القياس أحادية البعد: مقياس روزنبرغ لتقدير الذات ومقياس القلق (PSWQ)، في حين أن مقاييس القدرات العامة والاختبارات الشخصية مثل اختبار العوامل الستة عشر (16PF) تُعد أمثلة على أدوات قياس متعددة الأبعاد (Magno, et al, 2010, p.06).

8- طبيعة التقويم التربوي:

يُستند في عملية التقويم إلى نتائج القياس والتقييم معاً، إذ إن القرارات المتعلقة بانتقال الطالب أو بقاءه في مرحلته الدراسية تعتمد على درجاته وسلوكه، كما أن تقويم برنامج علاجي في الرياضيات يستلزم النظر إلى أكثر من جانب؛ مثل مدى تحسن مستوى الطلاب، وأداء المعلمين، والتغير في اتجاهات المتعلمين. ووفقاً لما طرحه سكريفن (1967) Scriven يُنظر إلى التقويم بوصفه عملية إصدار حكم حول قيمة أو جدوى البرامج والسياسات والأفراد والأحداث، بما يتيح لأصحاب المصلحة الاستناد إلى هذه الأحكام في اتخاذ القرارات المناسبة.

ويتطلب التقويم السليم وجود معايير يمكن من خلالها الحكم على الجودة، سواء كانت معايير نسبية أو مطلقة (Fitzpatrick, Sanders, and Worthen, 2004). وتُستخدم هذه المعايير لتقدير قيمة وفاعلية ما يُخضع للتقويم، فمثلاً يمكن الحكم على مستوى إحدى الجامعات بمقارنتها بمستوى الجامعات المرموقة عالمياً أو وفق متطلبات هيئات الاعتماد الأكاديمي كهيئة (PAASCU).

للتقويم خصائص أساسية، من أبرزها: أن غايته تتمثل في خدمة أصحاب المصلحة كالطلاب والمعلمين والإداريين، وأن نتائجه تأخذ شكل أحكام حول جدوى البرامج أو إمكانية استمرارها أو قبول الطلاب أو رفضهم، وأن دلالاته تنعكس في وصف البرامج والسياسات والمؤسسات والأفراد. أما الأجندة فهي غالباً تُحدد من قبل أصحاب المصلحة أنفسهم، في حين يكون التعميم مرتبطاً بالسياق الخاص الذي يُجرى فيه التقويم. وتقوم هذه العملية على مجموعة من المعايير تشمل الدقة، والمنفعة، وقابلية التطبيق، والملاءمة. وقد جرى اعتماد هذه المعايير الأربعة رسمياً سنة 1975 من قبل اللجنة المشتركة لمعايير التقويم التربوي بقيادة Daniel Stufflebeam in 1975، حيث عُرفت باسم: "معايير تقويم البرامج والمشروعات والمواد التعليمية. والجدول (1) يقدّم وصفاً للمعايير الأربعة (Magno, et al, 2010, p.07):

الجدول 01 يوضح المعايير الخاصة بتقويم البرامج والمشروعات والمواد التعليمية

المعيار	الملخص	المكونات
المنفعة (Utility)	يهدف إلى ضمان أن يخدم التقويم احتياجات وتقديم المعلومات للمستفيدين منه.	تحديد أصحاب المصلحة، مصداقية المقيم، نطاق المعلومات واختيارها، تحديد القيم، وضوح التقرير، توقيت التقرير ونشره، أثر التقويم.
الجدوى (Feasibility)	يهدف إلى ضمان أن يكون التقويم واقعياً، حكيمًا، دبلوماسيًا، واقتصاديًا.	الإجراءات العملية، الجدوى السياسية، الفعالية من حيث التكلفة.
الملاءمة (Propriety)	يهدف إلى ضمان أن يتم تنفيذ التقويم بطريقة قانونية، وأخلاقية، وبمراعاة رفاهية المشاركين في التقويم وكذلك المتأثرين بنتائجه.	التوجه الخدمي، الاتفاقيات الرسمية، حقوق الأفراد، التفاعل البشري، التقييم الكامل والعادل، الكشف عن النتائج، تضارب المصالح، المسؤولية المالية.
الدقة (Accuracy)	يهدف إلى ضمان أن يكشف التقويم وينقل معلومات تقنية كافية حول الخصائص التي تحدد قيمة أو جدوى البرنامج محل التقويم.	توثيق البرنامج، تحليل المحتوى، الأهداف والإجراءات الموصوفة، مصادر المعلومات الموثوقة، المعلومات الصالحة، المعلومات الموثوقة، المعلومات المنهجية، تحليل المعلومات الكمية، تحليل المعلومات النوعية، الاستنتاجات المبررة، التقارير الموضوعية، ما وراء التقويم (Metaevaluation).

(Magno, et al, 2010, p.08)

9- أشكال أو أنواع القياس والتقويم التربوي:

يُعدّ القياس والتقويم التربوي أداةً أساسية لتحسين جودة عمليتي التعليم والتعلم، حيث يُمكن جميع الأطراف المعنية – من معلمين، وطلبة، وأولياء أمور، وإدارة مدرسية، ومشرفين تربويين، وصناع قرار، بل والمجتمع بأسره – من الحصول على بيانات دقيقة ومؤشرات موثوقة حول مدى تقدّم العملية التعليمية وتحقيقها للأهداف المرجوة. ومن خلال هذه الأدوار يسهم التقويم في تشخيص الواقع التربوي، وتوجيه مساره وتطوير برامجه بما يتوافق مع احتياجات المتعلمين ومقتضيات المجتمع. وتتنوّع أشكال القياس والتقويم التربوي لتشمل ما يلي:

9-1- التقويم التشخيصي (القبلي/التمهيدي) (Diagnostic Evaluation):

يُعدّ التقويم التشخيصي أحد أهم أنواع التقويم التي تسبق العملية التعليمية، حيث يُستخدم قبل الشروع في تنفيذ البرنامج أو الدرس الجديد. يهدف هذا النوع من التقويم إلى الكشف عن مستوى

الطلبة الحالي من حيث المعارف والمهارات والخبرات السابقة ذات الصلة بالموضوع المزمع دراسته، وذلك بغرض تحديد نقاط القوة التي يمكن البناء عليها ونقاط الضعف التي تستدعي المعالجة. ويتخذ هذا التقويم أشكالاً متعددة، منها الأسئلة الشفوية التي يطرحها المعلم في بداية الدرس، أو الاختبارات القصيرة القبلية، أو مناقشات سريعة لاستكشاف خلفيات الطلبة العلمية ومدى استعدادهم للتعلم الجديد. وتتميز أدواته بالبساطة والعمومية والتركيز على المفاهيم الأساسية دون الدخول في تفاصيل دقيقة. ومن أبرز وظائف هذا النوع من التقويم:

- أ- تحديد الفروق الفردية بين الطلبة لتخطيط التعليم بما يتناسب مع حاجاتهم.
- ب- تشخيص صعوبات التعلم المحتملة قبل البدء بالبرنامج التعليمي.
- ج- مساعدة المعلم في اختيار الاستراتيجيات والأنشطة التعليمية الملائمة.
- د- توفير قاعدة مرجعية للمقارنة مع ما يحققه الطالب بعد انتهاء العملية التعليمية، بما يبرز قيمة التقدم المحقق.

وبذلك، يمثل التقويم التشخيصي أداة أساسية للمعلم في توجيه العملية التعليمية نحو مزيد من الفاعلية، لأنه يُمكنه من الانطلاق من مستوى الطلبة الفعلي لا الافتراضي (اليقوبي، 2013، ص.62).

9-2- التقويم البنائي أو التكويني (Formative Evaluation):

يُعد هذا النوع من التقويم من أهم الأدوات التي يعتمد عليها المعلم لمتابعة تعلم الطلبة أثناء سير العملية التعليمية، إذ يهدف إلى الكشف المستمر عن مدى إتقان التلميذ للمفاهيم والمعارف والمهارات التي يتلقاها. كما يساعد على تشخيص مواطن القوة والضعف لدى المتعلمين بشكل دوري، ومن ثم يتيح للمعلم تعديل طرائق التدريس أو إعادة شرح ما لم يتم استيعابه بصورة كافية.

يمتد دور التقويم التكويني منذ بداية الدرس وحتى مراحلها الأخيرة قبل التقويم الختامي، وهو بذلك يمثل حلقة تغذية راجعة دائمة تمكّن كلاً من المعلم والطالب من الوقوف على مستوى التقدم الفعلي. فقد يقوم المعلم باستخدام أسئلة شفوية أو تحريرية، أو تكليف الطلبة بواجبات عملية كإعداد تقرير، أو إجراء تجربة، أو عرض شفوي، أو أداء مهمة منزلية، وذلك بهدف التأكد من مدى تحقق الأهداف التربوية المرحلية. ومن أبرز وظائفه:

- أ- متابعة تطور مستوى التلاميذ أولاً بأول.
- ب- تمكين المعلم من تعديل استراتيجيات وأساليب التعليم بما يلائم قدرات التلاميذ.
- ج- تعزيز دافعية التعلم لدى الطلبة من خلال إشراكهم المستمر في عملية التقييم.

د- توفير تغذية راجعة فورية لكل من الطالب والمعلم تساعد على تحسين الأداء.

وبذلك فإن التقويم البنائي لا يُنظر إليه كأداة للحكم النهائي، بل كوسيلة لتوجيه العملية التعليمية نحو تحقيق أهدافها بأكبر قدر من الكفاءة والفاعلية (Sivakumar, et al, 2019, p.16).

9-2-1- التقويم التشخيصي (Diagnostic Evaluation):

يُعدّ التقويم التشخيصي أداة أساسية للكشف عن مواطن الضعف والصعوبات التي قد تعيق تعلم الطالب، والعمل على معالجتها بشكل فوري لضمان استمرارية العملية التعليمية بفاعلية. ويُمارس هذا النوع من التقويم منذ المراحل الأولى للتعلم ويستمر في مراحله اللاحقة كلما دعت الحاجة، حيث يساعد المعلم على تحديد طبيعة المشكلة بدقة، سواء كانت مرتبطة بالمعارف السابقة، أو المهارات الأساسية، أو حتى بالعوامل النفسية والاجتماعية المؤثرة في تعلم الطالب.

ويتميز التقويم التشخيصي بأن أسئلته وأنشطته مركزة ودقيقة، تُبنى خصيصاً للكشف عن طبيعة الصعوبة، وبالتالي فهي تختلف عن الأسئلة العامة أو الشمولية، إذ تركز فقط على النقطة التي تحتاج إلى علاج. ويمكن أن تكون أدوات هذا التقويم شفوية عبر المناقشة، أو تحريرية من خلال اختبارات قصيرة، أو عملية عبر مواقف تطبيقية وأنشطة أداء (اليقوبي، 2013، ص.62).

كما يمكن أن يُنجز هذا التقويم داخل القسم أثناء الحصة الدراسية، أو خارجه من خلال أنشطة إضافية، ويُوظف المعلم نتائجه مباشرة لوضع خطة علاجية سريعة تعيد الطالب إلى مسار التعلم الصحيح. ومن ثم فإن التقويم التشخيصي ليس مجرد وسيلة لاكتشاف الصعوبات، بل هو في جوهره عملية علاجية متكاملة تسعى لتحقيق مبدأ التعلم المستمر والمتوازن (Kumar, et al, 2016, p.16).

9-3- التقويم النهائي أو الختامي (Summative Evaluation):

يُقصد بالتقويم النهائي أو الختامي ذلك النوع من التقويم الذي يُجرى في نهاية المقرر الدراسي أو عند انتهاء عملية تعليمية كاملة، ويهدف أساساً إلى الوقوف على مدى تحقق الأهداف الكلية للموقف التعليمي. فهو وسيلة للحكم على فاعلية العملية التعليمية في مجملها، وتشخيص جوانب القوة والضعف فيها، واتخاذ قرارات تتعلق بتطويرها أو تعديلها في المستقبل.

كما يساعد هذا النوع من التقويم على إصدار أحكام تتعلق بمستوى تحصيل الطالب بصورة شاملة ونهائية، وتحديد مدى أهليته للانتقال إلى مستوى أعلى من التعليم أو البقاء في نفس المستوى، مما يجعله أداة أساسية في اتخاذ القرارات التربوية الكبرى ذات الطابع المصيري.

وتتميز أدوات وأسئلة التقويم الختامي بأنها أكثر عمومية وشمولية من أسئلة التقويم التكويني، ولكنها أكثر تفصيلاً وتحديداً من أسئلة التقويم التمهيدي. وقد تأخذ هذه الأدوات أشكالاً متعددة، مثل:

- أ- الأسئلة التحريرية (الامتحانات الكتابية بمختلف أنماطها).
- ب- الأسئلة الشفوية لمناقشة مدى عمق الفهم والتحليل.
- ج- الأسئلة العملية/الأدائية لتقويم المهارات التطبيقية.
- ويُجرى التقويم النهائي عادةً في أوقات معلومة ومحددة مسبقاً، مثل مواعيد الامتحانات الفصلية أو الامتحانات الوطنية، حيث تُعلن نتائجه رسمياً وتُبنى عليها قرارات إدارية وتعليمية مهمة. ويمثل الشكل التالي أهم العناصر الأساسية التي يتضمنها التقويم النهائي/الختامي:
- أ- قياس مدى تحقق الأهداف الكلية.
- ب- الحكم على مستوى تحصيل الطالب بشكل شامل.
- ج- إصدار قرارات تتعلق بالترقية أو النجاح.
- د- تقييم مدى فاعلية المنهج واستراتيجيات التدريس.
- هـ- تقديم تغذية راجعة شمولية للمؤسسة التعليمية (Sivakumar, et al, 2019, p.16) (اليقوي، 2013، ص.62).

الجدول 02 يوضح الفرق بين التقويم التكويني والتقويم الختامي

التقويم التكويني	التقويم الختامي
1- يُجرى أثناء عملية التدريس والتعلم، سواء خلال الحصة، أو أثناء الفصل الدراسي، أو خلال الجلسة التعليمية.	1- يُجرى في نهاية عملية التدريس والتعلم، مثل نهاية الحصة، أو نهاية الفصل الدراسي، أو عند اختتام البرنامج.
2- يقيس مستوى الإنجاز (التحصيل) في مهام صغيرة يتم تعلمها خلال فترة قصيرة.	2- يقيس مستوى الإنجاز (التحصيل) في مهام كبرى يتم تعلمها على مدى فترة زمنية أطول.
3- يتم بشكل متكرر ومنتظم أثناء الدرس أو المقرر أو الجلسة.	3- يتم في ختام المقرر أو البرنامج الدراسي.
4- نتائجه تعطي تعميمات محدودة.	4- نتائجه تعطي تعميمات واسعة وشاملة.
5- يغطي مجاًلاً محدوداً من المحتوى والقدرات.	5- يغطي مجاًلاً واسعاً من المحتوى والقدرات.

(Kumar, et al, 2016, p.16)

- العوامل الواجب مراعاتها لضمان تقويم فعّال:

- أ- **طريقة المعاينة:** ينبغي اختيار أسلوب معاينة مناسب وملائم لأهداف التقويم.
- ب- **تنظيم عملية التقويم:** يجب أن تُدار عملية التقويم بصورة منظمة، مع توفير:

- معالجة مناسبة للبيانات؛

- بيئة وجو ملائمين؛

- مراعاة النتائج المتوقعة وغير المتوقعة وما يترتب عليها من آثار.

ج- موضوعية أدوات القياس: وتشمل ضمان ما يلي:

- قابلية الأداة للتحقيق والتطبيق؛

- معالجة الجوانب الأخلاقية المرتبطة بالتقويم؛

- ثبات الاختبار من حيث دقة البيانات واستقرارها وإمكانية تكرارها؛

- صدق الأداة بحيث تقيس ما وضعت لقياسه وتعكس الخصائص المستهدفة.

د- مبررات استخدام أدوات التقويم: لا بد من وجود أسس واضحة ومبررة لاختيار الأداة.

هـ عدالة الأداة: ينبغي أن تكون الفروق في أداء الطلاب مرتبطة بمحتوى الاختبار نفسه وليس بطريقة التطبيق.

و- الكفاءة والاقتصاد: يجب اعتماد النمط الأكثر توفيراً للوقت والجهد والموارد مع تحقيق الكفاءة.

ز- إعداد المعلمين: من الضروري أن يتلقى المعلمون التدريب الكافي وأن يكونوا مؤهلين لتدريس

المواد التي يتولونها (Sivakumar, et al, 2019, p.09).

10- خصائص التقويم التربوي:

أ- يشمل التقويم تقييم جميع نواتج عملية التعليم والتعلم من حيث التغيرات السلوكية الشاملة، حيث يتجاوز الأهداف المعرفية ليغطي الأهداف المرتبطة بالمهارات، والتطبيقات، والاهتمامات، والاتجاهات، والتقدير. وبذلك أصبح مجال وأفق اختبار الأهداف التعليمية أوسع بكثير مع اعتماد هذا المفهوم.

ب- يتضمن التقويم إصدار أحكام واتخاذ قرارات تتعلق بتقدم المتعلم، والصعوبات التي يواجهها، والبحث عن الإجراءات التصحيحية المناسبة لتحسين تعلمه.

ج- يتطلب التقويم تفسير البيانات بطريقة دقيقة وحذرة.

د- يُعد التقويم عملية مستمرة، فهو لا يقتصر على صف دراسي معين أو مرحلة تعليمية محددة، بل يرافق المتعلم منذ التحاقه بالصف عبر تقويم التوزيع (التمهيدي)، ويمتد خلال مراحل تعلمه من وحدة إلى أخرى عبر التقويم البنائي والتشخيصي، وصولاً إلى التقويم الختامي عند نهاية مرحلة دراسية معينة.

هـ- يتسم التقويم بالشمولية، إذ لا يقتصر على الجوانب الأكاديمية للمتعلم، بل يغطي جميع جوانب نموه، بما في ذلك الجوانب المعرفية وغير المعرفية.

و- يعتمد التقويم على جميع الوسائل المتاحة لجمع المعلومات حول تعلم المتعلم، كالاختبارات، الملاحظة، المقابلة، مقاييس التقدير، قوائم الفحص، وأحكام القيمة، وذلك للحصول على معلومات وافية وموثوقة (Sivakumar, et al, 2019, p.10).

الفصل الثاني

المفاهيم الأساسية للاختبارات النفسية والتربوية: التصنيف، التقنين، المعايير، ومستويات القياس

الأهداف (Objectives):

يهدف هذا الفصل إلى تمكين الطالب من الإلمام بالجوانب التطبيقية والتقنية للقياس والتقويم، حيث يسعى إلى تعريفه بتصنيفات الاختبارات والمقاييس النفسية والتربوية، وفهم أسس تقنياتها وضوابط تطبيقها، إلى جانب توضيح مفهوم المعايير وأهميتها وعلاقتها بعينة التقنين. كما يهدف الفصل إلى تمكين الطالب من التمييز بين أنواع المعايير والدرجات المحولة، والتعرف على كيفية بناء الصفحة النفسية أو التربوية. كذلك يتناول الفصل تصنيف الاختبارات التحصيلية، ويعرض لمستويات القياس التربوي والنفسي وخصائصها، مع بيان علاقتها بالإحصاء، بما يساعد الطالب على بناء فهم شامل ودقيق لكيفية استخدام الاختبارات والمعايير في الممارسات التربوية والنفسية.

المحتوى (Contents):

- 1- تصنيف الاختبارات والمقاييس النفسية والتربوية.
- 2- تقنين الاختبارات التربوية والنفسية.
- 3- معنى المعايير وأهميتها.
- 4- المعايير وعينة التقنين.
- 5- أنواع المعايير.
- 6- أنواع المعايير أو الدرجات المحولة (Norms).
- 7- المعايير والصفحة النفسية أو التربوية (Psychological or Educational Profile).
- 8- تصنيف الاختبارات التحصيلية.
- 9- مستويات القياس التربوي والنفسي.
- 10- خصائص مستويات القياس.
- 11- مستويات القياس والإحصاء.

تمهيد:

يُعدّ الجانب التطبيقي للقياس والتقويم من أهم الركائز التي يقوم عليها البحث التربوي والنفسي، إذ لا يقتصر الأمر على فهم المفاهيم النظرية، بل يتطلب معرفة عملية بالاختبارات والمقاييس وطرق تصنيفها وتقنينها، بما يضمن دقة النتائج وموضوعيتها. كما تمثل المعايير بأنواعها المختلفة أداة أساسية لفهم درجات الأفراد ومقارنتها، سواء من خلال الدرجات الخام أو المحوّلة، أو من خلال بناء الصفحة النفسية أو التربوية التي تقدم صورة شاملة عن خصائص الفرد. ومن جهة أخرى يُعتبر التعرف على مستويات القياس وخصائصها، وكذلك علاقتها بالإحصاء، خطوة جوهرية لفهم كيفية التعامل مع البيانات وتحليلها بشكل علمي. ومن هنا تأتي أهمية هذا الفصل إذ يزود الطالب بالمعارف والمهارات اللازمة للتعامل مع الاختبارات والمعايير والمستويات الإحصائية في مجال القياس النفسي والتربوي.

1- تصنيف الاختبارات والمقاييس النفسية والتربوية:

نظرًا لما تنتم به الظواهر التربوية والنفسية من تعقيد وتشابك، وتعدد التطبيقات الميدانية للقياس في مجالات متنوعة، مثل: انتقاء الأفراد لشغل الوظائف والمهام المختلفة، وتسكينهم وتصنيفهم، إضافةً إلى التوجيه والإرشاد النفسي، وتقويم الأفراد والبرامج والمؤسسات وغيرها، فقد تنوعت أدوات القياس والتقويم بما يتوافق مع طبيعة الظاهرة المراد قياسها وما تتضمنه من متغيرات. وبفضل هذا التنوع أصبح في متناول الباحثين والممارسين عدد كبير من أدوات القياس التي يمكنهم اختيار الأنسب منها للفئة المستهدفة وللسمّة المطلوب قياسها. فنجد مثلاً أدوات قياس الذكاء والاستعدادات والتحصيل تختلف عن أدوات قياس القلق أو الاكتئاب، أو قياس الميول والاتجاهات، أو الجوانب النفسية الحركية. ومن المهم الإشارة إلى أنه لا يوجد تصنيف واحد جامع لأدوات القياس، بل يمكن تصنيفها وفق عدة أبعاد. وفيما يلي عرض لأحد التصنيفات الممكنة (علام، 2000، ص44-48).

أ- أدوات قياس الجوانب المعرفية مقابل أدوات قياس الجوانب الوجدانية:

تتطلب الجوانب المعرفية للشخصية، كالقدرات العقلية والتحصيل الدراسي، أدوات قياس يسعى فيها الفرد المختبر إلى بذل أقصى جهده للحصول على درجة تعكس بدقة مستوى قدرته الفعلية. فعلى سبيل المثال، يهدف اختبار الاستعداد اللغوي إلى قياس قدرة الفرد من خلال الإجابة عن أكبر عدد ممكن من المفردات – غالبًا بأسلوب الاختيار من متعدد – خلال زمن محدد. أما الجوانب الوجدانية والسمات المزاجية للشخصية، فتركز على قياس السلوك المعتاد للفرد في مواقف انفعالية أو اجتماعية

معينة، أي السلوك الذي يميزه عن غيره. وتشمل أدوات القياس هنا عبارات مفتوحة الاستجابة، يجب عنها الفرد بما يتفق مع خصائصه وسلوكه اليومي في مواقف مشابهة. ومن أمثلة هذه العبارات: "أحب تنسيق الزهور"، "أفضل الاختلاط في الجامعة"، "أعتقد أن الناس لا تحبني"، أو "تنتابني وساوس كثيرة". ونظرًا لعدم وجود إجابة صحيحة أو خاطئة لهذه العبارات، بل اختلافها تبعًا لمشاعر الفرد أو آرائه أو معتقداته، فإن قياس الجوانب الوجدانية يعد أكثر تعقيدًا من قياس الجوانب المعرفية. ولهذا، تعتمد أدواته على أساليب متعددة مثل الملاحظة الموضوعية، والتقارير الذاتية، والاستبيانات، وغيرها من أساليب التقييم الشخصي (Personal Assessment).

ب- أدوات القياس الجماعية مقابل أدوات القياس الفردية:

تُستخدم أدوات القياس الجماعية لتطبيق الاختبارات على عدد كبير من الأفراد في الوقت نفسه، مثل اختبارات الذكاء الجماعية، أو اختبارات التحصيل، أو اختبارات القبول بالكليات، حيث يمكن تطبيقها على فصل دراسي كامل أو على مجموعة من المتقدمين للخدمة العسكرية دفعة واحدة. في المقابل، صُممت أدوات القياس الفردية لتطبيقها على شخص واحد في كل مرة، وتُستخدم غالبًا في مواقف الإرشاد والتوجيه التربوي والمهني، أو عند قياس ذكاء الأطفال. يتم إجراء هذه الاختبارات بواسطة أخصائي متمرس ومدرب، يقوم بملاحظة أداء المستجيب وتوجيهه أثناء الاختبار، والتفاعل معه لفهم استجاباته بدقة، مما يوفر فرصًا غنية لجمع الملاحظات الإكلينيكية. وتشمل هذه الأدوات اختبارات فردية متعددة لقياس الذكاء والشخصية. ورغم أهميتها في اتخاذ القرارات الفردية الحاسمة كتشخيص الاضطرابات النفسية أو تحديد الفصول الدراسية المناسبة، فإن أدوات القياس الفردية تتطلب وقتًا أطول للتطبيق، إضافة إلى الحاجة لغرفة مجهزة وظروف خاصة، مما يجعل استخدامها مقصورًا على الحالات التي تبرر هذا الجهد الإضافي.

ج- أدوات القياس كعينة مقابل أدوات القياس كمؤشر:

في بعض أدوات القياس وخاصة اختبارات التحصيل يتم بناء الاختبار على اختيار عشوائي لعينة من المفردات تمثل جميع المحتوى أو الأهداف التعليمية المراد قياسها. من خلال هذه العينة يمكن الاستدلال على مدى تحقق الهدف لدى الفرد. فعلى سبيل المثال إذا أراد المعلم قياس قدرة الطالب على التعرف على الرموز الكيميائية، فإنه يختار عشوائيًا مجموعة ممثلة من هذه الرموز ويختبر الطالب بها، ومن نتائج هذه العينة يستنتج تحقق الهدف. في هذه الحالة يُعد الاختبار **عينة سلوكية** (Sample Test). أما في حالات أخرى فقد يُستخدم الاختبار لقياس سمة لا تتعلق مباشرة بما يقيسه فعلاً بل للاستدلال على سمة أخرى، ففي هذه الحالة يُعد الاختبار **مؤشرًا** (Signe) للتمييز بين

مجموعات مختلفة. فمثلاً إذا تضمن اختبار عقلي مفردات لقياس قدرة التلميذ على تذكر أرقام بترتيب معين، وأظهرت الدراسات أن ضعف هذه القدرة يميز بين المتخلفين عقلياً والعاديين، فإن الهدف هنا ليس قياس التذكر بحد ذاته، بل استخدام الاختبار كمؤشر للتمييز بين الفئتين (علام، 2000، ص44-48).

د- أدوات القياس اللفظية مقابل أدوات قياس الأداء:

تعتمد معظم أدوات القياس على استخدام اللغة، سواء في صياغة التعليمات أو في مفردات الاختبار، ويُطلق عليها الاختبارات اللفظية (Verbal Tests)، مثل غالبية اختبارات الجوانب المعرفية والوجدانية. في المقابل توجد أدوات لا تتطلب استخدام اللغة، حيث يمكن للمفحوص الاستجابة عبر الإشارة أو باستخدام الأشكال والرسوم والصور أو المواد العينية، وتعرف بالاختبارات غير اللفظية (Non-Verbal). وغالباً لا تتطلب هذه الأدوات إجابات كتابية، بل تعتمد على أداء عملي مرتبط بالخبرات الحسية المقدمة. ومن أمثلتها بعض الاختبارات الفردية لقياس الذكاء، أو القدرات الميكانيكية، أو تنفيذ مهام عملية معينة. والجدير بالذكر أن كثيراً من اختبارات الذكاء الفردية تجمع بين أجزاء لفظية وأخرى أدائية، مما يتيح قياس الذكاء غير اللفظي، والمقارنة بين الجزأين لتحديد جوانب القصور اللغوي لدى الفرد.

هـ أدوات القياس التي تتطلب السرعة مقابل أدوات القياس التي تتطلب القوة:

تختلف أدوات القياس في مدى تحديدها لزمان الإجابة. فاختبارات السرعة (Speed Tests) تُحدد بزمان دقيق، وغالباً تحتوي على مفردات سهلة نسبياً، لكن قصر الوقت يجعل معظم الأفراد غير قادرين على إكمالها، باستثناء ذوي السرعة العالية في الأداء. في هذا النوع من الاختبارات يظهر تباين الدرجات بسبب اختلاف سرعة الإجابة، لا صعوبة المفردات. وتشمل أمثلتها بعض اختبارات الذكاء والاستعدادات. في المقابل تهتم اختبارات القوة (Power Tests) بقياس قدرة الفرد على حل مشكلات صعبة دون التقيد بالزمان، حيث يُعطى وقتاً كافياً للتفكير والإجابة. ويشيع هذا النوع في اختبارات التحصيل الدراسي التي تركز على ما اكتسبه الطالب من معارف ومهارات. تجدر الإشارة إلى أنه لا يوجد حد فاصل صارم بين النوعين؛ فقد يكون الاختبار سرعةً لمجموعة، وقوةً لأخرى، تبعاً لمدى صعوبة مفرداته بالنسبة لهم. لذا من المهم تحديد طبيعة الاختبار بدقة خاصة إذا كان يعتمد جزئياً على السرعة.

و- أدوات القياس الموضوعية مقابل أدوات القياس الذاتية:

تشير الموضوعية (Objectivity) والذاتية (Subjectivity) في القياس التربوي والنفسي إلى طريقة تصحيح الاختبارات، وليس إلى نوع الاختبار نفسه. تتميز الاختبارات الموضوعية، مثل أسئلة الاختيار من متعدد، بأن تصحيحها لا يتأثر بالأحكام الذاتية للمصحح، نظرًا لوجود مفتاح إجابة ثابت يمكن تطبيقه يدويًا أو آليًا عبر الحاسوب، مما يضمن حصول الفرد على نفس الدرجة بغض النظر عن يقوم بالتصحيح. أما الاختبارات الذاتية كالأسئلة المقالية أو الأسئلة المفتوحة، فإن درجاتها قد تختلف باختلاف المحكمين نظرًا لتأثر التقدير بالاجتهاد الشخصي. ويمكن الحد من هذا التأثير عبر صياغة الأسئلة بوضوح وتحديد المطلوب بدقة، ووضع معايير أو محكات أساسية للإجابة تُستخدم كمرجع عند تقدير الدرجات، مما يقلل من الذاتية ويزيد من اتساق النتائج. وبشكل عام فإن تقنين إجراءات التطبيق، وتعليمات الأداء، وآلية التصحيح، يساهم في رفع درجة الموضوعية بغض النظر عن نوع أداة القياس المستخدمة.

ز- أدوات القياس التي يعدها المعلم مقابل أدوات القياس المقتنة:

يستخدم المعلم أدوات القياس لمتابعة مدى تحقق أهداف التعلم لدى طلابه، ليس فقط لتحديد درجاتهم في نهاية فترة دراسية، بل أيضًا للكشف عن نقاط القوة والضعف في تحصيلهم، مما يساعده على اقتراح أساليب علاجية مناسبة للطلاب الذين يواجهون صعوبات. في هذه الحالة يقوم المعلم بإعداد اختبارات ذات وظيفة تعليمية، تهدف إلى تحسين دافعية الطلاب وتشجيعهم على بذل مزيد من الجهد لتعزيز تعلمهم. وغالبًا تكون هذه الاختبارات مرجعية المحك، ويعتمد نجاحها على إلمام المعلم بأسس بناء الاختبارات الصفية. أما عندما يرغب المعلم في مقارنة مستوى تحصيل طلابه بمستوى مجموعة أوسع تمثل المجتمع المرجعي للطلاب على المستوى الوطني – والمعروفة باسم الجماعة المرجعية (Norm Group) فإنه يحتاج إلى استخدام اختبارات مقننة (Standardized Tests). تُبنى هذه الاختبارات على يد خبراء في القياس التربوي والمناهج، ويتم تقنينها من خلال تطبيقها وتصحيحها في ظروف موحدة على الجماعة المرجعية، للوصول إلى معايير (Norms) يمكن على أساسها إجراء المقارنات بين المجموعات المختلفة. تُستخدم الاختبارات المقننة أيضًا على نطاق واسع في المقاييس النفسية، حيث تُقارن نتائج الفرد بأداء أقرانه في الجوانب العقلية والوجدانية والمزاجية، ويُفسر أداء الفرد فقط في ضوء معيار الجماعة المرجعية (علام، 2000، ص44-48).

2- تقنين الاختبارات التربوية والنفسية:

يشير **مصطلح التقنين (Standardization)** إلى عملية توحيد الظروف المحيطة بجميع المفحوصين وضبط كل العوامل والمتغيرات التي قد تؤثر على أدائهم في الاختبار، بحيث يتم تقليل تأثير هذه العوامل إلى أدنى حد ممكن. والهدف من ذلك هو قياس المتغير الذي صُمم الاختبار لقياسه بدقة، وإظهار حجمه الحقيقي لدى العينة التي يُطبَّق عليها، دون أن تتأثر النتائج بمتغيرات خارجية. ومن الخطأ اعتبار التقنين مجرد خطوة نهائية أو لاحقة لعملية بناء الاختبار؛ بل هو جزء أساسي يرافق جميع مراحل إعداد الاختبار. فالتقنين بمفهومه الشامل يقوم على وضع خطة واضحة ومتكاملة تشمل جميع خطوات تصميم الاختبار، وإجراءاته، وطريقة تطبيقه وتصحيحه، وتفسير درجاته، إضافةً إلى تحديد السلوك المطلوب من المفحوص، والظروف المحيطة به أثناء التطبيق، مع توفير معايير دقيقة لتفسير النتائج. ويميز خبراء القياس بين معنيين للتقنين، وهما:

المعنى الأول: أن تكون تعليمات الاختبار، وصياغة بنوده، والزمن المخصص له، وشروط تطبيقه، وطريقة تصحيحه موحدة في جميع المواقف، بما يضمن الحصول على نتائج متشابهة في حال إعادة الاختبار. وبدون هذا النوع من التقنين يفقد الاختبار أساسه العلمي والموضوعي. ومن هذه الناحية التقنين يعني التوحيد (مخايل، 2020، ص94).

المعنى الثاني: التقنين يشير إلى إخضاع الاختبار لعملية تطبيقه على عينة ممثلة للمجتمع الأصلي، وذلك بهدف اشتقاق معايير تحدد دلالة الدرجة الخام التي يحصل عليها الفرد، بالاستناد إلى موقعها وانتشارها بين درجات أفراد عينة التقنين. ومن الواضح أنه لا يمكن تحقيق هذا النوع من التقنين دون استيفاء شروط التقنين بالمعنى الأول، حيث تمثل عملية استخراج المعايير المرحلة الأخيرة في تطوير الاختبار، وتوفر فوائد عديدة، منها تحديد **المكانة النسبية** للفرد داخل المجموعة، وتفسير الدرجة التي حصل عليها. وتُعد عملية استخراج المعايير مهمة معقدة ومجهدّة من الناحية التطبيقية والإحصائية، وتتطلب تعاون عدد من المختصين والخبراء، إضافةً إلى كونها عملية مكلفة مادياً وتستغرق وقتاً طويلاً. ومن الجدير بالذكر أن الاختبار الذي تم تقنيته بالمعنى الأول فقط، أي من حيث توحيد إجراءات التطبيق والتصحيح، يمكن أن يكون مفيداً في التشخيص والبحث العلمي، وفي جميع الحالات التي لا يركز فيها الفاحص على تفسير الدرجة الخام أو تحديد موقع الفرد النسبي داخل المجموعة (مخايل، 1996).

جدول 03 يوضح الفروق بين الاختبارات المقننة والاختبارات غير المقننة.

المعيار	الاختبارات المقننة (Standardized Tests)	الاختبارات غير المقننة (Non-Standardized Tests)
التعريف	أدوات قياس تم إعدادها وفق إجراءات موحدة ومعايير دقيقة لضمان الموضوعية والثبات والصدق.	أدوات قياس تُعد غالباً لأغراض محلية أو خاصة دون التقيد الصارم بإجراءات ومعايير موحدة.
إجراءات الإعداد	تعتمد على خطوات علمية منظمة تشمل تحديد الأهداف، صياغة البنود، التجريب الميداني، التحليل الإحصائي، وبناء المعايير.	غالباً تُعد بسرعة لتلبية حاجة محددة، مع تقليل أو غياب خطوات التحليل الإحصائي والمعياري.
التطبيق	يتم تطبيقها وفق تعليمات موحدة على جميع الأفراد في نفس الظروف.	قد تختلف طريقة التطبيق بين الأفراد أو المجموعات، ولا يشترط التوحيد التام.
المعايير	لها معايير (Norms) مستمدة من عينة ممثلة، تُستخدم لمقارنة أداء الفرد بأداء المجموعة المرجعية.	غالباً لا تملك معايير إحصائية دقيقة، وتعتمد على تقديرات أو معايير محلية.
الصدق والثبات	عالية لأن الاختبار يخضع لتحليل إحصائي وتجريب قبل اعتماده.	أقل دقة من حيث الصدق والثبات بسبب ضعف أو غياب الإجراءات العلمية في الإعداد.
الأمثلة	اختبار الذكاء وكسلر، اختبار SAT، اختبارات TIMSS، اختبارات التحصيل الوطنية المقننة.	اختبارات المعلم الصفية، أسئلة المراجعة اليومية، الاختبارات القصيرة الخاصة بالمقرر.
مجالات الاستخدام	القياس على نطاق واسع، المقارنات الوطنية والدولية، الأبحاث العلمية.	التقييم الفوري للأداء، متابعة التعلم داخل الصف.

3- معنى المعايير وأهميتها:

لفهم معنى المعايير وأهميتها في ميدان القياس النفسي والتربوي، ينبغي أولاً التوقف عند الدرجة الخام التي يحصل عليها الفرد في اختبار معين، إذ تمثل هذه الدرجة النتيجة الأولية المباشرة للأداء على الاختبار.

أ- في اختبارات الأداء الأقصى (أو اختبارات القوة): مثل اختبارات الذكاء والقدرة العقلية العامة، واختبارات القدرات الخاصة، والاختبارات التحصيلية بمختلف أنواعها، تُحسب الدرجة الخام عادةً بجمع عدد الإجابات الصحيحة التي يقدمها المفحوص على البنود.

ب- في اختبارات الأداء النمطي (أو العادي): مثل اختبارات الشخصية الشاملة أو مقاييس الاتجاهات والقيم والميول، فإن الدرجة الخام تُحسب بجمع عدد الاستجابات التي تتماشى مع الاتجاه العام أو المنحنى الذي يهدف إليه الاختبار.

غير أن هذه الدرجات الخام – سواء في اختبارات الأداء الأقصى أو النمطي – لا تُعد وحدات قياس متساوية أو منتظمة، إذ أن الفروق بينها لا تعكس بالضرورة فروقاً حقيقية بين الأفراد في

الخاصية المقاسة. ومن هنا تظهر أهمية المعايير التي تسمح بتحويل الدرجات الخام إلى درجات معيارية أو مشتقة يمكن تفسيرها بصورة أكثر دقة، إذ تتيح المقارنة بين أداء الفرد وأداء مجموعة معيارية، ما يمنح المعنى الحقيقي لنتائج الاختبار ويجعلها قابلة للاستخدام في التشخيص والتوجيه واتخاذ القرارات التربوية أو النفسية (مخايل، 2020، ص100).

4- المعايير وعينة التقنين:

تستمد المعايير من عينة التقنين التي تُعد تمثيلاً صادقاً للمجتمع الأصلي الذي يُستهدف بالدراسة. ويتم ذلك خلال الخطوات التجريبية التي يمر بها المقياس حتى يصل إلى صورته النهائية. ففي هذه المرحلة، تُحوّل الدرجات الخام التي حصل عليها أفراد عينة التقنين إلى درجات معيارية، لتصبح مرجعاً يمكن من خلاله الحكم على أداء الأفراد الآخرين. وبمعنى آخر عندما نريد تفسير درجة المفحوص في اختبار معين، فإننا نرجع إلى من يماثله في عينة التقنين.

وعليه فإن المعايير ما هي إلا حصيلة تطبيق الاختبار على عينات التقنين. وعند نقل اختبار من بيئة إلى أخرى أو من مجتمع لآخر، يصبح من الضروري إعادة تقنيه على عينات جديدة موازية للعينة الأصلية من حيث الخصائص، وذلك لاشتقاق معايير تتلاءم مع الواقع الجديد. ويرجع ذلك إلى أن الأفراد قد يشهدون تغيرات ناتجة عن عوامل متعددة كالفرقات الثقافية، والتقدم التكنولوجي، والتحول الاقتصادي والاجتماعي. ومن ثم فإن عملية التقنين لا تُعد حدثاً ثابتاً لمرة واحدة، بل هي عملية دورية تستوجب المراجعة وإعادة التقييم على فترات زمنية منتظمة، لضمان بقاء الاختبار صالحاً ودقيقاً في قياسه للظاهرة المستهدفة (زياد، 2023، ص288).

5- أنواع المعايير:

تشكل المعايير أساساً لتفسير الدرجة الخام التي يحصل عليها الفرد في الاختبار، إذ يتم من خلالها تحويل هذه الدرجة إلى عدة أنواع من المؤشرات أبرزها: المكافئ العمري، المكافئ الصفي، الدرجة المئينية، الدرجة المعيارية. وتتيح هذه المؤشرات فهم مستوى أداء المفحوص على الاختبار. وعندما يخضع الفرد لعدة اختبارات، يمكن بفضل هذه المعايير توحيد نتائجه في صفحة نفسية واحدة، مما يساعد على تكوين صورة وصفية عن مستوى أدائه النسبي في مجالات متعددة. وبهذا توفر المعايير إطاراً عاماً يوضح المستوى الكلي للمفحوص ويرسم الخطوط العريضة التي تسهم في تشخيص حالته بدقة وتجنب أخطاء التفسير (Angoff, 1971).

لذلك تبرز أهمية التعرف إلى أنواع هذه المعايير، والوقوف على مميزات كل نوع منها وما يقابله من جوانب قصور، وهو ما سيتناوله العرض التالي (بركات، 2018، ص336).

أولاً: المعايير الصفية (Grade Norms):

يقصد بالمعايير الصفية متوسط الدرجات التي يحققها أفراد صف دراسي معين. ولإعداد هذه المعايير يُطبَّق الاختبار على عينة من الطلبة تمثل عدداً من الصفوف المتتالية، ثم يُحسب المتوسط الحسابي لدرجاتهم في كل صف، وتُحدد القيم المعيارية الصفية (Normative values) التي تقابل الدرجات الواقعة بين هذه المتوسطات. غير أن استخدام هذا النوع من المعايير يواجه تحفظين أساسيين:

أ- **عدم تساوي الوحدات الصفية:** فالمعايير الصفية لا تقوم على وحدات متكافئة حقيقية، بل على وحدات تقريبية مصطنعة تُستخدم فقط لأغراض تفسير الدرجات الخام المحولة إلى المكافئ الصفي (Grade Equivalent) الذي يمثل متوسط أداء الطلبة في كل صف.

ب- **محدودية الدلالة التفسيرية:** حصول المفحوص على درجة تعادل صفًا معيناً لا يعني أنه أتقن جميع محتويات ذلك الصف بشكل كامل، بل يشير فقط إلى الدرجة التي حصل عليها دون توضيح كيفية الوصول إليها (علام، 2000، ص254).

وبناءً على ذلك يظل المكافئ الصفي موضع شك من حيث صلاحيته للمقارنة بين مواد دراسية مختلفة. ورغم سهولة استخلاص هذه المعايير لكونها تعتمد على مجموعات متوفرة أصلاً داخل المدرسة، إلا أنه لا يمكن الاعتماد عليها كإطار مرجعي دقيق لمقارنة أداء طالب فردي بأداء الطالب المتوسط في الصفوف المختلفة، خاصة مع غياب الثقة في تكافؤ وحدات الصف بين مادة وأخرى.

- مزايا المعايير الصفية:

تتميز المعايير الصفية بعدة جوانب إيجابية، من أبرزها:

أ- توفر إطاراً مرجعياً يساعد على تفسير مستوى التحصيل الدراسي للطلبة.

ب- سهولة استخدام المعادلات الإحصائية المختلفة معها، مثل: حساب المتوسط الحسابي والانحراف المعياري والنسب المئوية وغيرها.

ج- إمكانية تحويل الدرجات الخام إلى درجات معيارية مكافئة، مما يسهل مقارنتها وتفسيرها.

د- تتيح أساساً معقولاً للمقارنة بين أداء المفحوص ومتوسط أداء الصف.

ح- يمكن الاستفادة منها في عمليات التنبؤ والتشخيص للكشف عن مواطن القوة والضعف في العملية التعليمية.

- عيوب المعايير الصفية:

على الرغم من المزايا التي توفرها المعايير الصفية، إلا أنها تعاني من عدد من أوجه القصور، من أبرزها:

أ- تفاوت مستوى التعليم بين الصفوف المختلفة.

ب- لا يمكن اعتبار حصول المفحوص على مكافئ صفي أعلى دليلاً قاطعاً على امتلاكه مستوى ذلك الصف بشكل حقيقي وكامل.

ج- وحدات المكافئ الصفية ليست متساوية بالضرورة، رغم ارتباطها بالسنوات الدراسية.

د- تقوم هذه المعايير على افتراض أن النمو يحدث بشكل منتظم طوال العام الدراسي من شهر لآخر، وهو افتراض غير مبرر علمياً.

هـ- يختلف معدل النمو بين الصفوف، إذ ينمو الطلبة في الصفوف الأولى بوتيرة أسرع مقارنة بالصفوف العليا، وقد يُعزى ذلك أحياناً إلى النمو الطبيعي والخبرات المصاحبة له أكثر من التعلم نفسه.

و- تباين قدرات الطلبة بين الصفوف المختلفة وكذلك بين المواد الدراسية يحد من فعالية المكافئات الصفية.

ز- غياب التجانس داخل الصف الواحد، حيث تتواجد فروق فردية واسعة بين الطلبة قد تصل إلى وجود متفوقين جداً مقابل متأخرين جداً.

ح- صعوبة تحديد تدريجات صفية دقيقة للدرجات المتطرفة (العليا أو الدنيا) نظراً لضعف ثباتها.

ط- اختلاف أساليب التدريس وطرائقه بين المعلمين قد يؤثر بشكل مباشر على مستوى التحصيل الدراسي لطلبة الصفوف المختلفة.

ي- احتمال التركيز في العملية التعليمية على بعض الصفوف أكثر من غيرها، مما يؤدي إلى اختلاف معاملات الارتباط بين نتائج الاختبارات في الموضوعات المختلفة.

وبسبب هذه العيوب غالباً ما يلجأ معدّو الاختبارات التحصيلية إلى الاستعانة بأنواع أخرى من

المعايير أو تعديل المعايير الصفية ذاتها. ومن أبرز المقترحات في هذا المجال ما قدّمه كيلي (Kelley,

1981) من اعتماد معيار العمر المتوالي (Modal-Age Norm) عبر تحديد متوسط درجات الصف

واستبعاد الدرجات البعيدة عنه، إضافة إلى استخدام المعايير العمرية النسبية والميئنية (بركات،

2018، ص337).

ثانياً: المعايير العمرية (Age Norms):

المعايير العمرية يقصد بها متوسط الدرجات التي يحصل عليها الأفراد ضمن فئة عمرية محددة، كأعمار الخامسة أو السادسة أو السابعة أو الثامنة، ويتم التوصل إليها عبر فحص عينات ممثلة من كل فئة عمرية. وهي تشبه المعايير الصفية من حيث الإجراءات، إلا أنها تواجه إشكالية مشابهة تتعلق بمدى اعتبار السنة العمرية وحدة معيارية ثابتة ومنظمة. فهل النمو بين السنة الرابعة والخامسة يعادل النمو بين الخامسة والسادسة من حيث الكمية والنوعية؟ يبقى هذا السؤال مطروحاً خصوصاً أن قوانين النمو تشير إلى أن وحدة السنة تصبح أقل دلالة مع التقدم في العمر؛ فبعد سن العشرين مثلاً يتباطأ النمو إلى درجة تفقد معها السنة قيمتها كوحدة مناسبة للمقارنة.

- خطوات استخراج المعايير العمرية:

أ- اختيار عينات من كل فئة عمرية بحيث تشمل أعماراً تزيد أو تنقص عن العمر المحدد بنحو ستة أشهر.

ب- تطبيق الاختبار على الأفراد في مختلف الأعمار، وقد تُعتمد عدة اختبارات بحيث تُدرج فقرات مشتركة بين الأعمار المختلفة تسمى فقرات الإسناد (Anchor Items)، تُستخدم في عملية معادلة الاختبارات (Equating) (Ru & Holland, 2009).

ج- حساب المتوسط الحسابي لدرجات الأفراد في كل فئة عمرية بعد عملية المعادلة.

د- رسم منحنى يوضح العلاقة بين نتائج الاختبار وأعمار الأفراد بالسنوات.

هـ- ربط كل عمر بالدرجة الخام المكافئة له، بحيث يُعتبر الطفل البالغ 6 سنوات مثلاً ذا "عمر عقلي" يعادل 8 سنوات إذا حصل على درجة خام مساوية لمتوسط أداء الأطفال ذوي 8 سنوات، رغم أن عمره الزمني الفعلي هو 6 سنوات (علام، 2000، ص253).

- استخدامات وحدود المعايير العمرية:

أ- تُعد مناسبة للسمات التي تتأثر بالنمو الطبيعي العام، مثل: الطول، الوزن، الذكاء، والاستعدادات المختلفة، حيث تتخذ هذه السمات منحنى نمو مستمر (Continuous).

ب- أما السمات المرتبطة بالخبرات التعليمية المكتسبة، مثل: المهارات الحسابية واللغوية، فإنها تعكس تأثير العوامل التربوية أكثر من النمو البيولوجي.

ج- ورغم أن هذه المعايير توفر إطاراً يسهل من خلاله تفسير أداء الفرد مقارنة بالفئة العمرية التي ينتمي إليها، إلا أن الشكوك تبقى قائمة حول مدى تساوي وحدات العمر مع التقدم في العمر، مما يحد من صلاحيتها في بعض المجالات (Angoff, 1982).

- مزايا المعايير العمرية:

- أ- تتمتع المعايير العمرية بعدد من الإيجابيات، من أبرزها:
- ب- سهولة استخدامها وبساطة تطبيقها دون الحاجة إلى إجراءات معقدة.
- ج- شيوخ استخدامها على نطاق واسع في العديد من اختبارات الذكاء والقدرات العقلية العامة.

- عيوب المعايير العمرية:

- أ- على الرغم من مزاياها، فإن المعايير العمرية تعاني من عدة جوانب قصور، أهمها:
- ب- محدودية فائدتها في تقدير سمات الأفراد المتطرفين سواء في التفوق أو في التأخر.
- ج- اختلاف معدل سرعة النمو من عمر إلى آخر بل ومن وظيفة لأخرى ومن فرد لآخر مما يقلل من دقة الاعتماد على العمر كمعيار للمقارنة.
- د- العلاقة بين العمر ونتائج الاختبارات ليست دقيقة ولا تامة كما أن منطق هذه العلاقة يفتقر إلى التبرير الكافي.
- هـ- خطوط الانحدار التي تُرسم لبيان العلاقة بين العمر ونتائج الاختبار لا تأخذ في الاعتبار حجم التباين بين الأعمار المختلفة مما يعني غياب ثبات التباين في الدرجات.
- و- العمر العقلي يزداد تدريجياً مع التقدم في العمر حتى يصل إلى مستوى معين ثم يبدأ في التراجع وهذا يضعف صلاحية الاعتماد على المعيار العمري بشكل دائم.
- ز- الحكم على العمر العقلي بمقارنته بالعمر الزمني يظل حكماً مضللاً إذ لا يعني أن الفرد الذي يحصل على عمر عقلي أعلى من عمره الزمني قد أتيقن جميع مهارات ذلك العمر بشكل كامل (بركات، 2018، ص340).

ثالثاً: المعايير المئينية (Percentile Norms):

يمكن فهم المئينيات بشكل مبسط من خلال تصور اختبار أجري على 100 شخص. يتم ترتيب هؤلاء الأشخاص تصاعدياً بحسب درجاتهم بحيث يقف صاحب أقل درجة في المرتبة الأولى ويقف صاحب الدرجة المتوسطة عند المرتبة 50 بينما يحتل صاحب أعلى درجة المرتبة 100. هذا الترتيب يوضح أن المئينيات تقوم أساساً على مبدأ الترتيب الرتبي. ونظراً لطبيعتها الترتيبية فإن المئينيات

تُعرض عادةً في صورة أعداد صحيحة فقط. ومن المهم التمييز بين المئينيات والنسب المئوية، إذ يختلفان عن بعضهما رغم وجود ارتباط بينهما.

تساعد المئينيات في مقارنة درجة الفرد بدرجات مجموعة مرجعية، حيث توفر مقياساً نسبياً للمقارنة مع المتوسط الذي يُمثل بالمئين 50. فمثلاً الحصول على المئين 75 يعني أن أداء الفرد أعلى من المتوسط، بينما المئين 30 يشير إلى أداء أقل من المتوسط. أما نسبة الأشخاص الذين حصلوا على درجات أدنى من درجة معينة فتُعرف باسم الرتبة المئينية (Percentile Rank)، ويُطلق عليها أحياناً ببساطة "المئين". ولتحديدها يجب معرفة نسبة الأفراد الذين سجلوا درجات أقل من الدرجة المعنية (Coaley, et al, 2010, p.66).

وتحسب الرتبة المئينية للدرجة الخام باستخدام الصيغة الرياضية التالية:

$$p = \frac{(n_b + .5(n_e))}{n_t} \times 100$$

حيث أن:

n_b هو عدد الطلبة الذين حصلوا على أقل من الدرجة التي نريد إيجاد المئيني لها.

n_e هو عدد الطلبة الذين لهم نفس الدرجة مع الدرجة التي نريد إيجاد المئيني لها.

n_t هو العدد الكلي للطلبة الممتحنين على الاختبار.

مثال: إذا كان عندنا $n_t = 200$ ؛ و $n_b = 98$ ؛ و $n_e = 4$ ؛ والدرجة الخام التي نريد إيجاد الميني لها تساوي 15.

الحل:

$$p = \frac{(98 + 0.5(4))}{200} \times 100 = 50$$

المئين هو 50.

6- أنواع المعايير أو الدرجات المحوّلة (Norms):

تُعد الدرجات المحوّلة من الأدوات الأساسية في ميدان القياس النفسي والتربوي، إذ تُستخدم لتحديد المركز النسبي للفرد داخل توزيع معيّن، بحيث يمكن وصف أدائه مقارنةً بأداء أقرانه في اختبار يقيس سمة أو قدرة معينة. ويُعتبر هذا المفهوم وثيق الصلة بالاختبارات مرجعية الجماعة

(Norm-Referenced Tests) والتي تهدف بالأساس إلى تقدير موقع الفرد بالنسبة إلى بقية الأفراد، حيث يكون الاهتمام منصّباً على الترتيب النسبي بينهم أكثر من تركيزه على مقدار الأداء الفعلي. في هذا السياق لا تُفسّر الدرجة التي يحصل عليها المفحوص إلا في ضوء معيار الجماعة (Group Norm) أي متوسط أداء الأفراد وانحرافاتهم عن هذا المتوسط. ومن ثمّ تصبح الدرجة عديمة الدلالة إذا لم تُردّ إلى هذا المرجع لأن معناها مرتبط بموقعها في التوزيع الجماعي (علام، 2000، ص. 97).

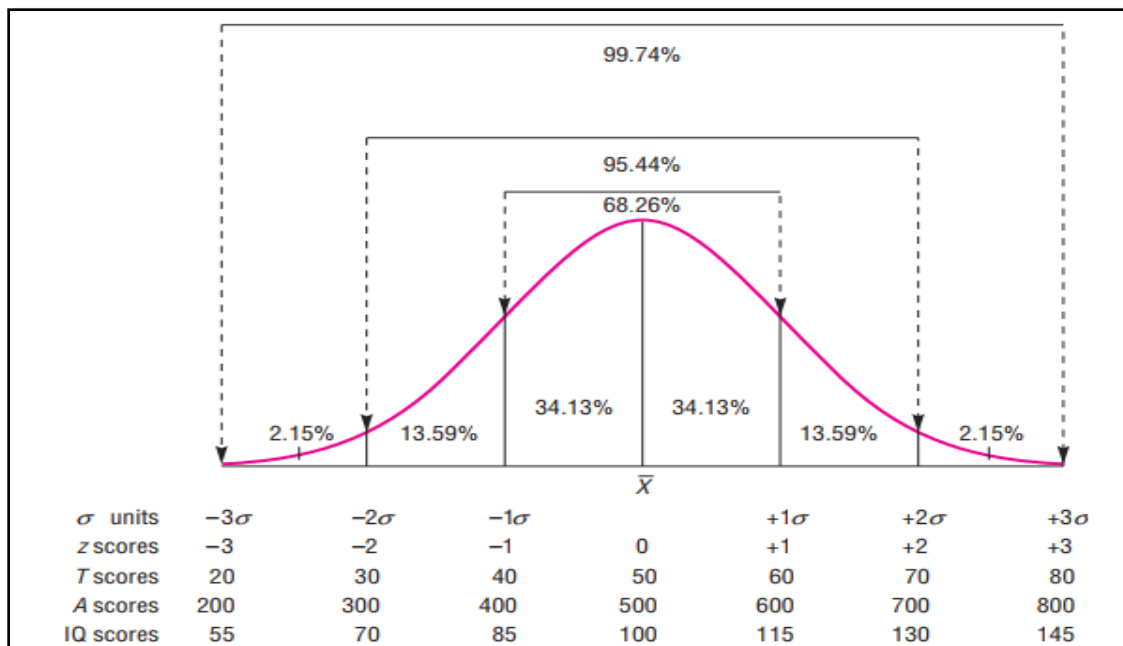
وتوجد عدة أنماط من الدرجات المحوّلة غير أن أكثرها شيوعاً في القياس التربوي والنفسي هما:

أ- **الدرجات المعيارية (Standard Scores):** والتي تتيح تحويل الدرجات الخام إلى درجات ذات معنى موحد تمكّننا من المقارنة بين اختبارات مختلفة.

ب- **الدرجات التائية (T-Scores):** وهي شكل خاص من الدرجات المعيارية يسهّل التفسير ويُستخدم على نطاق واسع في الاختبارات النفسية والتربوية.

ويعتمد كلا النوعين على التحويلات الخطية للدرجات الخام باستخدام كلّ من المتوسط الحسابي والانحراف المعياري معاً، بما يسمح بتحقيق قدر أكبر من الدقة في التفسير، كما يوضحه الشكل التالي:

شكل 01 يوضح بعض الدرجات المعيارية المحوّلة



نلاحظ أن القيم المعتمدة لدرجات اختبار الذكاء (IQ) غالبًا ما تفترض وجود متوسط حسابي مقداره 100 وانحراف معياري يساوي 15، وهي المعايير الأكثر شيوعًا في كثير من مقاييس الذكاء الحديثة. غير أن هذه القيم لا تنطبق بالضرورة على جميع اختبارات الذكاء إذ قد تُبنى بعض الاختبارات على متوسطات أو انحرافات معيارية مختلفة.

وفي حال أسفر اختبار معين عن نتائج ذات متوسط حسابي لا يساوي 100 أو انحراف معياري يختلف عن 15، فإن القيم المستخلصة من هذا الاختبار لا يمكن تفسيرها مباشرة بالرجوع إلى المعايير المتعارف عليها عالميًا. بل ينبغي في هذه الحالة تعديل القيم المحوَّلة والمعروضة بما يتناسب مع الخصائص الإحصائية للاختبار المعني، حتى تكون عملية التفسير أكثر دقة وموضوعية (Swerdlik, 2009).

أولاً: الدرجات المعيارية (Standard Scores):

تُعد الدرجات المعيارية من أهم الأدوات الإحصائية المستخدمة في ميدان القياس التربوي والنفسي، إذ يمكن توظيفها في مختلف التحليلات الإحصائية، مثل: المقارنة بين متوسطات المجموعات أو حساب معاملات الارتباط بين الدرجات. ونظرًا لشهرتها ومرونتها وسهولة تفسيرها، فقد أصبحت تُستخدم على نطاق واسع في مواقف التقييم والبحث التربوي والنفسي. وتستند فكرة الدرجة المعيارية إلى التحويل الخطي للدرجة الخام إلى وحدة قياس موحدة تجعل المقارنات أكثر موضوعية، حيث يُصبح أداء الفرد قابلاً للمقارنة مع غيره بغض النظر عن الاختلافات في مقاييس أو وحدات الاختبارات. تبدأ الخطوة الأولى في هذا التحويل بحساب الدرجة الزائنية (z-score) والتي تُمكن الباحث من تحديد موقع الدرجة الخام بالنسبة للمتوسط الحسابي، معبرًا عنها بعدد الانحرافات المعيارية التي تفصلها عنه. وتمثل هذه الخطوة الأساس الذي تُبنى عليه بقية أشكال الدرجات المعيارية مثل الدرجات التائية (T-scores) أو درجات الذكاء (IQ scores) وغيرها، وتعطي بالمعادلة التالية:

$$z = \frac{x - \bar{x}}{s}$$

حيث أن:

X : الدرجة الخام.

\bar{X} : المتوسط الحسابي للعينة.

S: الانحراف المعياري.

تعتمد الدرجة الزائفة (Z) على متوسط حسابي يساوي صفرًا وانحراف معياري يساوي واحدًا، حيث يتم تحويل الدرجة الخام إلى درجة زائفة تبين عدد الانحرافات المعيارية التي تفصلها عن المتوسط. فالقيمة (-2) تعني أن الدرجة الخام أقل من المتوسط بانحرافين معياريين، بينما تشير القيمة (1) إلى أنها أعلى بانحراف معياري واحد. ويسمح هذا التحويل بوضع جميع الدرجات على مقياس موحد يمكن من خلاله مقارنة الأداء بين مختلف الاختبارات، مثل الذكاء اللفظي مقابل الذكاء المكاني، كما يوفر مؤشرًا دقيقًا لموقع الفرد مقارنةً بالعينة الكلية. وتتميز وحدات (Z) بالثبات إذ يعكس الفرق بين درجتين متتاليتين نفس الفاصل عبر المقياس بأكمله. ورغم مزاياها في حل مشكلات الدرجة الخام والمئينيات، إلا أن الدرجات الزائفة لا تُستخدم عادة في تقارير النتائج النهائية بسبب طبيعة مقياسها (زياد، 2023، ص301).

ثانيًا: الدرجات المعيارية المعدلة أو الدرجات التائية (T-Scores):

تُستخدم الدرجات المعيارية المعدلة للتغلب على المشكلات المرتبطة بالدرجات المعيارية التقليدية. فالفرق الأساسي بينها هو اختيار متوسط جديد بدلاً من الصفر وانحراف معياري جديد بدلاً من الواحد، ما يساعد على التخلص من الأعداد السالبة والكسور العشرية. كما أن مدى توزيع الدرجات المعيارية محدود إذ لا يزيد عادة عن ست درجات مما يقلل من الفروق الفردية بين المتعلمين ويجعلها غير واضحة خاصة في العينات الكبيرة. هذا يجعل استخدام الدرجات المعيارية في التقييم التربوي العملي صعبًا فقد يواجه أولياء الأمور أو المعلمون صعوبة في تفسير درجة زائفة مثل (-1.5) لفهم أداء الطالب. لذلك يُعد تعديل الدرجات المعيارية الحل الأمثل حيث يتم تحويل جميع الدرجات إلى قيم موجبة وتسوية المنحنى لتسهيل المقارنة ما يضمن وضوحًا أكبر للفروق الفردية وإمكانية الاستفادة العملية منها في التقييم التعليمي (أبو ناهية، 1994، ص. 395).

ولذلك يتم تحويل الدرجات الزائفة إلى مقياس آخر أكثر فائدة عند إعداد تقارير الدرجات المدرسية إذ تُعد هذه الطريقة المفضلة لسهولة توصيل معلومات الأداء على الاختبار لجمهور واسع مثل المعلمين وأولياء الأمور. ويُجرى هذا التحويل من خلال ضرب الدرجة الزائفة المعيارية في 10 للتخلص من الكسور العشرية، ثم إضافة 50 كمتوسط حسابي للتخلص من الإشارات السالبة، وفق الصيغة الرياضية التالية:

$$\text{Standard score} = z * \sigma + \mu$$

حيث أن:

Z : الدرجة الزائفة.

σ : هو الانحراف المعياري المطلوب للدرجة.

μ : هو المتوسط المطلوب.

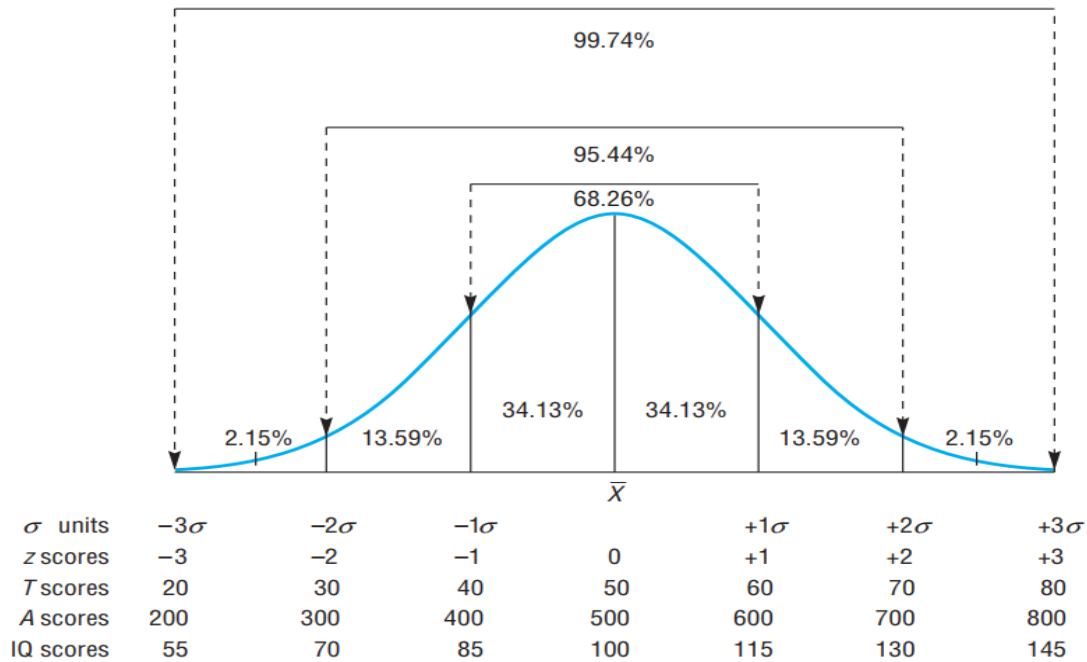
على سبيل المثال إذا أردنا تحويل الدرجة الزائفة (Z) التي تساوي (-1.5) إلى درجة معيارية معدلة (T) باستخدام الصيغة السابقة، نحصل على الآتي:

$$T = -1.5 * 10 + 50 = 35$$

وبالمثل عند التعامل مع درجات معدل الذكاء (IQ) والتي لها متوسط حسابي 100 وانحراف معياري 15، يُحوّل أولاً الدرجة الخام إلى درجة زائفة (Z)، ثم تُطبق المعادلة لتحويلها إلى درجة معيارية معدلة.

ومن الجدير بالذكر أنه بالإمكان اختيار متوسط جديد وانحراف معياري جديد حسب الحاجة، شريطة أن يتم التحويل بضرب الدرجة المعيارية في مقدار ثابت لإزالة الكسور العشرية، وإضافة مقدار ثابت آخر لإلغاء الإشارات السالبة. ورغم وجود عدة طرق للتحويل تبقى الدرجة الناتجة (T -score) من أهم وأكثر الطرق استخداماً لتحويل الدرجات الخام إلى درجات معيارية معدلة سهلة الفهم والتفسير (زياد، 2023، ص304).

شكل 02 يوضح بعض الدرجات المعيارية المحولة



تجدر الإشارة إلى أن القيم المعروضة لدرجات اختبارات الذكاء (IQ) تعتمد على فرضية أن متوسط الدرجات يساوي 100 والانحراف المعياري يساوي 15، وهو ما ينطبق على العديد من اختبارات الذكاء لكنه ليس قاعدة عامة لجميعها. وفي حال كانت نتائج اختبار معين تظهر متوسطًا مختلفًا عن 100 أو انحرافًا معياريًا مختلفًا عن 15، فإنه يلزم تعديل قيم درجات الذكاء لتتوافق مع هذه الخصائص الإحصائية (Cohen, et al, 2013, p105).

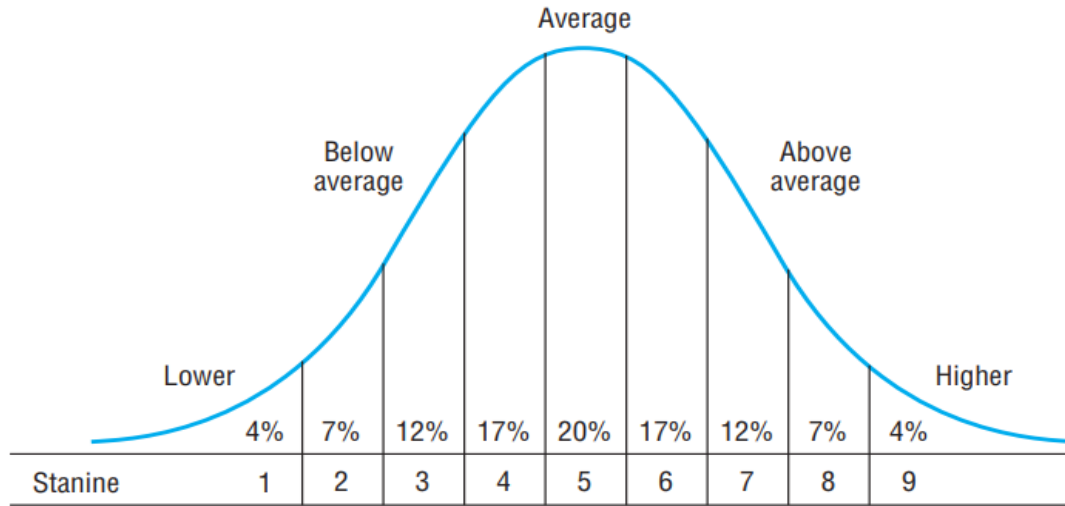
ثالثاً: الدرجات العشارية المعيارية (Sten Scores):

تُعرف درجات Sten أو Standard Ten على نطاق واسع خصوصاً في تقييم الشخصية، حيث استخدمها كاتل (Cattell) عند تصميم النسخة الأصلية من اختبار العوامل الستة عشر للشخصية (16PF). يمتد المقياس من 1 إلى 10 ويُعتبر النقطة الوسطى 5.5 هي المتوسط، مع إمكانية ظهور درجات مثل 0.5 أو 10.5 عند بناء البروفایل الشخصي، والتي يتم تعديلها لتصبح 1 و 10 على التوالي. يقسم المقياس الطبيعي إلى 10 درجات ويتميز بالبساطة والحساسية المناسبة لتمييز الفروق بين درجات السمات المختلفة. يبلغ الانحراف المعياري للمقياس 2، بحيث تقع الدرجة التي تبعد انحرافين معياريين تحت المتوسط عند 1.5، وتلك التي تبعد انحرافين معياريين فوق المتوسط عند 9.5. وتُعد درجات Sten مفيدة في تقديم التغذية الراجعة للأفراد إذ يُسهل فهمها وتفسيرها بفضل بساطتها واعتمادها على مقياس من 1 إلى 10 (Coaley, et al, 2010, p.93).

رابعاً: الدرجات التساعية (Stanine):

توجد العديد من أنظمة الدرجات المعيارية المختلفة. فخلال الحرب العالمية الثانية قام باحثون بتطوير درجة معيارية ذات متوسط حسابي يساوي 5 وانحراف معياري يقارب 2، وتم تقسيم هذا المقياس إلى تسع وحدات وسُميت هذه الدرجة Stanine، وهو مصطلح مركب من كلمتي Standard وNine. يستخدم نظام التساعيات (Stanine) غالباً في اختبارات التحصيل بالمرحلة الابتدائية والثانوية حيث تُعرض درجات الطلاب على شكل تساعيات. وتتميز التساعيات عن الدرجات المعيارية الأخرى بأنها قيم صحيحة تتراوح بين 1 و 9، وتمثل كل واحدة منها نطاق أداء بعرض نصف انحراف معياري. وتشير التساعية الخامسة إلى الأداء المتوسط، أي من ربع انحراف معياري تحت المتوسط إلى ربع انحراف معياري فوقه، وتشمل 20% الوسطى من الدرجات في التوزيع الطبيعي. أما التساعيتان الرابعة والسادسة فكل منهما بعرض نصف انحراف معياري أيضاً، وتمثل على التوالي 17% من الحالات الواقعة تحت وفوق التساعية الخامسة (Cohen, et al, 2013, p103).

شكل 03 يوضح التسايعات والتوزيع الطبيعي



أهمية هذه التحويلات:

تبرز أهمية تحويل الدرجات إلى مقاييس معيارية معدلة في عدة نقاط رئيسية:
أولاً: تجعل المتوسط الحسابي والانحراف المعياري والدرجات كلها أعداداً موجبة، مما يسهل فهم النتائج.

ثانياً: تجعل جميع الدرجات واقعة على مقياس مئوي يمتد بين الصفر والمائة، ما يوفر وضوحاً أكبر للفروق بين الأفراد.

ثالثاً: تتيح توحيد المتوسط والانحراف المعياري لجميع الدرجات إمكانية المقارنة المباشرة بين المتعلمين، سواء على مستوى نفس الاختبار أو بين اختبارات مختلفة.

7- المعايير والصفحة النفسية أو التربوية (Psychological or Educational Profile):

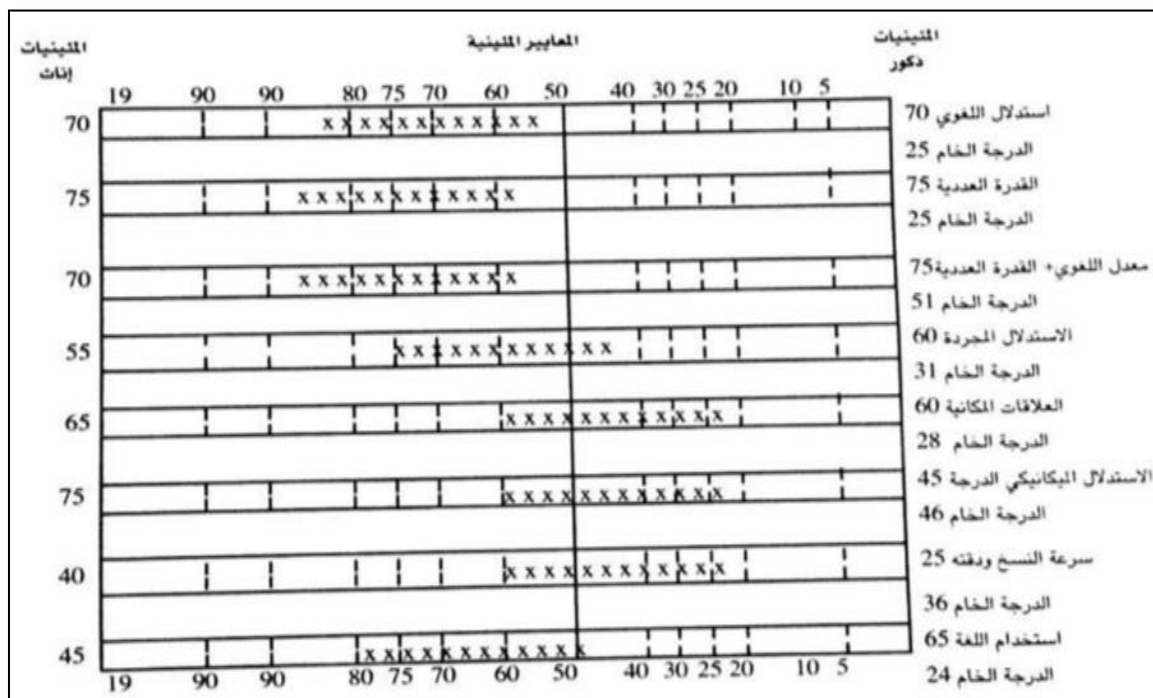
تُعد المعايير أداة أساسية ذات فائدة كبيرة في مجال القياس النفسي والتربوي، إذ تتيح لنا أساساً موحدًا للمقارنة بين الخصائص أو مظاهر السلوك التي يقيسها أي اختبار منفرد أو مجموعة من الاختبارات التي يخضع لها الفرد أو مجموعة من الأفراد. وقد أشار "ثورندايك" و"هيجن" إلى أن تنوع المعايير يزودنا بوسيلة للتعبير عن الدرجات المتحصلة من اختبارات مختلفة كلياً باستخدام وحدات قياس مشتركة الأمر الذي يجعل المقارنات أكثر وضوحاً ودقة. ويتجسد ذلك بشكل عملي في رسم ما يُعرف بـ الصفحة النفسية (البروفایل) للفرد أو لمجموعة من الأفراد حيث تمثل هذه الصفحة مجموعة متجانسة من الدرجات المعبر عنها بوحدات موحدة قابلة للمقارنة (ثورندايك وهيجن، 1989، ص125).

وبذلك يمكن القول إن الصفحة النفسية ليست سوى تمثيل بياني للدرجات التي يحصل عليها المفحوص عبر مجموعة من الاختبارات المستقلة أو الاختبارات الفرعية التي تكون اختباراً كلياً واحداً. ويكمن الهدف الأساسي من هذا التمثيل في توضيح الدرجات بصرياً عبر رسم بياني يأخذ عادةً شكل منحني، مما يجعل الفروق والأنماط أكثر وضوحاً للباحث أو الأخصائي.

ومن خلال الصفحة النفسية يمكن تحديد موقع درجات الفرد بالنسبة إلى معيار واحد أو أكثر من المعايير المعتمدة، مثل المتوسطات الحسابية، المئينيات، أو الدرجات المعيارية. كما تساعد هذه الأداة على التعرف على النمط العام للأداء، واكتشاف السمات أو مظاهر السلوك التي يظهر فيها الفرد تفوقاً أو ضعفاً مقارنةً بالمتوسط. وبهذا تصبح الصفحة النفسية وسيلة فعالة للكشف عن جوانب القوة والقصور في أداء المفحوص، سواء بمقارنة نتائجه مع نفسه عبر مجالات متعددة، أو بمقارنتها مع متوسط أداء الجماعة التي ينتمي إليها (مخائيل، 2021، ص125).

يعرض الشكل (04) نموذجاً لصفحة نفسية تتضمن الدرجات الخاصة بأحد الأفراد في بطارية اختبارات الاستعدادات أو القدرات الفارقة. ويبرز هذا النموذج بشكل واضح مجالات التفوق ونقاط التميز لدى المفحوص، بما يتيح تكوين صورة شاملة عن قدراته. ويمثل المثال بروفياً لأحد طلاب الصف الأول الثانوي في بطارية اختبارات الاستعدادات الفارقة (DAT)، حيث جرى تفسير نتائجه بالاستناد إلى معايير المئينيات المصنفة وفقاً للجنس (ذكور – إناث)، مما يضيف على التحليل دقة أكبر ويُمكن من مقارنته بأقرانه ضمن الفئة نفسها.

شكل 04 بروفيل أحد طلاب الصف الأول ثانوي في بطارية اختبارات الاستعدادات الفارقة (DAT).



تتسم الصفحات النفسية أو الرسوم البيانية التربوية بثلاث خصائص أساسية، هي: المستوى (Level)، والانتشار (Scatter)، والشكل (Shape).

1- يشير المستوى إلى الارتفاع العام للأعمدة في الرسم البياني، وهو ما يعكس مستوى الأداء الكلي للفرد أو متوسط درجاته المعيارية في الاختبارات الفرعية.

2- أما الانتشار فيدل على درجة التباين أو التشتت بين النقاط التي تمثل الدرجات حول المتوسط العام (المستوى)، الأمر الذي يوضح مدى تجانس أو تباين أداء الفرد عبر الاختبارات المختلفة.

3- بينما يعبر الشكل عن النمط العام للاستعدادات، ويظهر من خلال ترتيب الدرجات التي حصل عليها الفرد في الاختبارات، حيث يُبرز بوضوح مناطق القوة ومواطن الضعف (علام، 2000، ص259).

ومن الجدير بالذكر أن هذه الخصائص قد تتشابه أو تختلف فيما بينها فقد يتفق بروفيان نفسيان لفردين في خاصيتي المستوى والانتشار، ولكنهما يختلفان في الشكل أي في صورة الترتيب النسبي لدرجات الاختبارات الفرعية، مما يمنح كل بروفيان طابعاً مميزاً.

وتكمن أهمية هذه الخصائص في المجالين النفسي والتربوي في أنها تتيح فهماً معمقاً لبنية القدرات الفردية؛ إذ تساعد الأخصائي النفسي أو التربوي على تحديد مستوى الأداء العام والتعرف على مدى تجانس أو تباين القدرات وفهم نمط الاستعدادات المميز لكل فرد. وهذا بدوره يوفر أساساً علمياً لاتخاذ قرارات دقيقة في التوجيه الأكاديمي والمهني كاختيار المسارات التعليمية الملائمة أو تحديد المجالات التي تتطلب دعماً إضافياً لتنمية مهارات الطالب وتعزيز فرص نجاحه.

8- تصنيف الاختبارات التحصيلية:

حرص علماء القياس والتقويم التربوي على تطوير أساليب متنوعة لقياس مدى تحقق الأهداف التعليمية، وتقييم فعالية البرامج التربوية المقدمة، إضافةً إلى قياس مستوى التحصيل الدراسي للتلميذ والحكم عليه وفق أسس محددة. ومن هذا الاهتمام ظهرت أنماط متعددة في بناء الاختبارات التحصيلية، تُصنّف تبعاً للغرض منها وطريقة تفسير نتائجها، حيث يرتبط ذلك بكيفية تحديد معنى الدرجة التي يحصل عليها الفرد. ويبرز من بين هذه الأنماط اختباران رئيسيان، هما:

8-1- الاختبارات مرجعية المعيار (المرجعية الجماعية) NRT أو الاختبارات السيكومترية:

والهدف من هذه الاختبارات هو تحديد مقدار المعلومات والمهارات التي سبق وأن حصل عليها التلميذ في موضوع معين. وتفسر درجة التلامذة في ضوء متوسط أداء الجماعة التي ينتمي إليها للحصول على تقدير لتحصيله الدراسي. وقد وجه إلى هذا النوع من الاختبارات الانتقاد بأنها لا

تقيس الأهداف التعليمية التي يسعى المعلم إلى تحقيقها في حجرة الدراسة، كما أنها لا تفيد في تشخيص جوانب القوة والضعف في التحصيل الدراسي للتلميذ.

8-2- الاختبارات مرجعية المحك (المرجعة للهدف) CRT أو الاختبارات الإيديومترية:

يستخدم مصطلح "الاختبار محكي المرجع" والذي شاع بفضل غلاسر (1963) Glaser، أحياناً كمترادف لمفاهيم أخرى مثل، الاختبار المرجعي للمجال، أو المحتوى، أو الأهداف، أو اختبار الكفاءة. ويرجع هذا الالتباس جزئياً إلى أن تفسير الاختبارات محكية المرجع والذي يعتمد على مجموعتين أساسيتين من المعايير:

أ- المعايير القائمة على مقدار المعرفة في مجال محتوى محدد كما يظهر في الاختبارات الموضوعية المقننة.

ب- المعايير القائمة على مستوى الكفاءة في مجال مهاري، كما يتضح من جودة الأداء الفعلي أو المنتج الناتج عن تطبيق المهارة (Urbina, 2004, p.105).

وأحياناً يُستخدم المصطلح أيضاً للإشارة إلى التفسيرات المبنية على العلاقة المحددة مسبقاً بين درجات الاختبار والمستويات المتوقعة للأداء على محك معين، مثل نشاط مستقبلي أو حتى اختبار آخر. في هذا الاستخدام يكون "المحك" نتيجة محددة قد ترتبط أو لا ترتبط بالمهام التي يقيسها الاختبار. وهذا يختلف بشكل واضح عن الاختبارات المحكية المحتوى أو الأداء حيث تكون المهام المقاسة عينات مباشرة من السلوك المرتبط بالمحك.

إلى جانب ذلك فإن طريقة استخدام الاختبارات محكية المرجع تختلف باختلاف الموقف. ففي بعض الأحيان تكون المحكات كمية بحتة، مثل اشتراط تحقيق نسبة محددة من الإجابات الصحيحة (80% أو 90%) لإثبات الإتقان. وفي أحيان أخرى تكون المحكات نوعية أو ذات طبيعة تقديرية أكثر. كما قد تُقيّم النتائج على أساس "تحقق الإتقان أو عدمه" بشكل مطلق، أو قد يُسمح بتحديد مستويات وسيطة من الكفاءة. ورغم اختلاف المصطلحات وطرق التركيز بين هذه الاختبارات فإنها تشترك في خصائص أساسية، أهمها:

أ- أنها تُستخدم لقياس مدى إتقان الأفراد لمهارات أو معارف محددة.

ب- تصحيحها لا يعتمد على مقارنة أداء شخص بآخر بخلاف الاختبارات مرجعية الجماعة التي تهدف إلى ترتيب الأفراد أو تصنيفهم بالنسبة لبعضهم البعض (Urbina, 2004, p.105).

والهدف من هذه الاختبارات هو التعرف على مستوى التمكن أو الإتقان (Mastery) في الأداء، وبالتالي فإن درجة التلميذ تفسر في ضوء محك أو درجة قطع (Cut-off score) سابقة التحديد وليس

متوسط أداء الجماعة التي ينتمي إليها وبالتالي فهي تحدد مستويات الأداء المطلوب تحقيقها (علام، 2000).

- أوجه الاختلاف بين الاختبارات التحصيلية مرجعية المحك ومرجعية المعيار:

يوجد عديد الاختلافات بين الاختبارات مرجعية المعيار والاختبارات مرجعية المحك والتي يمكن أن تقدم على النحو التالي:

أ. من حيث الهدف الرئيس للاختبار: يهدف الاختبار مرجعي المعيار إلى مقارنة أداء التلميذ بأداء مجموعته المعيارية من خلال تحديد عدد الأسئلة التي أجاب عليها التلميذ بشكل صحيح، أما الاختبار مرجعي المحك يهدف إلى مقارنة أداء التلميذ بمحك محدد مسبقاً وذلك لمعرفة الأهداف التي حققها والتي لم يحققها التلميذ.

ب. من حيث استخدام الاختبار: يستخدم الاختبار مرجعي المعيار لإظهار الفروق الفردية بين التلامذة، أما الاختبار مرجعي المحك يستخدم للتأكد من تحقيق التلميذ للأهداف السلوكية المحددة.

ج. من حيث خصائص أسئلة الاختبار: تنتشر الأسئلة في الاختبار مرجعي المعيار بشكل واسع حول نطاق الأهداف، بينما تتجمع الأسئلة في الاختبار مرجعي المحك حول عدد محدد من الأهداف.

د. من حيث تفسير الأداء: تفسر درجة التلميذ في الاختبار مرجعي المعيار بناءً على درجات معيارية، بينما تفسر الدرجة في الاختبار مرجعي المحك بناءً على درجة قطع معينة.

هـ. من حيث بناء المفردات: يعتمد بناء المفردات في الاختبار مرجعي المعيار على تباين الدرجات ومعامل التمييز، بينما يعتمد بناء المفردات في الاختبار مرجعي المحك على مدى تحقيق المفردة للهدف السلوكي.

جدول 04 يبين أهم الفروق بين الاختبارات مرجعية الجماعة والاختبارات محكية المرجع.

العنصر	الاختبارات مرجعية الجماعة (Norm-Referenced)	الاختبارات محكية المرجع (Criterion-Referenced)
الهدف	معرفة موقع الفرد مقارنة بالآخرين.	معرفة إذا أتقن الفرد مهارة أو حقق معيار (محك) معين.
تفسير الدرجة	مقارنة الأداء بدرجات مجموعة معيارية (العينة).	مقارنة الأداء بمعيار (محك) ثابت أو حد نجاح معين.
التركيز	الموقع النسبي داخل المجموعة	المستوى المطلق للأداء.
مثال	اختبار ذكاء، قبول جامعي، ترتيب بالمئينيات.	اختبار قيادة، اختبار إملاء (نجاح $\geq 80\%$).
طريقة التحليل	الترتيب والتصنيف النسبي بين الأفراد.	تحقيق الأهداف أو تجاوز معيار (محك) محدد.

9- مستويات القياس التربوي والنفسي (Levels of Educational and Psychological)

:(Measurement)

اقترح ستيفنز (1946) Stevens في إطار تصوره للقياس وتصنيف مقاييسه، والذي يُعرف أيضًا بمستويات القياس، مفهوم مقاييس القياس والتي يمكن تصنيفها بناءً على نوع التحويلات الرياضية التي يمكن إجراؤها عليها دون أن تفقد خصائصها الأساسية. وقد حدد أربعة أنواع رئيسية من المقاييس: الاسمي، والرتبي، والفتري، والتسبي، ولكل من هذه المقاييس خاصية مميزة من حيث النوع، والترتيب، وتساوي الفترات، ونسب القيم الفردية على التوالي. يمكن إخضاع كل مقياس لعمليات تحويل شريطة أن تحافظ هذه التحويلات على خصائصه، وهي السمة التي أطلق عليها ستيفنز اسم الثبات (invariance) أي بقاء خصائص المقياس دون تغيير بعد التحويل. وتُعتبر المقاييس الاسمية الأقل تقييدًا نظرًا لإمكانية تطبيق تحويلات متنوعة عليها مع الاحتفاظ بخصائصها بينما تعد المقاييس النسبية الأكثر تقييدًا إذ لا يمكن تحويلها إلا بطرق محدودة دون المساس بخصائصها (DiIorio, 2005, p.9).

9-1- المقياس الاسمي (Nominal Scale):

يُستخدم المقياس الاسمي في تصنيف المتغيرات إلى فئات استنادًا إلى مبدأ التكافؤ بين عناصر كل فئة. ومن بين الخصائص الأربع للمقاييس يتميز هذا المستوى بخاصية النوع فقط أي أن الفئات تختلف فيما بينها من حيث النوع وليس الكم. لا يتضمن المقياس الاسمي ترتيبًا ذا معنى وبالتالي فإن الانتماء إلى فئة معينة لا يعكس امتلاك مقدار أكبر أو أقل من المتغير. كما أن الفترات بين القيم ليست متساوية والنسب بين القيم الفردية لا تحمل دلالة كمية. على سبيل المثال يُعد النوع الاجتماعي (الجنس) متغيرًا اسميًا يتكوّن من فئتين: ذكر وأنثى. جميع الأفراد في الفئة الواحدة متماثلون في هذا المتغير لكنهم يختلفون عن أفراد الفئة الأخرى. ومن الأمثلة الأخرى الحالة الاجتماعية التي قد تشمل فئات مثل: متزوج، أعزب، أرمل، أو مطلق. الأشخاص ضمن الفئة نفسها يشتركون في نفس الصفة بينما يختلفون عن الأشخاص في الفئات الأخرى. بما أن فئات المقياس الاسمي تختلف في النوع فقط وليس في الدرجة أو الكمية فإنه يمكن ترميز كل فئة بأي رقم شريطة أن يختلف عن أرقام الفئات الأخرى. فمثلاً يمكن تمثيل الذكور بالرقم 0 والإناث بالرقم 1، أو العكس. وبما أن هذه الأرقام لا تحمل أي دلالة كمية يمكن أيضًا ترميز الذكور بالرقم 1003 والإناث بالرقم 62 دون أن يؤثر ذلك على طبيعة المقياس. ومع ذلك جرت العادة على استخدام الأرقام 0 و1 أو 1 و2 عند ترميز متغير الجنس. يُعد المقياس الاسمي الأقل تقييدًا فيما يتعلق بـ **محك الثبات (invariant)**

(criterion) إذ تظل خاصية اختلاف الفئات في النوع قائمة بغض النظر عن الأرقام المختارة بينما لا تنطبق الخصائص الأخرى للمقاييس على هذا النوع من المتغيرات. وبذلك لا يوجد ترتيب ذو معنى يمكن أن يوحي بأن أحد الجنسين "يمتلك" هذه الصفة أكثر من الآخر (DiIorio, 2005, p.9).

9-2- المقياس الرتبي (Ordinal):

يُستخدم المقياس الترتيبي لقياس المتغيرات التي تمتلك فئات مرتبة ترتيباً ذا معنى وبشكل هرمي. على سبيل المثال يُعد المستوى الدراسي الجامعي مثلاً متغيراً ترتيبياً، حيث تُصنّف الفئات إلى: سنة أولى، سنة ثانية، سنة ثالثة، وسنة رابعة. تمثل هذه الفئات تدرجاً في عدد الساعات المعتمدة المنجزة فطلاب السنة الأولى لديهم أقل عدد من الساعات، بينما يمتلك طلاب السنة الرابعة أكبر عدد منها. مع ذلك قد يوجد اختلاف في السمات داخل الفئة الواحدة؛ ففئة السنة الأولى قد تضم طلاباً في بداية دراستهم لم يكملوا أي ساعات وآخرين أنجزوا الحد الأقصى المسموح به في هذه المرحلة. الأمر نفسه ينطبق على السنة الثانية التي قد تشمل طلاباً بأعداد متفاوتة من الساعات. ومن الأمثلة الأخرى تصنيف درجات لغط القلب أو النفخة القلبية (heart murmur)، حيث تُرتب الحالات من الدرجة 1 حتى الدرجة 6 تبعاً لحدة الحالة. ويظهر هذا التدرج أن كلما ارتفعت الدرجة زاد مستوى الخلل الوظيفي في القلب لكن الأرقام لا تمثل مقدار الزيادة الفعلي في شدة الحالة بل تشير فقط إلى ترتيبها (DiIorio, 2005, p.10).

كما تُصنّف مقاييس التقدير التي تتضمن استجابات مثل: "أوافق بشدة"، "أوافق"، "لا أوافق ولا أعارض"، "أعارض"، و"أعارض بشدة" ضمن المقاييس الترتيبية (Ordinal Scales)، حيث تكون فئات المقياس منفصلة وتختلف في الدرجة ويتم ترتيبها بحيث تدل استجابة "أوافق" على اتجاه أكثر إيجابية من "أعارض". ومع ذلك قد يختلف شعور الأفراد داخل نفس الفئة فالشخص الذي يختار "أوافق" قد يكون أقرب في رأيه إلى "أوافق بشدة" أو أقرب إلى "لا أوافق ولا أعارض". الأرقام المخصصة لهذه الفئات تعكس فقط ترتيبها ولا تمثل الفروق الدقيقة في الاتجاه بينها. تحافظ المقاييس الترتيبية على ترتيب الفئات حتى عند إجراء تحويلات عددية كإضافة قيمة ثابتة لجميع الأرقام أو ضربها في ثابت. فعلى سبيل المثال يمكن ترميز سنوات الدراسة الجامعية بالأرقام 1، 2، 3، و4 لتمثيل السنوات من الأولى إلى الرابعة مع ضرورة أن يعكس الترميز ترتيب السنوات دائماً؛ أي أن الرقم 1 يجب أن يخصص للسنة الأولى، والرقم 4 للسنة الرابعة. كما هو موضح في الشكل 1.1، إذا أضفنا رقم 18 إلى كل رقم أو ضربنا كل رقم في 44، ستتغير القيم العددية الفعلية لكن ترتيبها سيبقى كما هو. وبالتالي فإن أي تحويل يحافظ على هذا الترتيب الأصلي يكون مقبولاً.

$$(1 + 18), (2 + 18), (3 + 18), (4 + 18) = 19, 20, 21, 22$$

$$(1 \times 44), (2 \times 44), (3 \times 44), (4 \times 44) = 44, 88, 132, 176$$

وفي حالة تحديد مقاييس النزعة المركزية في البيانات الترتيبية يُفضَّل استخدام المنوال أو الوسيط إذ لا يمكن تعريف المتوسط الحسابي في مثل هذه البيانات.

المستوى الرتبي يتميز بخاصيتي التمايز (Identity) والترتيب (Magnitude). فكل قيمة على هذا المقياس تحمل معنى وحيد كما أن القيم مرتبة في علاقة متسلسلة فيما بينها. ومن الأمثلة الكلاسيكية على ذلك نتائج سباق الخيل والتي تُعرض عادةً بصيغة الفائز الأول، الثاني وهكذا. من خلال هذا المقياس يمكننا معرفة ترتيب المتسابقين. ومع ذلك فإن المقياس الرتبي لا يقدّم معلومات حول المسافة أو الفروق بين هذه المراتب فلا يمكننا أن نعرف إن كان السباق متقارباً أم أن المتسابق الفائز تقدّم بفارق كبير (Sivakumar, et al, 2019, p.12).

3-9- المقياس الفتري (Interval Scale):

المتغيرات المقاسة على مقياس الفترات (Interval Scale) تتميز بوجود ترتيب ذي معنى بحيث تمثل القيم الأعلى مستويات أكبر من الخاصية المقاسة. على عكس المقياس الاسمي والترتبي فإن الأرقام المستخدمة هنا ذات دلالة كمية فهي تعكس مقدار الخاصية وتزودنا بمعلومات عن كميتها. يمتاز هذا المقياس أيضاً بتساوي الفئات أي أن المسافات بين الوحدات المتتالية للمقياس متكافئة. فمصطلح "الفترات" يشير إلى "المسافات الفاصلة بين القيم" وهو ما يميز هذا النوع من المقاييس، إذ لا يقتصر على تحديد الترتيب فقط بل يُبين أيضاً قيمة الفارق بين كل عنصر وآخر. لكن مقياس الفترات يفتقر إلى نقطة صفر حقيقية (True Zero) تمثل انعدام الخاصية تماماً لذلك لا يمكن تفسير النسب بين القيم الفردية بشكل صحيح بينما تبقى النسب بين الفروق (Ratios of Differences) قابلة للتفسير. على سبيل المثال في قياس درجة الحرارة بالفهرنهايت درجة حرارة 100 أكبر من 60، والفارق بينهما (40 درجة) يساوي الفارق بين 70 و30 درجة (40 درجة أيضاً). كذلك يمكن القول إن الفارق بين 100 و60 درجة (40 درجة) يساوي ضعف الفارق بين 50 و30 درجة (20 درجة). لكن لا يمكن القول إن 60 درجة تساوي ضعف حرارة 30 درجة، لأن المقياس لا يملك صفراً حقيقياً (DiIorio, 2005, p.12). ومن أمثلة ذلك مقياس سلسيوس الذي لا يعكس وجود نقطة صفر مطلقة.

شكل 05 يوضح مثال على المسافات المتساوية



المثال الثاني افترض أن الباحث يعرف أن متوسط وزن المواليد في مجموعة معينة من الأطفال هو 7 كيلوغرام. ويرغب الباحث في تحديد مقدار انحراف كل طفل عن متوسط الوزن. إذا أعاد الباحث معايرة الميزان بحيث يكون متوسط الوزن (7 كيلوغرام) مساوياً للصفر فإن الوزن 7 كيلوغرام يُرمز بـ 0. والطفل الذي يزن 8 كيلوغرام يحصل على درجة +1 والذي يزن 9 كيلوغرام يحصل على درجة +2 وهكذا. أما الطفل الذي يزن 6 كيلوغرام فيحصل على درجة -1 والذي يزن 5 كيلوغرام يحصل على درجة -2 (Urbina, 2004, p.37-41).

باستخدام مقياس الفترات الجديد يمكننا القول إن الفرق في الوزن بين -2 و 0 هو 2 كيلوغرام يعادل الفرق في الوزن بين +2 و +4 وهو كذلك 2 كيلوغرام. لكن لا يمكن القول إن الطفل الذي حصل على درجة +4 والذي وزنه 11 كيلوغرام يزن ضعف الطفل الذي حصل على درجة +2 والذي وزنه 9 كيلوغرام.

يُنظر إلى المقياس الفترتي على أنه يمثل قياساً مستمراً بطبيعته في حين أن فئات المتغيرات المقاسة بالمقياس الرتبي تكون مقسمة إلى مجموعات منفصلة، مثل: (السنة الأولى، الثانية، الثالثة، الرابعة). ورغم إمكانية تحويل المتغيرات الفترية إلى فئات فإن تفسير هذه الفئات يجب أن يراعي سياق التحويل. فعلى سبيل المثال تصنيف درجات الحرارة إلى منخفضة، متوسطة، أو مرتفعة (مقياس رتبي) يعتمد على الموقف إذ يختلف هذا التصنيف عند قياس حرارة جسم الإنسان عنه عند قياس درجة حرارة الهواء. كما أن درجة الحرارة قد تُعرض بأرقام صحيحة لكن يمكن أيضاً قياسها بجزء من الدرجة بل وفي بعض الحالات يكون من المنطقي قياسها حتى العُشر من الدرجة أو جزء من المئة منها. يتيح المقياس المستمر مثل هذه الدقة في التمييز بينما لا يتيحها المقياس الرتبي. ونظراً لأن المقياس الفترتي يعتمد على نقطة صفر اعتباطي (غير حقيقي) فإن القيم السالبة تمثل كميات

حقيقية من الخاصية المقاسة ولها معنى. فعلى سبيل المثال درجة الصفر فهرنهايت ليست خلواً من الحرارة بل تظل هناك حركة جزيئية (حرارة) يمكن قياسها وقد سُجّلت درجات حرارة هواء منخفضة تصل إلى -50 فهرنهايت. وبالمثل إذا قام الباحث بترميز أوزان الرضع مستخدماً متوسط الوزن كنقطة صفر فإن الدرجة -3 تعكس قيمة حقيقية للوزن ولها مدلول واضح (DiIorio, 2005, p.13).
تكمّن أهمية مقاييس الفترات في أنها تسمح باستخدام مجموعة واسعة من التحليلات الإحصائية مثل مقاييس النزعة المركزية كالوسيط، المنوال، أو المتوسط الحسابي، وقياس التشتت كالانحراف المعياري، وبالتالي فإن مقاييس الفترات مفيدة جداً في التحليل الإحصائي لكنها محدودة من حيث عدم إمكانية حساب النسبة المئوية وهو ما يقودنا إلى النوع الأخير من مقاييس القياس.
يتميز مقياس الفترات بخصائص التمايز (Identity)، والترتيب (Magnitude)، والتساوي بين الفترات (Equal Intervals). ومن أبرز الأمثلة عليه مقياس فهرنهايت المستخدم لقياس درجة الحرارة. فهذا المقياس يتكوّن من وحدات متساوية بحيث يكون الفرق بين 40 و 50 درجة فهرنهايت مساوياً تماماً للفرق بين 50 و 60 درجة فهرنهايت (Sivakumar, et al, 2019, p.13).

9-4- المقياس النسبي (Ratio Scale):

المقياس النسبي (Ratio Scale) يتميز بجميع خصائص المقياس الفترتي (Interval Scale) لكنه يختلف عنه في وجود نقطة صفر حقيقي وليس افتراضي. تمثل هذه النقطة غياب الصفة أو الخاصية المقاسة تماماً ما يسمح بالتعبير عن النسب بشكل ذي معنى. فكما هو الحال في المقياس الفترتي يمكن القول إن الفرق بين القيمتين 2 و 4 يساوي الفرق بين 4 و 6 (أي نسبة الفروق). لكن على عكس المقياس الفترتي يمكننا أيضاً القول إن القيمة 8 تساوي ضعف القيمة 4 (أي نسبة القيم نفسها). يُقاس كلٌّ من الطول والوزن على مقياس نسبي إذ لا يبقى شيء لقياسه إذا كان الطول أو الوزن يساوي صفراً. فعلى سبيل المثال الشخص الذي يكون ضغط دمه صفراً لن يبقى على قيد الحياة دون تدخل طبي والجزيئات تتوقف عن الحركة وإنتاج الحرارة عند درجة حرارة -460 فهرنهايت (أي صفر كلفن، أو الحرارة المطلقة). وبما أن الوزن له نقطة صفر حقيقية فإن طفلاً يزن 10 كيلو غرام يكون ضعف وزن طفل آخر يزن 5 كيلو غرام. ويعرض الشكل (1.3) مخططاً يوضح الفروق بين المقاييس الأربعة للقياس (DiIorio, 2005, p.14).

الشكل 06 يوضح مخطط لمستويات القياس.

Nominal
2 5 6 1 4 3
Ordinal
1-----2-----3-4-----5-----6-----
Interval
-1-----0-----1-----2-----3-----4-----5-----
Ratio
0-----1-----2-----3-----4-----5-----6-----

شكل 07 يوضح هذا الجهاز لمثالين على مقاييس النسبة، وهما: الطول والوزن.



المقياس النسبي يوفر ثراءً من الإمكانيات في مجال التحليل الإحصائي. إذ يمكن جمع هذه المتغيرات وطرحها وضربها وقسمتها (النسب) بشكل ذي معنى. كما يمكن مقاييس النزعة المركزية باستخدام المنوال أو الوسيط أو المتوسط الحسابي؛ بالإضافة إلى إمكانية حساب مقاييس التشتت، مثل: الانحراف المعياري ومعامل الاختلاف من خلال المقاييس النسبية.

كما يحقق الخصائص الأربع جميعها: التمايز (Identity)، الترتيب (Magnitude)، التساوي بين الفترات (Equal intervals)، ووجود قيمة صغرى تساوي الصفر (True zero) (Sivakumar, et al, 2019, p.14).

10- خصائص مستويات القياس:

كل مستوى من مستويات القياس يحقق واحدة أو أكثر من الخصائص الآتية:

- 1- التمايز (Identity): لكل قيمة في كل مستوى قياس معنى وحيد ودقيق.
- 2- الترتيب (Magnitude): القيم في كل مستوى له علاقة الترتيب مع بعضه البعض؛ أي أن بعض القيم أكبر وأخرى أصغر.
- 3- الفترات المتساوية (Equal intervals): الوحدات على المقياس متساوية فيما بينها. على سبيل المثال، الفرق بين 1 و 2 مساوٍ للفرق بين 19 و 20.
- 4- الصفر الحقيقي (True zero): يحتوي المقياس على نقطة صفر حقيقية، لا توجد قيم أقل منها (Sivakumar, et al, 2019, p.10).

11- مستويات القياس والإحصاء:

وسّع ستيفنز (1959) تصنيفه لمستويات المتغيرات ليشمل العمليات الإحصائية المسموح استخدامها مع كل مستوى قياس. وبوجه عام توجد علاقة عكسية بين التحويلات المسموح بها والعمليات الإحصائية الممكنة فكلما قلّت القيود على التحويلات زادت القيود على الإجراءات الإحصائية والعكس صحيح. على سبيل المثال المتغيرات المقاسة بالمقياس الاسمي رغم أنها تسمح بتحويلات عديدة إلا أن إمكانيات تحليلها إحصائياً محدودة. ففي هذه الحالة يمكن الاكتفاء بوصف البيانات باستخدام التكرارات والنسب المئوية واعتماد المنوال كمؤشر للنزعة المركزية. أما الاختبارات الإحصائية الاستدلالية فتقتصر غالباً على اختبار كاي تربيع أو ما يشابهه. فعلى سبيل المثال يمكن تحديد عدد الرجال والنساء في عينة الدراسة وحساب نسبتهم المئوية لمعرفة الفئة الأكثر عدداً (المنوال)، كما يمكن مقارنة المتغيرات من خلال الجداول المتقاطعة وتحليلها باستخدام اختبار كاي تربيع لمعرفة ما إذا كانت نسبة النساء المتزوجات مثلاً أعلى من نسبة الرجال المتزوجين. وتقريباً نفس الشيء بالنسبة للمتغيرات المقاسة على المقياس الاسمي فإن إمكانيات المعالجة الإحصائية للمتغيرات على المستوى الترتيبي تبقى محدودة. ففي هذا المستوى يمكن حساب المنوال، والوسيط، والنسب المئوية، ومعاملات الارتباط الرتبي، إضافة إلى استخدام الإحصاءات اللابارامترية. أما في حالة المتغيرات المقاسة على المقياس الفكري فيتاح للباحث حساب المتوسط

الحسابي، والانحراف المعياري، ومعامل الارتباط بيرسون، وإجراء الاختبارات البارامترية كتحليل التباين (ANOVA). أما بالنسبة للمستوى النسبي فيمكن تطبيق جميع الأساليب الإحصائية تقريباً بما في ذلك حساب المتوسط الهندسي، والمتوسط التوافقي، ونسبة التباين، فضلاً عن الأساليب الإحصائية المتقدمة كتحليل الانحدار والنمذجة بالمعادلات البنائية (SEM).

ومن المهم أن يعلم الطالب أن بعض الاختبارات الإحصائية الاستدلالية، كاختبار (T) وتحليل التباين (ANOVA)، تتطلب وجود متغير مستقل تصنيفي ومتغير تابع مستمر (متصل). وفي حال كان المتغير المستقل مستمراً يمكن تحويله إلى متغير تصنيفي عن طريق التجميع كتقسيم الأفراد وفق أوزانهم إلى فئة تقل عن 100 كيلو غرام وأخرى أوزانهم أكثر من 100 كيلو غرام (DiIorio, 2005, p.15).

الفصل الثالث

الأهداف التعليمية أنواعها ومستوياتها

الأهداف (Objectives):

يهدف هذا الفصل إلى تعريف الطالب بمفهوم الأهداف التعليمية وأنواعها ومستوياتها، مع التركيز على مجالاتها الثلاثة: المعرفي، والانفعالي، والنفسي-حركي، بما يتيح له التمييز بين طبيعة كل مجال ومتطلباته. كما يهدف الفصل إلى إكساب الطالب القدرة على صياغة الأهداف السلوكية بشكل صحيح وفق شروط علمية دقيقة، وتمكينه من مهارة تحليل محتوى المادة الدراسية، وصولاً إلى إعداد جدول المواصفات الذي يربط بين الأهداف والمحتوى وأدوات القياس، بما يساهم في بناء اختبارات موضوعية وفعّالة.

المحتوى (Contents):

- 1- الأهداف التعليمية: أنواعها ومستوياتها
- 2- مستويات الأهداف التعليمية
- 1-2- المجال المعرفي أو الإدراكي (العقلي) The Cognitive Domain
- 2-2- المجال الانفعالي أو الوجداني (العاطفي) The Affective Domain
- 3-2- المجال النفسي حركي أو المهاري The Psychomotor Domain
- 3- شروط صياغة الأهداف السلوكية
- 4- تحليل محتوى المادة الدراسية
- 5- إعداد جدول المواصفات

تمهيد:

- ماذا نريد أن يتعلم الطلاب؟

قد يبدو هذا السؤال بسيطاً للوهلة الأولى لكنه في الحقيقة يمثل حجر الزاوية لكل عملية تعليمية ناجحة. فقبل أن نختار الأساليب التعليمية أو نصمم أدوات التقييم علينا أن نعرف بالضبط ما نريد أن يحققه طلابنا بنهاية الدرس أو الوحدة. بدون وضوح في هذه الرؤية تصبح كل جهودنا في التدريس والتقويم عشوائية وتفقد فعاليتها. لقد استخدم الباحثون والممارسون التربويون مصطلحات مختلفة لوصف هذه الغايات التعليمية، مثل: "مخرجات التعلم"، "نواتج التعلم"، "أهداف الدرس"، "توقعات التعلم"، أو "معايير الأداء". ورغم تعدد هذه المسميات فإن المعنى الجوهر هو نفسه أي تحديد ما نرغب أن يعرفه الطالب أو يستطيع القيام به عند انتهاء العملية التعليمية. تخيل على سبيل المثال، درساً في العلوم عن دورة الماء. إذا كان هدفك غير محدد قد يقضي الطلاب وقتاً في حفظ تفاصيل غير أساسية بينما الهدف الواضح – مثل قدرة الطالب على وصف مراحل دورة الماء وربطها بالطقس المحلي – يوجّه تعلمهم ويجعل الدرس أكثر تأثيراً وارتباطاً بالحياة الواقعية.

يشير (Stiggins et al (2007 إلى أن تحديد أهداف التعلم بشكل واضح ومحدد ليس ترفاً أكاديمياً بل هو شرط أساسي لنجاح العملية التعليمية. فهو يوجّه كل قرار تربوي من اختيار أسلوب التدريس إلى تصميم أدوات التقييم ويضمن أن يكون التعلم فعالاً والقياس دقيقاً والنتائج متوافقة مع الغايات المنشودة. باختصار معرفة ما نريد للطلاب أن يتعلموه يمنح العملية التعليمية اتجاهاً واضحاً ويساعد المعلمين على تصميم تجارب تعلم غنية تفاعلية ومثمرة تجعل الطلاب محور العملية التعليمية ويحققون التعلم المتوقع بنجاح.

1- الأهداف التعليمية: أنواعها ومستوياتها

عندما نشرع في إعداد أي اختبار تحصيلي أو تصميم منهج دراسي فإن الخطوة الأولى التي لا غنى عنها هي تحديد الأهداف التعليمية المراد تحقيقها. فالأهداف هي البوصلة التي توجه مسار العملية التعليمية من بدايتها حتى نهايتها. وقد عرّف "مايجر" (Robert F. Mager) الهدف التعليمي بأنه "رغبة في إحداث تغيير متوقع في سلوك المتعلم أو وصف دقيق للسلوك الذي نتوقع من التلميذ أن يكون قادراً على أدائه بعد إتمام عملية التعلم".

ورغم أن صياغة الأهداف بدقة قد تبدو مهمة معقدة – كما تشير بعض الدراسات في مجال القياس والتقويم – إلا أنها تصبح أكثر يسراً عندما تُربط بالنتائج المتوقعة من عملية التعليم. فالأهداف

في جوهرها ليست سوى مجموعة التوقعات التي نصوغها حول ما ينبغي أن يكتسبه الطالب من معرفة أو مهارة أو اتجاه بعد أن يمر بخبرة تعليمية محددة (Kumar, et al, 2016, p.20).

وتكمن أهمية الأهداف التعليمية في أنها تمثل الأساس والمعيار الذي تُبنى عليه الاختبارات التحصيلية بل وجميع مكونات العملية التعليمية. فاختيار المحتوى الدراسي على سبيل المثال يتأثر بشكل مباشر بنوعية الأهداف الموضوعية؛ فمحتوى يركّز على المهارات العملية يختلف عن محتوى يهدف إلى ترسيخ المعارف النظرية أو القيم والاتجاهات.

إذ، فإن السؤال الجوهرى الذي يفرض نفسه هنا هو: ما هو الهدف التعليمي؟ وما أنواعه ومستوياته التي ينبغي للمعلم والمصمم التربوي أن يضعها في اعتباره عند التخطيط للعملية التعليمية والتقويمية؟

1-1- مفهوم الهدف:

يشكل الهدف التربوي المرتكز الأساسي لكل عملية تعليمية، فهو البوصلة التي تحدد الاتجاه الذي ينبغي أن تسلكه الممارسات التدريسية والتقويمية. ويمكن تعريفه بأنه وصف للتغير المتوقع حدوثه في سلوك المتعلم نتيجة تزويده بخبرات تعليمية وتفاعله مع مواقف تعليمية محددة. إن الأهداف التربوية لا تقتصر على كونها عبارات مكتوبة، بل تمثل تصورات واضحة لما نطمح أن يبلغه المتعلم بعد مروره بخبرة تعليمية معينة.

ولتحقيق التوازن والشمولية درج التربويون على تصنيف الأهداف التربوية إلى ثلاثة مستويات أساسية تتكامل فيما بينها:

- المستوى الأول: الأهداف التربوية العامة

هي الأهداف الأكثر شمولاً وتجريداً، إذ تعبّر عن الغايات الكبرى للتربية والتعليم. وتركّز هذه الأهداف على التغيرات العميقة والمستديمة في شخصية المتعلم وعلى ما يُفترض أن يتعلمه بشكل شامل على المدى الطويل. وغالباً ما ترتبط هذه الأهداف بفلسفة الدولة التعليمية وخصائص المجتمع وثقافته كإعداد المواطن الصالح، بناء الإنسان المؤمن المسؤول، تنمية روح المواطنة والانتماء، أو تعزيز التفكير النقدي والإبداعي.

- المستوى الثاني: الأهداف التعليمية

تمثل هذه الأهداف درجة أقل من التجريد وأكثر تخصصاً من المستوى الأول إذ تُعنى بالإنجازات التعليمية التي يُتوقع أن تتحقق خلال فترة زمنية محددة كفصل دراسي أو مرحلة تعليمية معينة.

وتشمل هذه الأهداف ما يُتوقع من الطلاب تعلمه في مواد دراسية محددة كمعرفة فروع الرياضيات المختلفة، القدرة على تطبيق القوانين الفيزيائية، أو التمكن من قواعد النحو الأساسية في اللغة العربية.

- المستوى الثالث: الأهداف السلوكية المحددة

تُعَدّ الأهداف السلوكية أدق مستويات الأهداف وأكثرها تحديداً فهي تصف سلوكيات أو استجابات ملموسة يمكن ملاحظتها وقياسها لدى الطالب سواء كانت عقلية (معرفية)، أو حركية (مهارية)، أو انفعالية (وجدانية). ويُعَدّ صياغة هذه الأهداف من المهام الجوهرية للمعلم إذ تمثل الأساس الذي تُبنى عليه استراتيجيات التدريس وأدوات التقويم. ومن أمثلة ذلك: أن يعدد الطالب خطوات حل المعادلة من الدرجة الثانية أو أن يكتب جملة صحيحة مستخدماً أداة الشرط أو أن يؤدي مهارة تمرير الكرة في كرة السلة بدقة أو أن يظهر احتراماً لآراء زملائه أثناء المناقشة الصفية.

وبهذا التقسيم نلاحظ أن الأهداف التربوية تبدأ من الغايات الكبرى ذات الطابع الفلسفي والمجتمعي مروراً بالأهداف التعليمية الأكثر تخصصاً ووصولاً إلى الأهداف السلوكية الدقيقة التي تشكل الممارسات اليومية في الصف الدراسي لتتكامل جميعها في خدمة العملية التعليمية الشاملة.

جدول 05 يوضح ملخص الفروق بين المستويات الثلاثة للأهداف التربوية.

المستوى	الطابع	المدى الزمني	القابلية للقياس	أمثلة	الجهة المسؤولة عن الصياغة
الأهداف التربوية العامة	شمولية وتجريدية، تهتم بالغايات الكبرى	طويل المدى (المرحلة التعليمية أو مستوى المجتمع)	صعبة القياس المباشر، تُقاس عبر مؤشرات غير مباشرة	إعداد مواطن صالح، بناء إنسان مؤمن ومسؤول	السياسات التعليمية والمناهج الوطنية
الأهداف التعليمية	متخصصة وأقل تجريدًا، ترتبط بالمواد الدراسية	متوسط المدى (فصل دراسي أو مرحلة تعليمية)	قابلة للقياس جزئيًا، عادة عبر اختبارات تقويمية	معرفة فروع الرياضيات، تطبيق القوانين الفيزيائية	المعلم والمصمم التربوي
الأهداف السلوكية المحددة	دقيقة ومحددة، تركز على السلوكيات والمهارات	قصير المدى (درس أو نشاط محدد)	قابلة للقياس بسهولة ووضوح	حل معادلة من الدرجة الثانية، أداء مهارة رياضية، التعبير عن احترام زملاء الصف	المعلم مباشرة أثناء التخطيط اليومي للدرس

1-2- مكونات الهدف السلوكي:

يُعدّ الهدف السلوكي أدق مستويات الأهداف التعليمية إذ يصف سلوكيات أو استجابات محددة يمكن ملاحظتها وقياسها لدى الطلاب. ولكي تكون صياغة الهدف السلوكي واضحة وقابلة للتقويم يُنصح بالالتزام بالمكونات الأساسية التالية:

الهدف السلوكي = أن + فعل سلوكي + الطالب + محتوى علمي + شرط الأداء + معيار الأداء.

أ- **الفعل السلوكي:** يشير إلى السلوك أو المهارة التي يُتوقع أن يؤديها الطالب، مثل "يكتب"، "يقيس"، "يحلّ".

ب- **المحتوى العلمي:** يشير إلى المعرفة أو المهارة أو المفهوم الذي يتعلق به السلوك.

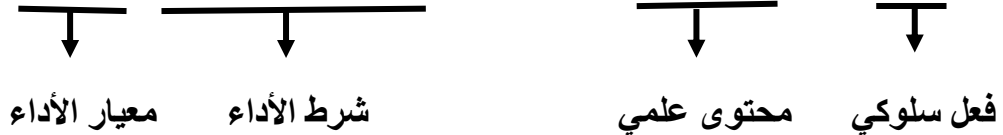
ج- **الطالب:** الفاعل الذي سيؤدي السلوك.

د- **شرط الأداء:** الظروف أو الوسائل التي يتم فيها تنفيذ السلوك، مثل "باستخدام المسطرة"، "دون الرجوع إلى الكتاب المدرسي".

هـ- **معيار الأداء:** المستوى أو الدقة التي يجب أن يصل إليها الطالب، مثل "كتابة صحيحة"، "قياساً دقيقاً".

- أمثلة لصياغة الأهداف السلوكية:

1- أن يبسط التلميذ تعبيراً جبرياً معقداً باستخدام قوانين الجمع والطرح دون أخطاء.



2- أن يكتب التلميذ تعريف ظاهرة التلوث بدون الرجوع إلى الكتاب المدرسي كتابة صحيحة.



3- أن يقيس التلميذ طول خارطة ما باستخدام المسطرة قياساً صحيحاً.



2- مستويات الأهداف التعليمية:

يُعدّ تصنيف الأهداف التربوية أو التعليمية من الأمور الجوهرية التي ينبغي لكل معلم معرفتها والإلمام بها فهو يمثل المفتاح الأساسي لاختيار الأساليب التدريسية الملائمة والوسائل التعليمية المناسبة ويضمن تحقيق أهداف كل درس بفعالية. فالمعلم الذي يفهم تصنيفات الأهداف يستطيع توجيه تعلم الطلاب بشكل أكثر دقة وتخطيط الأنشطة التعليمية بما يتوافق مع الغايات المرجوة. ومن بين أهم التصنيفات والأكثر شيوعاً واستخداماً في المجال التربوي يبرز تصنيف بلوم (Bloom) الذي يعتبر مرجعاً مهماً لتحديد الأهداف التربوية على مستويات متعددة وفي مجالات مختلفة. يقوم هذا التصنيف على افتراض أساسي مفاده أن نتائج التعلم يمكن وصفها في صورة تغييرات محددة وملموسة في سلوك الطلاب كما أنه يساعد المعلمين على صياغة الأهداف التعليمية في عبارات سلوكية واضحة وقابلة للقياس. وقد قسم بلوم الأهداف التعليمية إلى ثلاث مجالات رئيسية:

- 1- **المجال المعرفي (الإدراكي/العقلي):** يتعلق باكتساب المعرفة والمفاهيم والقدرة على التفكير والتحليل والتطبيق والاستنتاج.
 - 2- **المجال الانفعالي (العاطفي/الوجداني):** يهتم بالقيم والمواقف والاتجاهات التي يتبناها الطلاب تجاه موضوعات التعلم كالاهتمام، التقدير أو الالتزام.
 - 3- **المجال النفس حركي (المهاري):** يركز على تنمية المهارات الحركية والقدرات العملية للطلاب مثل تنفيذ المهام اليدوية أو أداء الحركات الرياضية بدقة.
- باختصار يُوفّر هذا التصنيف إطاراً متكاملًا للمعلم يمكنه من تخطيط التعلم بصورة متوازنة بحيث يشمل المعرفة، المهارات، والاتجاهات، ويتيح له صياغة أهداف تعليمية قابلة للقياس والتقييم بما يضمن فاعلية العملية التعليمية وتحقيق النتائج المرجوة.

2-1- المجال المعرفي أو الإدراكي (العقلي) (The Cognitive Domain):

يشمل المجال المعرفي أشكال النشاط الفكري لدى الإنسان ويركز بشكل خاص على العمليات العقلية كالحفظ، والفهم، والتحليل، والاستنتاج. يضم هذا المجال الأهداف التربوية التي تهدف إلى تنمية المعرفة والقدرات الذهنية لدى المتعلمين سواء من حيث القوانين والحقائق والنظريات والمفاهيم العامة أو المهارات العقلية المرتبطة بالتحليل والتفكير المنطقي.

وقد قام بلوم (Bloom) بتقسيم المجال المعرفي إلى ست مستويات فرعية مميزة ومرتبطة بشكل هرمي تبدأ من البسيط إلى الأكثر تعقيداً بحيث يشترط إتقان المستوى الأدنى قبل الانتقال إلى المستوى

الأعلى. هذا التقسيم يساعد المعلمين على صياغة أهداف تعليمية دقيقة ويعزز قدرة الطلاب على التقدم التدريجي في التفكير والمعرفة.

1-1-2- مستويات المجال المعرفي لبلوم (Cognitive Domain Levels – Bloom, 1956):

قدّم بلوم (Bloom, 1956) مع مجموعة من زملائه تصنيفاً شهيراً للأهداف التعليمية في المجال المعرفي (Cognitive Domain) وهو المجال الأول والأكثر استخداماً في ميدان التربية والتعليم. ويُعنى هذا المجال بالعمليات العقلية التي يقوم بها المتعلم ابتداءً من العمليات البسيطة كالتذكر والتعرف، وصولاً إلى المستويات الأكثر تعقيداً كالتحليل والتركيب والتقييم.

ويُعدّ المجال المعرفي حجر الزاوية في العملية التعليمية لأنه يركّز على تنمية القدرات العقلية والإدراكية التي تمكّن المتعلم من اكتساب المعرفة وفهمها وتطبيقها في مواقف جديدة ثم تحليل عناصرها وتركيبها في بنى جديدة وأخيراً إصدار الأحكام والتقييم.

وقد اعتمد بلوم وزملاؤه في تصنيف هذا المجال على المنطق الهرمي بحيث تُبنى كل مرحلة معرفية على ما قبلها ويصعب الانتقال إلى مستوى أعلى دون إتقان المستويات الأدنى. يبدأ الهرم بالمستويات السهلة البسيطة في القاعدة ويتدرج حتى يصل إلى المستويات المعقدة والمتقدمة في القمة. وتتمثل المستويات الستة التي صنفها بلوم على النحو الآتي:

1- مستوى التذكر أو الحفظ أو المعرفة (Knowledge / Remembering):

يعرف التذكر بأنه القدرة على استدعاء المادة التي سبق تعلمها ويشمل تذكر معلومات تتراوح من حقائق محددة إلى نظريات كاملة. يمثل هذا المستوى أقل مستويات المجال المعرفي من حيث التعقيد لكنه الأساس الذي يبنى عليه المستويات الأعلى.

- أمثلة على الأفعال المستخدمة في هذا المستوى: يحدد، يصف، يذكر، يسمي، يختار، ينسب، يعرف، يسترجع، يعدد.

- مثال تطبيقي: أن يعرف الطالب مفهوم الأمطار التضاريسية.

2- مستوى الفهم والاستيعاب (Comprehension / Understanding):

يعني القدرة على إدراك معنى المادة الدراسية ويظهر ذلك عبر تفسير أو شرح المعلومات أو تحويلها من شكل إلى آخر كال كلمات إلى أرقام. يعد هذا المستوى خطوة أبعد من مجرد التذكر.

- أمثلة على الأفعال المستخدمة: يشرح، يلخص، يعبر، يفسر، يميز، يرتب، يستدل، يترجم، يحسب، يعيد صياغة، يؤيد، يستنتج، يعلل، يعطي.

- مثال تطبيقي: أن يستنتج الطالب أسباب صراع الشرق والغرب على الوطن العربي منذ القدم.

3- مستوى التطبيق (Application):

يعني قدرة الطالب على استخدام ما تعلمه من مفاهيم وقوانين ومبادئ في مواقف جديدة وحل المشكلات المألوفة وغير المألوفة.

- أمثلة على الأفعال المستخدمة: يطبق، ينتج، يعد.

- مثال تطبيقي: أن يحل الطالب مسائل في التفاضل.

4- مستوى التحليل (Analysis):

يشير إلى قدرة المتعلم على تفكيك مادة التعلم إلى مكوناتها الجزئية لفهم تنظيمها البنائي أي القدرة على تحليل المشكلة أو الفكرة وفهم العلاقات بين عناصرها.

- أمثلة على الأفعال المستخدمة: يجزي، يفرق، يميز، يتعرف على، يوضح، يستنتج، يختار، يفصل، يقسم.

- مثال تطبيقي: أن يحلل الطالب أهمية موقع الوطن العربي بالنسبة إلى العالم.

5- مستوى التركيب (Synthesis / Creating):

يعني الجمع بين العناصر والأجزاء لتكوين بناء جديد بحيث يتم ربط هذه العناصر بطريقة مبتكرة أو منظمة لتنتج فكرة أو منتجًا جديدًا.

- أمثلة على الأفعال المستخدمة: يصنف، يؤلف، يجمع، يبتكر، يصمم، يشرح، يعدل، ينظم، يعيد الترتيب، يعيد البناء، يربط بين، يراجع، يلخص، يحكي، يكتب موضوعًا، يقترح.

- مثال تطبيقي: أن يكتب الطالب موضوعًا إنشائيًا.

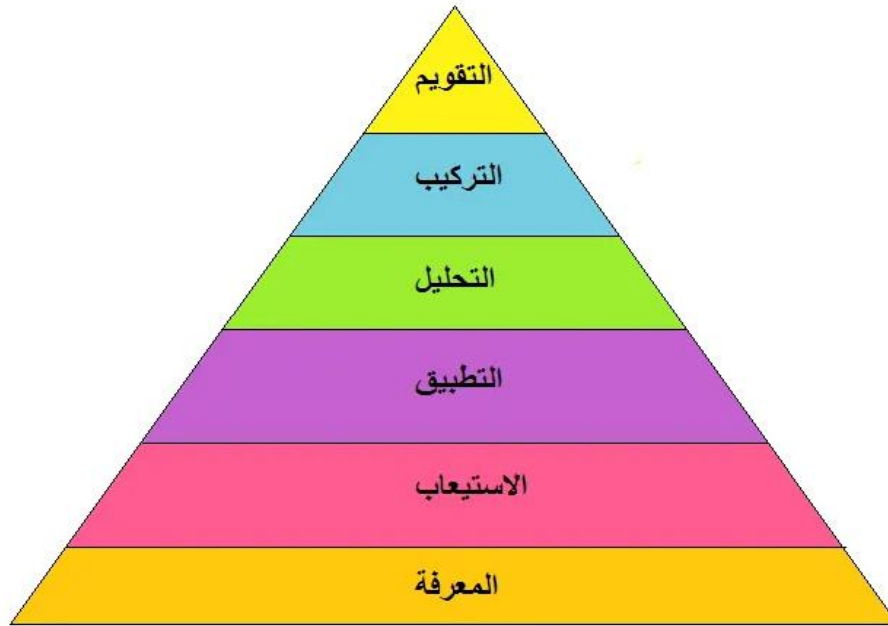
6- مستوى التقويم (Evaluation):

يعرف التقويم بأنه القدرة على إصدار حكم على قيمة المادة أو الظاهرة وفق معايير محددة سواء كانت داخلية (ذاتية) أو خارجية (موضوعية) مرتبطة بالغرض أو الهدف.

- أمثلة على الأفعال المستخدمة: ينقد، يقيم، يبدي رأيه، يحكم، يقرر، يثمن، يستخلص، يقوم، يدعم، يقدر، يبرز.

- مثال تطبيقي: أن يحكم الطالب على حركة التغيير الاجتماعي في ضوء القيم والمبادئ التي يؤمن بها.

شكل 08 يوضح صنفات بلوم (Bloom's Taxonomy).



(جودة، 2001، ص.162)

2-1-2- تصنيف بلوم المعدل (Bloom's Revised Taxonomy):

شهد تصنيف بلوم تطوراً مهماً تمثل فيما عُرف بـ مصفوفة التعلم والتدريس والتقييم: تنقيح مصفوفة بلوم للأهداف التربوية الذي قام به أندرسون بالتعاون مع زميله كراثول (Anderson & Krathwohl, 2001) ضمن فريق ضم نخبة من الخبراء في مجالات علم النفس المعرفي والمناهج وطرائق التدريس والاختبارات التربوية والقياس والتقييم. وقد مثل هذا التنقيح نقلة نوعية في الفكر التربوي والتعليمي إذ أعيدت صياغة التصنيف ليتلاءم مع التحولات المعرفية والتربوية المعاصرة في مطلع القرن الحادي والعشرين. فبينما اتسم التصنيف الأصلي (1956) ببنية هرمية جامدة جاء التصنيف المعدل أكثر ديناميكية وترابطاً ليعكس طبيعة عمليات التفكير لدى المتعلم بوصفها عمليات متداخلة ومتفاعلة وليست مجرد خطوات خطية متتابعة (السعدوي، 2018، ص.111).

ومن أبرز التغييرات التي أدخلت على هذا التصنيف التحول من استخدام الأسماء، مثل: (المعرفة، الفهم، التطبيق...) إلى استخدام الأفعال التي تعبر بشكل أوضح عن السلوك التعليمي المتوقع، وبذلك أصبحت المستويات الجديدة كالتالي:

1- تذكر (Remembering)

2- الفهم (Understanding)

3- التطبيق (Applying)

4- التحليل (Analyzing)

5- التقويم (Evaluating)

6- الابداع / الابتكار (Creating)

شكل 09 يوضح صنافه بلوم المعدلة (Revised Bloom's Taxonomy).



(Magno, et al, 2010, p.08)

ويلاحظ أنّ التغيير الأبرز يتمثل في إضافة الابداع أو الابتكار (Creating) كأعلى مستوى في الهرم، بعد أن كان التقويم (Evaluation) هو القمة في التصنيف القديم. شمل التنقيح كذلك إضافة بُعد جديد للبعد المعرفي، كما هو موضح في الجدول الآتي حيث تضمن هذا البعد أربع تصنيفات فرعية أساسية، وهي: المعرفة الحقائقية (Factual Knowledge)، والمعرفة المفاهيمية (Conceptual Knowledge)، والمعرفة الإجرائية (Procedural Knowledge)، وأخيرًا المعرفة ما وراء المعرفية (Metacognitive Knowledge) (السعدوي، 2018، ص113).

جدول 06 يمثل مستويات البعد المعرفي

البعد المعرفي (Knowledge Dimension)						بعد العمليات المعرفية (Cognitive Process Dimension)					
						يتذكر	يفهم	يطبق	يحلل	يقوم	يبتكر
الحقائق المعرفية (Factual Knowledge)											
المفاهيم المعرفية (Conceptual Knowledge)											
المعرفة الإجرائية (Procedural Knowledge)											
المعرفة التأملية (Metacognitive Knowledge)											

وفيما يلي عرضٌ لمستويات البعد المعرفي:

1- المعرفة الحقائقية (Factual Knowledge): تمثل المعارف الأساسية التي ينبغي على الطالب

تذكرها للتعرف على موضوعات المادة وتشمل المصطلحات والمعارف المتخصصة المرتبطة بالمجال إضافةً إلى الحقائق أو العناصر الجوهرية التي يستند إليها المتخصص عند الحديث عن

موضوع معين (Pickard, 2007).

2- المعرفة المفاهيمية (Conceptual Knowledge): تُعد أكثر تعقيداً من المعرفة الحقائقية وتتكون

من ثلاثة عناصر رئيسية:

أ- المعرفة بالتصنيفات والفئات.

ب- المعرفة بالمبادئ والتعميمات.

ج- المعرفة بالنظريات والنماذج والأطر (Anderson & Krathwohl, 2001).

ويظهر امتلاك الطالب للمعرفة المفاهيمية عندما يتمكن من شرح المفاهيم بأسلوبه الخاص ونقلها إلى مواقف جديدة. كما يمكن تحديد خصائص المفاهيم ووضع تعريفات لها في حين تُستخدم التعميمات لتوضيح العلاقات فيما بينها. ويُشكّل تصنيف المفاهيم وتبويبها الأساس الذي تُبنى عليه المبادئ والتعميمات والتي بدورها تمثل قاعدة النظريات والنماذج والأطر النظرية. وتشغل هذه العناصر حيزاً واسعاً من التفكير ضمن مختلف المواد الدراسية (السعدوي، 2018، ص115).

3- المعرفة الإجرائية (Procedural Knowledge):

بينما ترتبط المعرفة الحقائقية والمفاهيمية بالمنتجات المعرفية فإن المعرفة الإجرائية تتمثل في إدراك سلسلة من الخطوات والإجراءات اللاحقة وفهم المحركات التي تحكم استخدام كل إجراء. وتشمل

هذه المعرفة القدرة على تنفيذ خطوات عملية أو تطبيق قواعد وإجراءات محددة في سياقات مختلفة. فعلى سبيل المثال قد تتجسد المعرفة الإجرائية في معرفة مكونات المقال أو في تطبيق قواعد لعبة معينة.

4- المعرفة التأملية/ ما وراء المعرفة (Metacognitive Knowledge):

تُعرف أيضاً بالوعي المعرفي وتمثل وعي الفرد بعمليات تفكيره وإدراكه لها وهي حجر الأساس لفهم عمليات التعلم. وقد أصبح الاهتمام في الآونة الأخيرة منصباً على تنمية وعي الطلاب بعملياتهم المعرفية والتفكيرية وجعلهم أكثر مسؤولية عن الطرق والإجراءات التي يتبعونها عند مواجهة المشكلات وحلها. فقد يكتسب الطالب معارف من دراسته لكن قد يفتقر إلى القدرة على مراقبة شروط تعلمه أو التكيف مع استراتيجيات التعلم بما ييسر اكتساباً أعمق وأبعد من الفهم السطحي (Anderson & Krathwohl, 2001).

وتتمثل الطريقة المثلى لمساعدة الطلاب على تنمية هذه المهارات في تدريبهم على التأمل في جهودهم المبذولة كالتفكير في مقدار الجهد الذي صرفوه في إنجاز الواجبات أو الاستعداد للاختبارات. فحين يتأمل الطلاب حجم الجهد المبذول ويربطونه بالنتائج المحصلة يصبحون أكثر وعياً بالعلاقة بين الجهد والتحصيل ويدركون أن الإخفاق في بذل الجهد الكافي يؤدي بالضرورة إلى مستويات أدنى من الإنجاز (Marzano, Norford, Paynter, Pickering, & Gaddy, 2001) (السعدوي، 2018، ص115).

ويعكس هذا التعديل قناعة تربوية حديثة مفادها أن القدرة على الإبداع والابتكار تمثل أعلى صور النشاط العقلي وأكثرها تعقيداً إذ تجمع بين جميع المهارات السابقة وتوظفها في إنتاج معرفة أو حل جديد لمشكلة أو تصميم فكرة مبتكرة. كما أن التصنيف المعدّل لم يعد ينظر إلى المستويات بوصفها درجات منفصلة بقدر ما هي حلقة متصلة حيث يمكن للمتعلمين الانتقال صعوداً وهبوطاً بين المستويات وفقاً لمتطلبات الموقف التعليمي. وهذا ما يجعل التصنيف الجديد أكثر مرونة وواقعية في توصيف مسارات التعلم.

ومن الناحية التطبيقية يوجّه هرم بلوم المعدّل المعلمين إلى تصميم أنشطة تعليمية وتقييمات تدمج بين المعرفة النظرية والممارسة العملية ويشجع على استخدام مهارات التفكير العليا كالتفكير النقدي حل المشكلات والإبداع. ومن ثمّ فهو يُعتبر أداة تربوية أساسية في القرن الحادي والعشرين خصوصاً مع تزايد الحاجة إلى خريجين يمتلكون مهارات مركّبة تمكّنهم من التكيف مع بيئات معرفية ومهنية متغيرة باستمرار.

2-2- المجال الانفعالي أو الوجداني (العاطفي) (The Affective Domain):

يشمل هذا المجال الأهداف التي يسعى لتحقيقها أن يظهر المتعلم سلوكاً انفعالياً متوازناً كالتعبير عن مشاعر الحب والاحترام وتقدير العلماء والمبدعين وممارسة التسامح والتقبل وعدم التعصب أو التعجب إضافةً إلى الاستجابة الملائمة للمواقف المختلفة. كما يركز هذا المجال على تنمية المشاعر الإنسانية لدى المتعلم وتطويرها بما يعزز القدرة على التكيف الاجتماعي وبناء علاقات إيجابية مع الآخرين.

ترتبط الأهداف في هذا المجال بمستوى قبول المتعلم أو رفضه لأشياء أو قيم معينة ويتميز السلوك الوجداني بالثبات النسبي على المدى الطويل حيث يظهر في الميول والاتجاهات والقيم وأشكال التقدير المختلفة. تكمن أهمية هذا المجال في كونه يعكس جوهر السلوك الإنساني ويشكل جزءاً لا يتجزأ من نمو الشخصية. ومن الخطأ عزل الجوانب الانفعالية عن الجوانب المعرفية فهما متكاملتان بشكل تام؛ إذ يمثل العقل البشري المدخل الأساسي إلى الوجدانيات ويمكن القول بأن الطريق إلى قلب وعقل المتعلم يبدأ من تنشيط وتوجيه إدراكه ومعرفته مما يجعل المجال المعرفي أداة تمهيدية لتحقيق الأهداف الانفعالية.

2-2-1- مستويات المجال الانفعالي أو الوجداني أو العاطفي (Krathwohl, Bloom, & Masia,)

(1964) (Affective Domain Levels – Krathwohl):

قدم كراثول (Krathwohl, 1964) تصنيفاً متكاملًا للأهداف التعليمية في المجال الوجداني أو العاطفي أو الانفعالي (Affective Domain) باعتباره المجال الثاني المهم إلى جانب المجال المعرفي. فإذا كان المتعلم في المجال المعرفي يُكَلَّفُ بالتعامل مع العمليات العقلية على اختلاف مستوياتها من معرفة وفهم وتحليل وتركيب وتقويم، فإن المطلوب منه في المجال الوجداني أن يتعامل مع ما يختزنه القلب من اتجاهات ومشاعر وأحاسيس وقيم تؤثر بصورة مباشرة في سلوكه الظاهر وأنشطته المتنوعة وطريقة تفاعله مع ذاته والآخرين.

وقد استند كراثول في بنائه لهذا التصنيف إلى المنطق الهرمي ذاته الذي استخدمه بلوم في المجال المعرفي حيث يبدأ التصنيف من المستويات البسيطة التي تمثل استجابات أولية وسطحية ثم يتدرج نحو المستويات الأعلى التي تعكس نضجاً وجدانياً وقيماً أعمق إلى أن يبلغ قمته التي تمثل الثبات القيمي والشخصي في السلوك. وبهذا جاء تصنيفه في صورة هرم يتدرج من السهل إلى الصعب ومن البسيط إلى المركب، وتتمثل المستويات الخمسة لهذا المجال في الآتي:

1- مستوى الاستقبال أو التقبل (Receiving):

يعرف هذا المستوى بأنه استعداد المتعلم للاهتمام بظاهرة أو مثير معين كنشاط تعليمي في الفصل أو وسيلة تعليمية أو الكتاب المقرر. ويعتبر التقبل أدنى مستويات نواتج التعلم في المجال الانفعالي حيث يظهر استعداد المتعلم للاستماع والانتباه والاهتمام قبل أن يشارك بفاعلية.

- **الأفعال المرتبطة بهذا المستوى:** ينتبه، يسأل، يصغي، يتابع، يتعرف، يبدي، يختار، يجيب
- **مثال:** أن ينتبه الطالب لشرح المعلم حول خطورة التدخين ويُظهر استعداداً للاستماع من دون مقاطعة.

2- مستوى الاستجابة (Responding):

يشمل هذا المستوى المشاركة الإيجابية والتفاعل الحسي مع الموقف التعليمي بحيث يعكس ميول المتعلم واهتماماته ويسعى للبحث عن الأنشطة التي تشبع حاجاته. ويعد خطوة متقدمة بعد التقبل حيث يظهر السلوك الانفعالي في شكل أفعال ملموسة.

- **الأفعال المرتبطة بالاستجابة:** يجيب، يساير، يشعر، يقرر، يعاون، يناقش، يؤدي، يبدي، يسمع، يشترك.

- **مثال:** أن يجد الطالب متعة في قراءة الانتصارات العسكرية التي حققها المسلمون على الروم أيام الخلفاء الراشدين.

3- مستوى التقدير أو التقويم (Valuing):

يقصد به تقييم المتعلم لقيمة ظاهرة أو سلوك معين بحيث يبدأ بقبول الظاهرة ثم التعبير عن رأيه فيها وصولاً إلى إصدار أحكام مدروسة والدفاع عنها. ويتميز هذا المستوى بأن القيم تتشكل تدريجياً وتتعمق مع الوقت.

- **الأفعال المرتبطة بالتقدير:** يبادر، يبرز، يعمل، يقترح، يمارس، يتابع، يقدر، يشارك، يساهم، يدعو.
- **مثال:** أن يقدر الطالب جهود العلماء العرب المسلمين في تطوير علم الطب ويعبر عن رأيه في أهميتها.

4- مستوى التنظيم القيمي (Organization):

يتضمن هذا المستوى جمع القيم المختلفة وإعادة ترتيبها داخلياً بحيث تتماشى مع السلوك أو الظاهرة محل الدراسة مع القدرة على تعديلها عند إدخال قيم جديدة. ويشمل التنظيم القيمي مقارنة القيم وتحليلها وإعادة تركيبها بطريقة واعية.

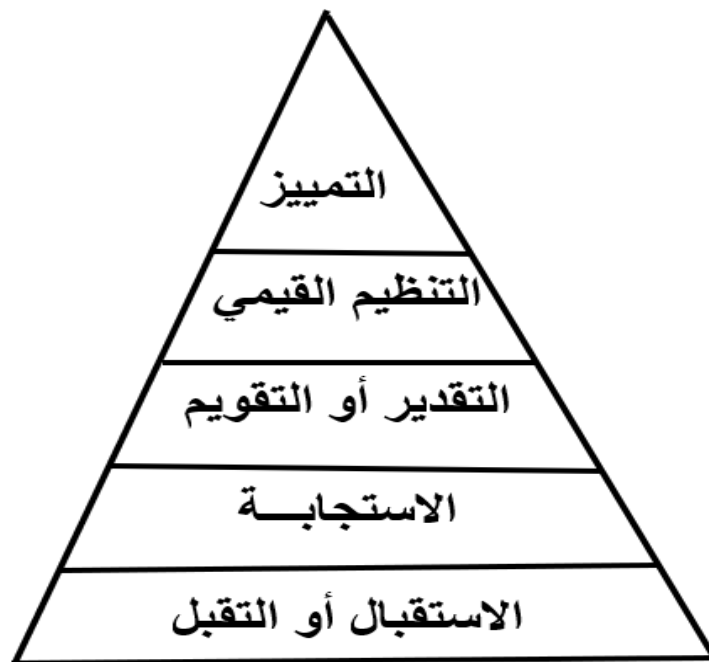
- الأفعال المرتبطة بهذا المستوى: ينظم، يصوغ، يفاضل، يصحح، يجمع بين، يرتب، يتمسك.
- مثال: أن يعيد الطالب ترتيب أولوياته اليومية بحيث يقدم أداء الصلاة والواجبات الدراسية على الأنشطة الترفيهية.

5- مستوى التمييز (بناء النسق القيمي) (Characterization):

يعد هذا المستوى أرقى مستويات المجال الانفعالي إذ يعكس ثبات السلوك ووضوح الشخصية والفردية في التعامل مع المواقف المختلفة بحيث يصبح السلوك أسلوب حياة للفرد وفلسفة متكاملة.
- الأفعال المرتبطة بالتمييز: يميز، يؤدي، يستخدم، يؤمن، يستحي، يقترح، يساهم، يظهر، يغير، يحل.

- مثال: أن يتبنى الطالب الصدق كقيمة أساسية في حياته فيظهر ذلك في سلوكه اليومي داخل المدرسة وخارجها حتى في المواقف الصعبة التي تتطلب الاختيار بين المصلحة الشخصية والالتزام بالمبادئ.

شكل 10 يوضح صناف كراثول (Krathwohl, 1964)



2-3- المجال النفس حركي أو المهاري (The Psychomotor Domain):

يشمل المجال النفس-حركي أو المهاري الأهداف التربوية التي تهدف إلى تنمية المهارات الحركية لدى المتعلم حيث يُتوقع منه أن يسلك سلوكًا يقوم فيه بأداء حركات منظمة باستخدام عضلات الجسم المختلفة. ويركز هذا المجال على الأنشطة التي تتطلب التنسيق العضلي-العصبي كالكتابة،

الرسم، تشغيل الأجهزة، ممارسة الرياضة، أو أداء أي عمل يتطلب الدقة والتأزر بين الحواس والعضلات.

إن إتقان هذه المهارات لا يعتمد فقط على الممارسة البدنية والتدريب العملي بل يتأثر كذلك بعوامل نفسية ووجدانية كالذوافع والاتجاهات والميول الأمر الذي يجعل هذا المجال متداخلاً ومتفاعلاً مع كل من المجال المعرفي والمجال الانفعالي. فعندما يمارس المتعلم مهارة معينة فإن جودة أدائه ومستوى تمكنه منها يتوقفان بدرجة كبيرة على معرفته النظرية السابقة ذات الصلة ومدى ارتباطها بخبراته إضافة إلى طبيعة التدريب الذي تلقاه ومدى شعوره بأهمية هذه المهارة واقتناعه بجوداها. ولهذا يمكن القول إن إتقان أي مهارة حركية يتطلب تكاملاً بين المعرفة النظرية والتدريب العملي والدافعية الداخلية مما يفسر لماذا يُقال إن الإنسان لا يحقق النجاح الحقيقي في عمله إلا إذا أحبه لأن الحب والرغبة الداخلية يمثلان قوة دافعة نحو الإبداع والإتقان.

2-3-1- مستويات المجال النفس-حركي أو المهاري (Dave, Psychomotor Domain Levels- (1970):

قدّم عدد من الباحثين – من أبرزهم ديف (Dave, 1970) وسيمبسون (Simpson, 1972) وهارو (Harrow, 1972) تصنيفات للمجال النفس-حركي باعتباره المجال الذي يهتم بقدرة المتعلم على استخدام العضلات والجهاز العصبي في أداء أنشطة عملية أو حركية تتطلب التناسق بين العقل والجسم (Kumar, et al, 2016, p.36).

ويعنى هذا المجال بتنمية المهارات العملية والحركية التي تبدأ من مجرد التقليد البسيط وتنتهي بقدرة الفرد على أداء سلوك أو نشاط بمهارة عالية ودقة وإبداع. وقد جرى تصنيف مستوياته على غرار المجالين المعرفي والوجداني في صورة هرمية تبدأ بالمهارات الأساسية البسيطة وتترج إلى المستويات العليا الأكثر إتقاناً وتعقيداً، ويتضمن هذا المجال خمسة مستويات كالآتي:

1- مستوى التقليد أو المحاكاة (Imitation):

يمثل مستوى التقليد الخطوة الأولى في اكتساب المهارات الحركية حيث يقوم المتعلم بملاحظة حركة أو مجموعة من الحركات ثم محاولة تقليدها بشكل واعٍ. ويُعد هذا المستوى مدخلاً أساسياً لعملية التعلم المهاري إذ يبدأ المتعلم من خلاله بترجمة ما يراه إلى استجابات حركية فعلية. ويتضمن هذا المستوى جانبين رئيسيين:

أ- **الدفع للمحاكاة:** ويُقصد به الاستجابة الأولية للعضلات عند بدء عملية التقليد حيث يحاول المتعلم تقليد ما شاهده ولو بصورة غير دقيقة أو متكاملة مما يعكس بداية تكوين الارتباط بين الإدراك البصري والاستجابة الحركية.

ب- **القدرة على التكرار:** وتعني قدرة المتعلم على إعادة الحركة أو مجموعة الحركات بشكل متكرر ومع كل محاولة تتحسن درجة الضبط والدقة وهو ما يُعد تمهيداً للوصول إلى مستويات أعلى من الإتقان.

ويلاحظ أن هذا المستوى يعتمد بدرجة كبيرة على انتباه المتعلم ودافعيته الداخلية ورغبته في التعلم فكلما كان المتعلم أكثر انخراطاً واهتماماً كان تقليده للحركات أكثر فاعلية ودقة الأمر الذي يضع الأساس للانتقال إلى مراحل أكثر تقدماً من الأداء المهاري.

2- مستوى المعالجة أو التناول (Manipulation):

يمثل هذا المستوى انتقال المتعلم من مرحلة التقليد إلى مرحلة الاعتماد على التعليمات في أداء الحركات. ففي حين كان الأداء في المستوى الأول قائماً على الملاحظة المباشرة للآخرين فإن المتعلم هنا يعتمد على التعليمات أو التوجيهات (الشفهية، المكتوبة، أو الرمزية) ليبدأ بتنفيذ المهارة المطلوبة. يُظهر هذا المستوى بداية الاستقلالية النسبية في الأداء حيث يبدأ المتعلم في التحكم الواعي في الحركة ويعمل على تعديل أدائه وفقاً لما يتلقاه من تعليمات وتغذية راجعة.

- مكونات هذا المستوى:

أ- **اتباع التوجيهات لتنمية المهارة:** يلتزم المتعلم بالإرشادات المحددة من قبل المدرب أو المعلم في كيفية أداء الحركة كطريقة الإمساك بالأداة أو خطوات حل تمرين عملي.

ب- **أداء الحركة التي يتم اختيارها:** يبدأ المتعلم بتنفيذ الحركات بشكل مستقل نسبياً حيث يصبح قادراً على اختيار الحركة الصحيحة وتنفيذها بشكل متكرر دون الحاجة للملاحظة المباشرة.

ج- **تثبيت الأداء بالتدريب:** من خلال التدريب المستمر يبدأ الأداء بالتحسن من حيث الدقة والسرعة والانسيابية مما يمهد الطريق لانتقال المتعلم إلى مستويات أعلى من الإتقان. يمكن القول إن مستوى المعالجة أو التناول يركّز على:

- التعلّم بالتوجيه بدلاً من الملاحظة.

- تنمية القدرة على الاستجابة الواعية للتعليمات.

- ترسيخ الأساس المهاري عبر التكرار والتدريب.

3- مستوى الإتقان أو الدقة (Precision):

يمثل هذا المستوى مرحلة متقدمة في تعلم المهارة حيث ينتقل المتعلم من مجرد أداء الحركة بتوجيه أو تدريب إلى مستوى أعلى من الدقة والانضباط في الأداء. في هذه المرحلة تصبح الحركات أكثر تحكماً وسلاسة وثباتاً مع انخفاض الحاجة إلى التوجيهات الخارجية أو المساعدة.

- مكونات هذا المستوى:

أ- إعادة الأداء ذاتياً: يستطيع المتعلم إعادة الحركة أو المهارة بشكل مستقل وبنفس الكفاءة دون الاعتماد على الملاحظة أو التعليمات المتكررة.

ب- الانضباط في الأداء: يتميز الأداء هنا بالاتساق والدقة العالية في التفاصيل مما يدل على أن المتعلم أصبح قادراً على التحكم التام في أبعاد الحركة أو المهارة المطلوبة مع تقليل الأخطاء إلى أدنى حد ممكن.

- خصائص هذا المستوى:

- زيادة مستوى الثقة بالنفس أثناء الأداء.
- القدرة على تكرار النتائج بدقة عند إعادة نفس المهمة.
- ظهور الاستقرار الحركي والانسجام في الأداء.
- يمثل أساساً للانتقال إلى مستويات أكثر تعقيداً كالتنسيق أو التآزر (Articulation).

4- مستوى التنسيق أو التآزر (Articulation):

يمثل هذا المستوى مرحلة متقدمة من التعلم الحركي حيث لا يقتصر الأداء على حركة واحدة منفردة بل يتعدى ذلك إلى دمج مجموعة من الحركات في سلسلة مترابطة ومنظمة بحيث يظهر المتعلم قدرة على التوفيق والتنسيق بين مكونات الحركة لتحقيق أداء متكامل ومتوازن.

- مكونات هذا المستوى:

- أ- تكوين السلسلة المناسبة من الحركات: يتمكن المتعلم من دمج عدة حركات جزئية في إطار واحد متكامل بحيث تتتابع بطريقة منطقية وسلسة تؤدي إلى تحقيق الهدف المطلوب.
- ب- تحقيق التآلف أو التوافق بين مجموعة من الحركات المختلفة: ينجح المتعلم في تنسيق الحركات المتنوعة لتعمل معاً بشكل متجانس ومترابط بما يعكس مستوى عالياً من الانسجام الداخلي للأداء الحركي.

- خصائص هذا المستوى:

- القدرة على الربط بين حركات بسيطة لتكوين مهارات مركبة.

- ظهور الانسجام والدقة عند أداء الحركات المتعددة معًا.
- يعكس مستوى من المرونة والقدرة على التكيف مع متطلبات المواقف المختلفة.
- يُعدّ خطوة انتقالية نحو أعلى مستوى في المجال النفس-حركي وهو التلقائية أو الأصالة (Naturalization / Origination).

5- مستوى التطبيع أو الأتمتة (Naturalization):

يمثل هذا المستوى أعلى درجات المجال النفس-حركي حيث يصل المتعلم إلى مرحلة الإتقان التام للحركة أو المهارة فيؤديها بصورة طبيعية وآلية دون حاجة إلى جهد واعٍ أو تفكير متكرر في خطواتها. في هذا المستوى تصبح المهارة جزءًا من السلوك الاعتيادي للمتعلّم وتتم بأقل قدر ممكن من الطاقة الجسدية والذهنية.

- مكونات هذا المستوى:

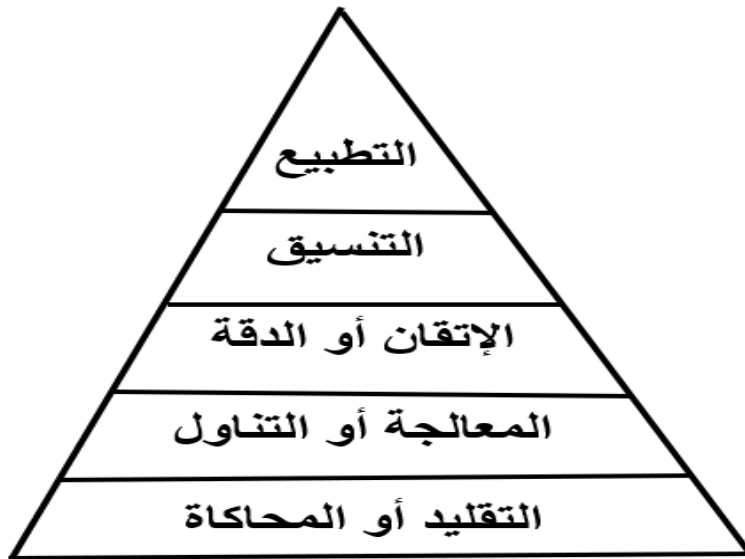
أ- **الحركة الذاتية أو الاستجابة التلقائية:** حيث يقوم المتعلم بأداء الحركة بشكل آلي وسريع دون تردد أو تفكير مسبق وكأنها صارت رد فعل طبيعيًا.

ب- **تكوين العادة في إجراء الحركة:** تصبح الحركة جزءًا من العادات اليومية للفرد حيث يؤديها تلقائيًا وبسهولة نتيجة الممارسة المستمرة والتكرار حتى الوصول إلى التطبيع الكامل.

- خصائص هذا المستوى:

- الأداء يصبح سلسًا وفعالًا دون بذل جهد زائد.
- المهارة تظهر في صورة إبداعية أو أصيلة يمكن أن يطورها المتعلم ويكيفها مع مواقف جديدة.
- الانتقال من مجرد تقليد أو محاكاة إلى مستوى التفرد والإبداع في الأداء.

شكل 11 يوضح مستويات المجال النفس-حركي أو المهاري لديف (Dave, 1970)



جدول 07 يوضح مستويات المجال النفس-حركي أو المهاري مع الأفعال المناسبة ومثال لكل مستوى

المستوى	أفعال مناسبة	مثال
التقليد أو المحاكاة (Imitation)	يقلد، يكرر، يتبع، يحاكي	تلميذ يقلد معلمه في طريقة مسك الكرة
المعالجة أو التناول (Manipulation)	ينفذ، يتناول، يطبق، يعالج	طالب يستخدم المسطرة لرسم خط مستقيم بعد توجيه
الإتقان أو الدقة (Precision)	يتقن، يضبط، يصحح، يعدل	لاعب يرمي السهام نحو الهدف بدقة متكررة
التنسيق أو التآزر (Articulation)	ينسق، يرتب، يجمع، يركب	ممرض ينسق بين يديه أثناء تركيب الحقنة
التطبيع أو الأتمتة (Naturalization)	يعتاد، يبدل، يكتيف، يظهر، يقود	سائق يقود السيارة تلقائياً دون وعي كبير بالخطوات

وهناك أيضاً تصنيفات أخرى للمجال النفس-حركي أو المهاري كتصنيف (Simpson, 1972):

- اقترح سيمبسون (Simpson, 1972) للمجال النفس-حركي (Psychomotor Domain) تصنيفاً مرتب المستويات من الأدنى إلى الأعلى كالآتي:

1- الإدراك الحسي (Perception)

2- التهيئة (Set)

3- الاستجابة الموجهة (Guided Response)

4- الآلية أو الميكانيكية (Mechanism)

5- الاستجابة المركبة أو المعقدة (Complex Overt Response)

6- التكيف والمواءمة (Adaptation)

7- الأوتوماتيكية أو الإبداع الحركي (Origination)

- اقترح كذلك هارو (Harrow, 1972) تصنيفاً للمجال النفس-حركي والذي يركز على الحركات الأساسية والحركات المهارية المعقدة:

1- الحركات الانعكاسية (Reflex Movements)

2- الحركات الأساسية (Basic Fundamental Movements)

3- القدرات الإدراكية (Perceptual Abilities)

4- القدرات البدنية (Physical Abilities)

5- الحركات الماهرة (Skilled Movements)

6- الاتصال غير اللفظي (Non-discursive Communication)

3- شروط صياغة الأهداف السلوكية:

تُعَدُّ صياغة الأهداف السلوكية خطوة أساسية في التخطيط التربوي إذ تمثل الأساس الذي تُبنى عليه العملية التعليمية وتُوجَّه من خلاله جهود المعلم والمتعلم معًا. وحتى تؤدي هذه الأهداف دورها بفعالية لا بد من مراعاة مجموعة من الشروط المنهجية والعلمية، من أبرزها ما يلي:

1- **وضوح الهدف ودقته:** ينبغي أن يُصاغ الهدف بلغة واضحة ومباشرة بحيث لا تحتمل عباراته أكثر من معنى مما يضمن فهمه وتفسيره بالطريقة نفسها من قبل جميع المعنيين.

2- **التركيز على سلوك المتعلم:** يتمحور الهدف حول ما يُتوقع من الطالب أن ينجزه أو يكتسبه وليس حول ما يقوم به المعلم من أنشطة أو إجراءات.

3- **قابلية الملاحظة والقياس:** يشترط أن يُعبر الهدف عن سلوك يمكن ملاحظته وقياسه بوسائل موضوعية حتى يمكن التحقق من مدى تحققه في الواقع.

4- **مناسبة مستوى الطلبة:** يجب أن تكون صياغة الهدف منسجمة مع مستوى نضج الطلبة العقلي والمعرفي لا أن تعكس مستوى المعلم أو معدّ المنهج.

5- **تحديد الحد الأدنى للأداء:** من الضروري أن يحدد الهدف المستوى الأدنى المقبول من الإنجاز ليكون بمثابة معيار أو محك للتقويم.

6- **إمكانية التحقيق والواقعية:** ينبغي أن يكون الهدف قابلاً للتحقيق ضمن الإمكانيات والموارد المتاحة وقريباً ما أمكن من واقع البيئة التعليمية.

7- **احتواؤه على فعل سلوكي محدد:** تتضمن صياغة الهدف فعلاً سلوكياً (أدائياً) يوضح نوع السلوك المطلوب ومستواه، مثل: يعرف، يحل، يقارن، يبرهن... إلخ.

8- **التركيز على ناتج تعليمي واحد:** الهدف السلوكي الجيد يصف نتيجة تعليمية واحدة محددة حتى لا يختلط على المعلم أو المتعلم غرضه الأساسي.

9- **تجنب التكرار والتداخل:** يجب أن تُكتب الأهداف بطريقة متميزة بحيث لا تتكرر أو تتعارض مع أهداف أخرى.

10- **الالتزام بالبناء اللغوي السليم للهدف:** يُفضل أن تُصاغ الأهداف وفق القاعدة التالية: (أن + فعل مضارع + حدث سلوكي محدد)،

- مثل: "أن يعدد الطالب خطوات حل المسألة"، أو "أن يميز الطالب بين المفاهيم الأساسية".

4- تحليل محتوى المادة الدراسية:

يُعدّ تحليل المحتوى من أبرز الأساليب التربوية التي تهدف إلى دراسة المادة الدراسية بشكل منهجي ومنظم. وقد عُرّف تحليل المحتوى بأنه "مجموعة من الأساليب والإجراءات الفنية التي وُضعت لتفسير وتصنيف مكونات المادة التعليمية بما في ذلك النصوص المكتوبة والرسومات والصور والأفكار التي يتضمنها الكتاب المدرسي" وذلك بغرض الكشف عن عناصرها الأساسية وعلاقاتها البنائية. وتتسم مهارات تحليل المحتوى بعدد من الخصائص والسمات التي تُميزها عن غيرها من الإجراءات التربوية، ومن أهمها:

أ- **الطابع الوصفي:** يركز تحليل المحتوى على دراسة الظاهرة النصية ذاتها أي العلاقات والترابط بين النصوص والمضامين دون الانشغال باستكشاف النوايا الخفية للمؤلف أو مقاصده. فهو يقتصر على الوصف الموضوعي للمادة كما هي بعيداً عن التفسير الشخصي أو إصدار الأحكام التقويمية.

ب- **المنحى العلمي المنظم:** يعتمد التحليل على خطوات منهجية واضحة تتسم بالدقة والموضوعية بحيث يتم وصف المادة الدراسية وتحليل عناصرها كما وردت فعلاً ومن ثم تفسير الظواهر المرتبطة بها ضمن إطارها المعرفي.

ج- **التركيز على المكونات المتنوعة للمحتوى:** لا يقتصر التحليل على النصوص المكتوبة فقط بل يشمل كذلك الرسومات والجداول والخرائط والصور والمفاهيم التي يعرضها الكتاب وذلك من أجل بناء صورة شاملة عن البنية المعرفية للمحتوى.

د- **الابتعاد عن الذاتية:** يبتعد التحليل عن التقديرات الانطباعية ويسعى إلى تقديم بيانات يمكن التحقق منها ومقارنتها مما يجعله أداة علمية يمكن استخدامها لأغراض متعددة كتطوير المناهج أو إعداد الاختبارات أو تقويم الكتب المدرسية.

4-1- أهمية تحليل محتوى الكتاب الدراسي:

يمثل تحليل محتوى الكتاب المدرسي أحد المرتكزات الأساسية في العملية التعليمية إذ يتيح للمعلمين والمخططين التربويين فهم المادة الدراسية بعمق والوقوف على بنيتها التنظيمية ومعطياتها المعرفية، ومن خلال هذا التحليل يمكن الإجابة عن سؤال محوري هو: لماذا نحلل المحتوى؟

تبرز أهمية تحليل المحتوى في مجموعة من الجوانب التربوية، من أبرزها:

أ- **إعداد الخطط التعليمية:** يساعد التحليل على وضع الخطط الدراسية طويلة المدى (الفصلية) وقصيرة المدى (اليومية) بطريقة منهجية بما يضمن التدرج المنطقي في عرض المادة وتوزيعها على الحصص الدراسية.

ب- اشتقاق الأهداف التعليمية: يُعدّ التحليل مدخلاً رئيساً لاستخراج الأهداف التربوية العامة والخاصة وصياغتها بدقة لتوجيه العملية التعليمية نحو نتائج قابلة للتحقق.

ج- اختيار الاستراتيجيات التعليمية المناسبة: يتيح فهم طبيعة المحتوى ومستواه للمعلم فرصة انتقاء أساليب التدريس الأكثر ملاءمة سواء أكانت استراتيجيات تقليدية أم حديثة قائمة على المشاركة والتفاعل.

د- انتقاء الوسائل التعليمية والتقنيات: يساعد التحليل على تحديد الوسائل والأدوات التعليمية الأكثر فاعلية في دعم عملية التعلم كالخرائط، النماذج، العروض التفاعلية، أو الوسائط الرقمية.

هـ- بناء الاختبارات التحصيلية: يُسهّم في صياغة أسئلة وتقويمات مبنية على أسس علمية بحيث تعكس بدقة ما ورد في الكتاب وتغطي مستويات مختلفة من الأهداف وفق تصنيفات كتصنيف بلوم. و- تصنيف عناصر الكتاب المدرسي: من خلال التحليل يمكن تنظيم وتبويب موضوعات الكتاب ومكوناته مما يسهل على المعلم تخطيط الدروس وتنفيذها بوضوح وانتظام.

ز- الكشف عن جوانب القوة والضعف: يُعدّ التحليل أداة فاعلة لتشخيص مدى شمولية الكتاب المدرسي ودقته وملاءمته لخصائص المتعلمين وبالتالي يقدم ملاحظات بناءة لتطوير المحتوى وتحسينه.

5- إعداد جدول المواصفات:

يُعرّف جدول المواصفات بأنه أداة تنظيمية تُبنى على شكل قائمة تربط بين الأهداف التعليمية والمحتوى الدراسي من جهة، وبين عدد فقرات الاختبار التي تمثل هذه الأهداف والمحتوى من جهة أخرى. وبمعنى آخر يُعدّ هذا الجدول خريطة توضح المجالات التي سيركّز عليها المعلم عند إعداد الاختبار من خلال ترجمة محتوى المادة الدراسية إلى عناوين محددة تُعكس في صورة أسئلة أو فقرات.

وتتجلى أهمية جدول المواصفات في كونه يضمن شمولية الاختبار وتمثيله للمحتوى التعليمي بصورة متوازنة كما يحقق مجموعة من الفوائد التربوية، من أبرزها:

أ- الحكم على صدق الاختبار: يتيح الجدول للمعلم والمقيم التأكد من أن الاختبار يقيس فعلاً ما أُعدّ لقياسه وبالتالي يُعطي حكماً دقيقاً على مدى صدق الأداة.

ب- تقدير مستوى تحصيل الطلبة: من خلال تمثيل الأهداف والمحتوى بشكل علمي يمكن الحكم على مستوى التحصيل الدراسي للمتعلمين بدقة وموضوعية.

ج- قياس الأهداف التعليمية: يوفّر جدول الموصفات مؤشراً واضحاً حول مدى تحقق الأهداف التعليمية حيث يربط كل فقرة اختبارية بهدف محدد مما يعزز دقة عملية التقويم.

د- تحديد نسبة تمثيل المحتوى: يكشف الجدول عن نسبة كل وحدة أو موضوع من محتوى المنهج ضمن الاختبار الأمر الذي يضمن العدالة والتوازن بين الموضوعات ويُجنّب التركيز العشوائي أو الانتقائي.

5-1- إعداد جدول موصفات:

يُعدّ جدول الموصفات من الأدوات التربوية الأساسية في بناء الاختبارات التحصيلية حيث يُصاغ في صورة مخطط تفصيلي يجمع بين جانبي المحتوى الدراسي والأهداف السلوكية في إطار واحد متكامل بحيث يُظهر بوضوح كيف يرتبط كل هدف سلوكي بجزء محدد من محتوى المادة الدراسية.

ويُشار في هذا الجدول عادةً إلى النسب المئوية (الأوزان النسبية) التي تعكس درجة الأهمية لكل مجال من مجالات المحتوى ولكل مستوى من مستويات الأهداف السلوكية. وبهذا يصبح جدول الموصفات أداة مرجعية أساسية تضمن شمولية وصدق الاختبار التحصيلي. ويتضمن جدول الموصفات مجموعة من الجوانب الرئيسية، من أبرزها:

أ- محتوى المادة الدراسية: يتم حصر موضوعات المادة المراد قياسها مع إمكانية تفصيلها إلى مفردات أصغر وأكثر تحديداً بما يضمن شمول جميع العناصر المهمة.

ب- الأهمية النسبية للمحتوى: يُعطى لكل مفردة أو موضوع نسبة مئوية تعبّر عن وزنه الحقيقي في المنهج ويتم تحديد هذه النسبة إما بالرجوع إلى آراء خبراء المناهج والقياس والتقويم أو بالاعتماد على عدد الحصص المقررة لكل مجال دراسي مقارنة بعدد الحصص الكلي للمادة، ويُحسب ذلك باستخدام المعادلة الآتية:

$$\text{الأهمية النسبية للمجال} = \frac{\text{عدد الحصص المقررة للمجال}}{\text{عدد الحصص الكلي للمادة}} \times 100\%$$

ج- تحديد نسبة الأهداف السلوكية: بعد تحديد الأهمية النسبية لمجالات المحتوى يقوم المعلم بتوزيع هذه الأهمية على الأهداف السلوكية المختلفة، مثل: (التذكر، الفهم، التطبيق، التحليل...) بحيث تعكس هذه النسب مستوى التركيز على كل هدف أثناء عملية التدريس. وتُحدد الأهداف كذلك في صورة نسب مئوية.

د- تحديد عدد الأسئلة في الاختبار: يُحوّل الوزن النسبي لكل مجال ولكل هدف إلى عدد محدد من الفقرات الاختبارية بما يضمن تحقيق التوازن بين جميع جوانب المحتوى والمستويات المعرفية المختلفة.

هـ- إعداد الجدول النهائي: يُصاغ الجدول بصورته النهائية على شكل شبكة تربط بين مجالات المحتوى (في الصفوف) ومستويات الأهداف السلوكية (في الأعمدة) مع الإشارة إلى الأهمية النسبية وعدد الفقرات المخصصة لكل خلية. وبهذا يتحول جدول المواصفات إلى مرجع عملي يُوجّه عملية إعداد الاختبار خطوة بخطوة (المندلاوي، 2021، ص151).

- تمرين تطبيقي:

يريد معلم الرياضيات للمرحلة المتوسطة إعداد اختبار تحصيلي يتضمن (24) سؤالاً موزعة على أربعة مستويات معرفية وفق تصنيف بلوم المعدل. وبعد التشاور تم الاتفاق على التوزيع النسبي للأسئلة كما يلي: (التذكر 20%؛ الفهم 30%؛ التطبيق 35%؛ التحليل وما فوقه (تحليل-تقويم-إبداع) 15%). لنفترض أن عدد الحصص المقررة لمادة الرياضيات للمرحلة المتوسطة هي 40 حصة.

-المطلوب:

- 1-صمّم جدول مواصفات للاختبار يوضح عدد الأسئلة المخصصة لكل مستوى معرفي.
- 2-وزّع الأسئلة على خمس موضوعات رئيسية من مقرر الرياضيات بحيث يراعى التناسب بين الموضوعات وعدد الأسئلة الكلي.

خطوات الحل:

أولاً- تخطيط جدول تفصيلي يشمل على مجالات المستوى والأهمية النسبية والأهداف السلوكية. ثانياً: نحدد الموضوعات وعدد الحصص اللازمة لتدريس كل موضوع ونحسب الأهمية النسبية لكل مجال أو موضوع (الوزن النسبي):

يجب تحديد الموضوعات الأساسية للمادة الدراسية وعدد الحصص اللازمة لتدريس كل موضوع. بعد ذلك يتم حساب الأهمية النسبية لكل موضوع أو ما يعرف بالـ الوزن النسبي والذي يعكس نسبة كل موضوع من إجمالي عدد الحصص المقررة للمادة. لحساب الوزن النسبي للموضوعات نستخدم المعادلة الآتية:

$$\frac{\text{عدد حصص الموضوع} \times 100}{\text{إجمالي عدد الحصص}} = \text{الوزن النسبي للموضوع (\%)}$$

جدول 08 يلخص نتائج حساب الوزن النسبي لكل مجال في مادة الرياضيات

الموضوع	عدد حصص التدريس	الوزن النسبي لكل موضوع
الأعداد والجبر	12	30%
الهندسة والقياس	10	25%
الإحصاء والاحتمالات	08	20%
النسبة والتناسب	06	15%
حل المشكلات الرياضية	04	10%
المجموع الكلي	40	100%

ثالثاً- نستخرج عدد الأسئلة لكل مجال من مجالات المحتوى ويتم ذلك باستخدام المعادلة الآتية:

$$\text{عدد الأسئلة لكل مجال} = \frac{\text{الأهمية النسبية للمجال (\%)} \times \text{عدد الأسئلة الكلي}}{100}$$

- الأعداد والجبر (30%):

$$7.2 \approx 7 \text{ أسئلة} = \frac{30}{100} \times 24$$

- الهندسة والقياس (25%):

$$6 \text{ أسئلة} = \frac{25}{100} \times 24$$

- الإحصاء والاحتمالات (20%):

$$4.8 \approx 5 \text{ أسئلة} = \frac{20}{100} \times 24$$

- النسبة والتناسب (15%):

$$3.6 \approx 4 \text{ أسئلة} = \frac{15}{100} \times 24$$

- حل المشكلات الرياضية (10%):

$$2.4 \approx 2 \text{ أسئلة} = \frac{10}{100} \times 24$$

رابعاً- نستخرج عدد الأسئلة لكل مستوى من مستويات الأهداف السلوكية في الجدول ويتم ذلك باستخدام المعادلة الآتية:

$$\text{عدد الأسئلة في كل مستوى} = \frac{\text{نسبة الهدف السلوكي (\%)} \times \text{عدد أسئلة المجال}}{100}$$

- خطوات الحساب لكل مستوى:

- الأعداد والجبر (7 أسئلة):

- تذكر: $2 \approx 1.4 = \frac{20}{100} \times 7$
- فهم: $2 \approx 2.1 = \frac{30}{100} \times 7$
- تطبيق: $2 \approx 2.45 = \frac{35}{100} \times 7$
- تحليل: $1 \approx 1.05 = \frac{15}{100} \times 7$

- الهندسة والقياس (6 أسئلة):

- تذكر: $1 \approx 1.2 = \frac{20}{100} \times 6$
- فهم: $2 \approx 1.8 = \frac{30}{100} \times 6$
- تطبيق: $2 \approx 2.1 = \frac{35}{100} \times 6$
- تحليل: $1 \approx 0.9 = \frac{15}{100} \times 6$

- الإحصاء والاحتمالات (5 أسئلة):

- تذكر: $1 = \frac{20}{100} \times 5$
- فهم: $1 \approx 1.5 = \frac{30}{100} \times 5$
- تطبيق: $2 \approx 1.75 = \frac{35}{100} \times 5$
- تحليل: $1 \approx 0.75 = \frac{15}{100} \times 5$

- النسبة والتناسب (4 أسئلة):

- تذكر: $1 \approx 0.8 = \frac{20}{100} \times 4$
- فهم: $1 \approx 1.2 = \frac{30}{100} \times 4$
- تطبيق: $1 \approx 1.4 = \frac{35}{100} \times 4$
- تحليل: $1 \approx 0.6 = \frac{15}{100} \times 4$

- حل المشكلات الرياضية (2 أسئلة):

- تذكر: $0 \approx 0.4 = \frac{20}{100} \times 2$
- فهم: $1 \approx 0.6 = \frac{30}{100} \times 2$
- تطبيق: $1 \approx 0.7 = \frac{35}{100} \times 2$
- تحليل: $0 \approx 0.3 = \frac{15}{100} \times 2$

خامساً: يجب أن تكون الأسئلة لكل مجال من مجالات المحتوى مع مجموع الأسئلة في الخلايا لينتج عنه عدد الأسئلة الكلي.

سادساً: لضمان الدقة في تحديد عدد أسئلة الاختبار لكل مجال وكل خلية في جدول المواصفات يُنصح بعدم تقريب الأعداد أثناء العمليات الحسابية الأولية. ثم يُستخرج عدد الأسئلة بدقة باستخدام المعادلات المخصصة لكل مجال ومستوى سلوكي مع الاحتفاظ بالقيم العشرية عند الحساب.

وبعد الانتهاء من جميع العمليات الحسابية يمكن للمعلم إجراء التقريب النهائي عند اتخاذ القرار النهائي بشأن توزيع عدد الأسئلة لكل مجال أو مستوى بما يحقق التوازن والملاءمة العملية في الاختبار، ونتائج الخطوات السابقة نلخصها في الجدول النهائي الآتي:

جدول 09 مواصفات لمادة الرياضيات (الصف الثاني متوسط)

عدد الفقرات الكلي	الأهداف السلوكية				الوزن النسبي %	مجال المحتوى
	تحليل %15	تطبيق %35	فهم %30	تذكر %20		
7	1	2	2	2	%30	الأعداد والجبر
6	1	2	2	1	%25	الهندسة والقياس
5	1	2	1	1	%20	الإحصاء والاحتمالات
4	1	1	1	1	%15	النسبة والتناسب
2	0	1	1	0	%10	حل المشكلات الرياضية
24 سؤال	4	8	7	5	%100	المجموع

بعد الانتهاء من إعداد جدول المواصفات للاختبار التحصيلي يصبح المعلم قادراً على تقويم تحصيل الطلاب بدقة وموضوعية، حيث يضمن الجدول أن عدد الأسئلة وتوزيعها يعكس الأهمية النسبية للمحتوى الدراسي ومستويات التفكير المختلفة وفق تصنيف بلوم المعدل. كما يسهم الجدول في تحقيق التوازن بين الموضوعات والمستويات المعرفية ويسهل إعداد الاختبارات بشكل علمي ومنهجي ويمكن المعلم من اتخاذ قرارات دقيقة بشأن توزيع الأسئلة بما يتوافق مع المنهاج الدراسي. لذلك يعد جدول المواصفات أداة أساسية لضمان فاعلية الاختبار وموثوقيته من أجل تحسين جودة العملية التعليمية والتقويم داخل الصف.

الفصل الرابع

النظرية الكلاسيكية في القياس (CTT)

الأهداف (Objectives):

يهدف هذا الفصل في تمكين الطالب من التعرف على نظريات القياس التربوي والنفسي وأهميتها في بناء الاختبارات والمقاييس مع التركيز على النظرية الكلاسيكية للقياس (Classical Test Theory CTT -) من حيث مميزاتها وأوجه قصورها وحدود استخدامها. كما يهدف إلى مساعدة الطالب على فهم مفهوم الخطأ المعياري للقياس وكيفية توظيفه في تقدير دقة الاختبار وموثوقية نتائجه بما يعزز قدرته على تقييم الاختبارات وتحليل نتائجها بصورة علمية دقيقة.

المحتوى (Contents):

- 1- نظريات القياس التربوي والنفسي
- 2- النظرية الكلاسيكية للقياس (CTT) Classical Test Theory
- 1-2- مميزات النظرية الكلاسيكية للقياس في بناء الاختبارات
- 2-2- أوجه قصور وعيوب النظرية الكلاسيكية للقياس
- 3- الخطأ المعياري للقياس (Standard Error of Measurement)

تمهيد:

يُعتبر فهم نظريات القياس التربوي والنفسي خطوة أساسية لفهم طبيعة الاختبارات ودقتها في قياس الخصائص الإنسانية. وتُعد النظرية الكلاسيكية للقياس (Classical Test Theory - CTT) من أقدم وأوسع النماذج استخدامًا حيث وضعت الأسس الأولى لفهم العلاقة بين الدرجة الحقيقية والدرجة الملاحظة وأسهمت في بناء العديد من الاختبارات التربوية والنفسية. وتمتاز هذه النظرية بالبساطة وسهولة التطبيق لكنها في المقابل تعاني من بعض أوجه القصور التي حدّت من قدرتها على مواكبة التطورات الحديثة. ومن المفاهيم الأساسية التي تناولتها النظرية الكلاسيكية مفهوم الخطأ المعياري للقياس الذي يساعد على تقدير دقة النتائج وضبط الأحكام المستخلصة من درجات الأفراد. وفي هذا الفصل سنتناول بالتفصيل مميزات وأوجه قصور النظرية الكلاسيكية بالإضافة إلى الخطأ المعياري للقياس على أن نتطرق في فصول لاحقة إلى نظرية الاستجابة للمفردة (Item Response Theory - IRT) بوصفها أحد التطورات الحديثة في ميدان القياس التربوي والنفسي.

1- نظريات القياس التربوي والنفسي:

تخضع نظريات القياس سواء في العلوم التربوية أو العلوم النفسية لكثير من المفاهيم الرياضية والإحصائية في أساسها وفي إجراءاتها وفي طرق استنتاج النتائج للتأكد من مطابقتها للنماذج النظرية الشائعة وتأتي أهمية نظريات القياس في النقاط الآتية:

أ- تساعد في تحديد مصادر خطأ القياس ومستوى التأثير التي تحدثه هذه المصادر.

ب- تمثل الإطار الذي يتم من خلاله البحث عن العوامل التي تؤثر على الدرجة التي يحصل عليها المفحوص ويكون لهذا الإطار مجموعة من الفروض وقد تتحقق هذه الفروض فتكون النتائج المبنية عليها واقعية وممثلة للحقيقة وقد تكون غير واقعية ومن ثم ستكون الاستنتاجات غير واقعية وخاطئة، ونظرا لأن جميع عمليات القياس يشوبها بعض الخطأ فإن العلماء يبحثون باستمرار عن طرق لزيادة دقة القياس ولتحقيق ذلك فإنهم كثيرا ما يحصلون على متوسط عدد مرات القياس التي تم تعريفها بواسطة عدة معايير. ويفيد هذا المتوسط كتقدير للقياس الحقيقي أو المثالي الذي يمكن أن نحصل عليه نظريا من جميع حالات القياس لصفة من الصفات.

ج- نظريات القياس تساعدنا في بناء أدوات القياس وفي تفسير الدرجات الناتجة عن التطبيق فنجد أن علماء القياس يهتمون بمدخلين رئيسيين في بناء أدوات القياس أولهما المدخل التقليدي الذي يعتمد على النظريات الكلاسيكية في بناء الاختبارات والثاني هو المدخل الحديث في بناء الاختبارات والذي يعتمد على نظرية الاستجابة للمفردة الاختبارية (IRT).

2- النظرية الكلاسيكية للقياس (CTT): Classical Test Theory

تُعد نظرية القياس نموذجًا يستخدمه الباحثون في مجال القياس لتفسير العوامل التي تؤثر على الدرجة التي يحصل عليها المفحوص في اختبار معين حيث تُعتبر الدرجة المحصل عليها ليست بالضرورة انعكاسًا دقيقًا لقدرة الفرد أو مستواه الحقيقي في الصفة أو الخاصية المقاسة (الطريبي، 1997، ص34).

تعود أصول النظرية الكلاسيكية للقياس إلى العالم جاليكسنا (Gulliksen) في عام 1950، الذي قام بتطويرها ونشر استخدامها بشكل واسع في مجالي القياس النفسي والتربوي. وقد عُرفت هذه النظرية بعدة مسميات منها: نظرية الدرجة الحقيقية (The True Score Theory)، ونظرية الخطأ في القياس (Theory of Measurement Error)، ونظرية الدرجة العددية الصحيحة، وكذلك نظرية الاستجابة للاختبار (Test Response). تُعتبر هذه النظرية من أقدم وأهم الطرق المستخدمة في القياس النفسي والتربوي وانتشر تطبيقها في بناء وتطوير مختلف أنواع الاختبارات سواء العقلية كاختبارات الذكاء والاستعدادات والقدرات الطائفية أو الخاصة، أو الميول والاتجاهات ومقاييس الشخصية وغيرها، وكذلك في الاختبارات المدرسية التحصيلية. وتعتمد النظرية في تفسيرها للثبات على وجود صور متكافئة للمقاييس المستخدمة (أمين، 2009، ص68).

ولغرض تفسير مدى صحة الدرجة التي يحصل عليها المفحوص تستند هذه النظرية إلى مجموعة من الافتراضات الأساسية وهي:

أ- الدرجة الملاحظة (س) (Observed Score) في الاختبار هي عبارة عن مجموع جزئين هما الدرجة الحقيقية (ح) (True Score) ودرجة خطأ القياس (خ) (Measurement Error)، ويمكن التعبير عليها بالمعادلة التالية:

$$\text{الدرجة الملاحظة} = \text{الدرجة الحقيقية} + \text{درجة خطأ القياس}$$

ب- لا يمكن قياس الدرجة الحقيقية بل يمكن الاستدلال عليها أو تقديرها من خلال الدرجة الملاحظة، وذلك بحساب متوسط الدرجة الملاحظة أي أن الدرجة الحقيقية للفرد تساوي متوسط الدرجات الملاحظة التي يحصل عليها الفرد عندما يطبق عليه عدد لا نهائي من نسخ الاختبارات المتكافئة أي تقيس نفس البعد وبنفس الصعوبة والخصائص الإحصائية، وبالتالي يكون متوسط خطأ القياس يساوي صفراً؛

ج- التوزيع الاعتدالي لدرجات الأفراد في مثل هذه الاختبارات؛

د- لا يوجد ارتباط بين الدرجات الحقيقية (ح) وأخطاء القياس (خ) للمفحوصين على الاختبار؛

هـ- لا يوجد ارتباط بين أخطاء القياس لاختبارين مختلفين على نفس المشاركون؛

و- درجات الخطأ على الاختبار الأول (خ1) لا ترتبط بالدرجة الحقيقية على الاختبار الآخر (خ2)؛

ز- يمكن بناء اختبارين متكافئين (متوازيين) يقيسان نفس السمة بحيث تتساوى درجاتهما في

المتوسط، والانحراف المعياري، والتباين، وصعوبة المفردات الاختبارية وتمييزها؛

ح- خطأ القياس هو خطأ عشوائي يحدد دقة القياس أو ما يسمى بثبات الاختبار، بينما الخطأ المنتظم

هو خطأ يتعلق بصدق القياس أو صدق الاختبار؛

ط- تفترض هذه النظرية تساوي تباين أخطاء القياس لجميع المشاركين الذين يطبق عليهم الاختبار أو

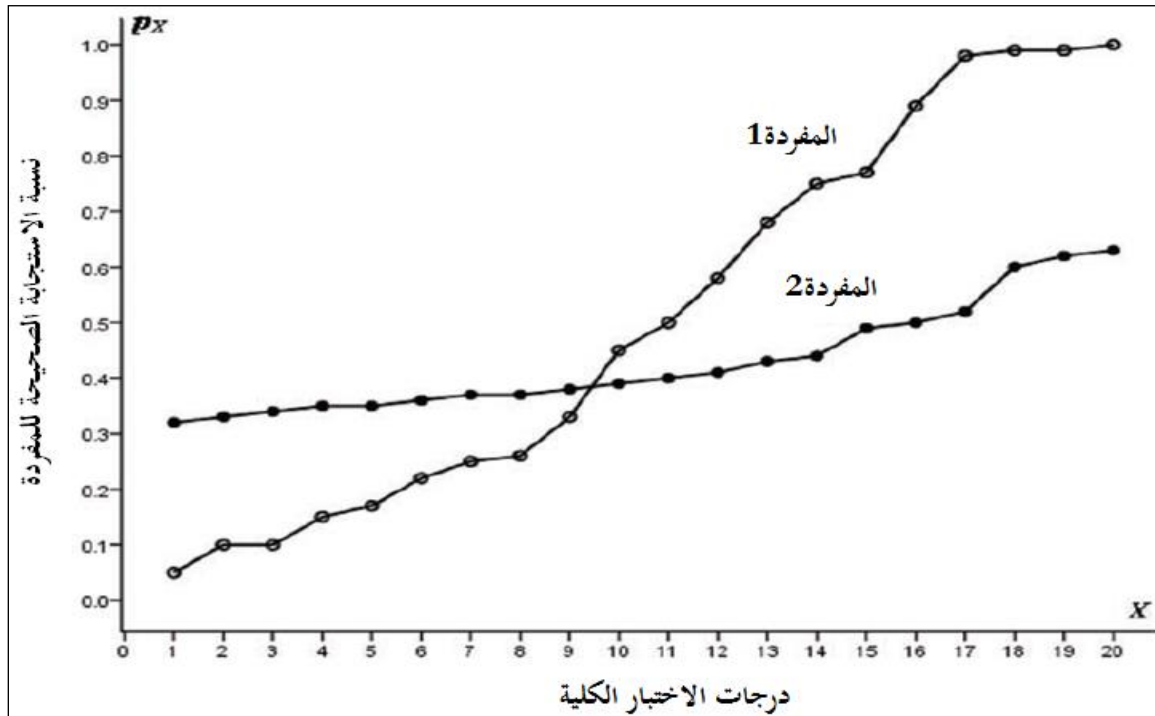
أداة القياس (محاسنة، 2013، ص95) (أمين، 2009، ص69).

كما ترى النظرية الكلاسيكية أن درجة الفرد في الاختبار أو المقياس هي دالة خطية مطردة

بمعنى كلما زادت درجة الفرد على الاختبار زاد مقدار وجود السمة لديه (Cronbach & Gleser, 1970, p116).

كما يوضحه الشكل الآتي:

شكل 12 منحنيات خصائص المفردة لمفردتين وفق النظرية الكلاسيكية في القياس.



يوضح الشكل 12 تمثيل لدرجات الاختبار الكلية للممتحنين على محور السينات (X)، ونسبة

الاستجابة (الإجابة) الصحيحة للممتحنين لمفردتين من مفردات الاختبار على محور الصادات (Px)،

صمم الاختبار لتمييز الأداء العالي مع تحديد الدرجة الكلية تزيد عن 17 على هذا الاختبار (نلاحظ أن المفردة 2 تفرق بشكل جيد نسبياً بين الممتحنين عند الدرجة $X < 17$). ومن خلال ملاحظة انحدار منحني خصائص المفردتين يمكننا تحديد تمييز وصعوبة المفردتين، أي أن الدرجة الكلية على الاختبار تقابل نسبة الإجابة الصحيحة 0.50، ما معناه (احتمال ما نسبته 50% للنجاح في الإجابة الصحيحة على المفردة). كما يتضح أن الدرجة الكلية على الاختبار التي تقابل 0.5 للمفردة 1 هي الدرجة $X=11$ ، وبالتالي مؤشر الصعوبة القائم على منحني خصائص المفردة هو $b_1 = 11$. ومن ناحية أخرى فإن الدرجة الكلية على الاختبار التي تقابل 0.5 للمفردة 2 هي $X=16$ ، مما يشير إلى أنها لمؤشر صعوبة المفردة 2 أي $b_2 = 16$. نستطيع القول بأنها في المتوسط ومنه فإن المفردة 2 أكثر صعوبة من المفردة 1 ($b_2 > b_1$). ومع ذلك لا يمكننا القول إن المفردة 2 هي دائماً أكثر صعوبة من المفردة 1. حيث يتضح من الشكل أن منحني خصائص المفردات للمفردتين يتقاطع تقريباً عند الدرجة $X=9$ أي أن الممتحنين الذين درجاتهم الكلية أقل من الدرجة 9 فإن المفردة 1 تكون بالنسبة لهم أكثر صعوبة أما الممتحنين الذين لديهم درجة كلية أكبر من 9 فإن المفردة 2 تكون بالنسبة إليهم هي الأكثر صعوبة (Dimitrov, 2012, p179).

2-1- مميزات النظرية الكلاسيكية للقياس في بناء الاختبارات:

أ- من السهل على غير المتخصصين في مجال القياس النفسي والتربوي من تحليل بيانات الاختبارات أي المعالجة الإحصائية لمفرداتها (حساب السهولة والصعوبة، وحساب معامل التمييز لمفرداتها) واستيعاب المفاهيم المتعلقة بها.

ب- بساطة الافتراضات التي تقوم عليها النظرية الكلاسيكية للقياس والتي يمكن أن تلائم بيانات مختلف أنواع الاختبارات النفسية والتربوية، بالإضافة إلى أنها لا تحتاج لنماذج رياضية معقدة.

ج- يمكن تحليل بيانات الاختبارات وفق هذه النظرية دون الحاجة لبرامج أو أشخاص متخصصين.

2-2- أوجه قصور وعيوب النظرية الكلاسيكية للقياس:

أ- جميع الخصائص السيكمترية للاختبارات (معامل الصدق، معامل الثبات)، وكذلك الخصائص السيكمترية لمفردات الاختبار مثل (معامل الصعوبة، معامل التمييز) تعتمد في خصائصها على عينة الأفراد التي يطبق عليها الاختبار.

ب- تفترض هذه النظرية أن درجات الاختبار (والتي تمثل السمة/ القدرة) المقاسة هي دالة خطية مطردة بمعنى أنه كلما زادت درجة المفحوص دل ذلك على زيادة السمة/ القدرة المقاسة ولكن الواقع

يشير إلى أن بعض الأفراد ذوي القدرة العقلية العالية يحصلون أحياناً على درجات منخفضة في الاختبارات والعكس صحيح.

ج- تفترض هذه النظرية أن الاختبارات تقيس متغير أحادي البعد هذا الفرض صحيح في العلوم الفيزيائية لأنها تقيس شيئاً واحداً فقط كالطول، الوزن، الحجم، الخ. إلا أن اختبارات الذكاء قد تقيس متغيرات أخرى بجانب الذكاء كثرء البيئة، أو الإمكانات المادية والاجتماعية وغيرها. وبالتالي يصعب تقدير أحادية البعد في كثير من الاختبارات العقلية.

د- اعتماد الدرجة الكلية للاختبار على مفردات (أسئلة) هذا الاختبار.

هـ- تغير تكوين الاختبار ومعنى مفردات الاختبار بمرور الزمن.

و- يعتمد تعريف ثبات الاختبارات على مفهوم الاختبارات المتكافئة رغم أن مفهوم التكافؤ يصعب تحقيقه عملياً.

3- الخطأ المعياري للقياس (Standard Error of Measurement):

يعد الخطأ المعياري من الخصائص السيكومترية الأساسية للاختبارات حيث أنه يستند إليه في تفسير كل درجة من درجاتها. فالخطأ المعياري للقياس يعد عاملاً أساسياً في تقرير نتائج الاختبارات والمقاييس وتفسيرها وهو مرتبط ارتباطاً وثيقاً بمفهوم الثبات ولكنه أيسر فهماً لمستخدمي الاختبارات والمختبرين ولا يتأثر كثيراً بنتائج الدرجات على عكس الثبات الذي يعتمد على الثبات (التباين). وتعتبر النظرية السيكومترية للقياس أن الخطأ المعياري ثابت بالنسبة لمجتمع معين من الأفراد المختبرين (علام، 2000، ص176). ويمكن التعبير عن ثبات الاختبار من حيث الخطأ المعياري للقياس (SEM) ويسمى أيضاً الخطأ المعياري للدرجة. وهذا المقياس يناسب بدرجة جيدة تفسير الدرجات منفردة ولذلك فإنه بالنسبة لكثير من الأغراض الاختبارية يكون أكثر فائدة من معامل الثبات. لذلك إذا أردنا المقارنة بين ثبات "اختبارات مختلفة" فإن معامل الثبات يكون هو المقياس الأفضل، ولتفسير "الدرجات منفردة" فإن الخطأ المعياري للقياس يكون أكثر ملاءمة. فالدرجة الحقيقية ثابتة والدرجة الملاحظة تتذبذب حولها. والخطأ المعياري للقياس يستخرج وفق المعادلة الآتية:

$$SEM = SD\sqrt{1-R}$$

حيث أن:

SEM = قيمة الخطأ المعياري للقياس؛

SD = قيمة الانحراف المعياري؛

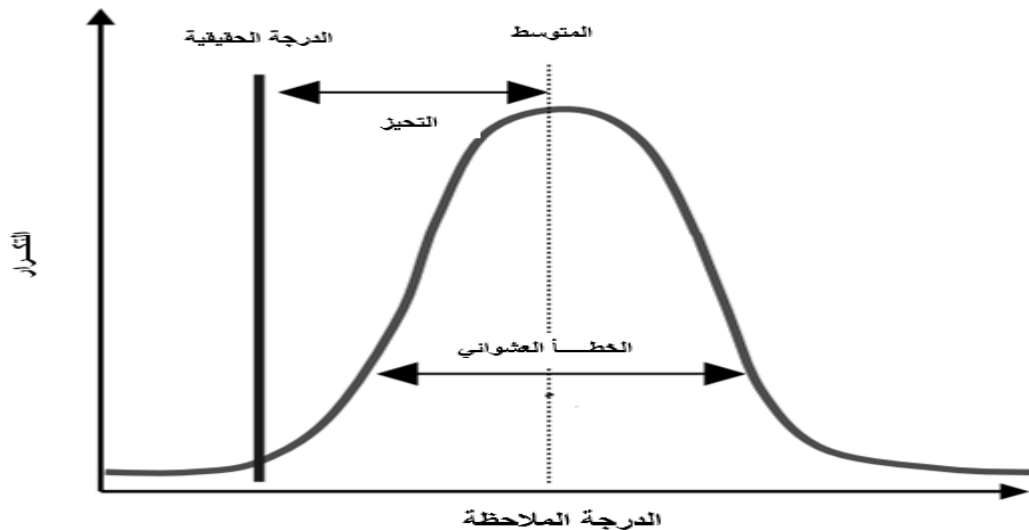
R = قيمة معامل الثبات.

فإذا افترضنا أن قيمة معامل ثبات درجات الاختبار تساوي (0.80)، وقيمة الانحراف المعياري للدرجات الملاحظة (6.3)، فإن قيمة الخطأ المعياري للقياس يساوي (2.82).

وبالتالي فإن حصل شخص ما في الاختبار على الدرجة (12) مثلاً فإنه يمكننا أن نستدل على أن درجته الحقيقية سوف تقع بين 12 - 2.82؛ و12 + 2.82، أي بين الدرجتين 9.18؛ و14.82 وثقتنا في هذا الاستدلال 68% تقريباً. كذلك يمكننا أن نستدل بدرجة ثقة 95% تقريباً من خلال ضرب قيمة الخطأ المعياري للقياس في قيمة مجال مستوى الثقة (1.96) على أن درجته الحقيقية سوف تقع بين (5.53 ±) أي بين 6.46؛ و17.53. كما يمكننا أن نستدل بدرجة ثقة 99% تقريباً كذلك، من خلال ضرب قيمة الخطأ المعياري للقياس في قيمة مجال مستوى الثقة (2.58) على أن درجته الحقيقية سوف تقع بين (7.27 ±) أي بين 4.73؛ و19.27 (أنا ستازي، 2015، ص142).

غير أنه لا يمكن افتراض أن معاملات الثبات أو أخطاء القياس تظل ثابتة عندما تباين مستوى القدرة تبايناً كبيراً والفروق في معاملات الثبات تظل ثابتة عندما يتم حساب أخطاء القياس عند مستويات مختلفة لنفس الاختبار. وأحد الحلول الشاملة لهذه المشكلة تقدمها أساليب نظرية الاستجابة للمفردة الاختبارية في تحليل المفردات. وبمراعاة هذه الأساليب مدى متسعاً من القدرة فإنها تقدم وسيلة للتعبير عن دقة قياس الاختبار كدالة لمستوى القدرة وينتج هذا الاجراء "منحنى معلومات الاختبار" الذي يعتمد فقط على المفردات المتضمنة في الاختبار ويسمح بتقدير خطأ القياس عند كل مستوى قدرة معين (أنا ستازي، 2015، ص142).

شكل 13 يوضح العلاقة بين الدرجة الحقيقية والدرجة الملاحظة والخطأ العشوائي والخطأ المنتظم.



- أ- **الدرجة الملاحظة (Observed Score):** هي الدرجة التي يحصل عليها الفرد فعليًا عند أداءه للاختبار وهي تتأثر بعدة عوامل ولا تمثل بالضرورة مستوى الفرد الحقيقي بدقة.
- ب- **الدرجة الحقيقية (True Score):** تمثل المستوى الحقيقي أو القدرة الفعلية للفرد في السمة أو الخاصية التي يقيسها الاختبار وهذه الدرجة هي التي نرغب في معرفتها وقياسها.
- ج- **الخطأ العشوائي (Random Error):** هو الانحراف غير المنتظم أو المتغير في الدرجة الملاحظة الناتج عن عوامل غير متوقعة كتقلبات المزاج أو تغيرات بيئة الاختبار أو أخطاء في التقدير. والخطأ العشوائي يمكن أن يزيد أو ينقص الدرجة الملاحظة بشكل غير متوقع.
- د- **التحيز (Bias) أو الخطأ المنتظم (Systematic Error):** هو الانحراف المستمر والمتكرر في الدرجة الملاحظة باتجاه معين بسبب عوامل ثابتة تؤثر على القياس كسوء صياغة الأسئلة أو ظروف معينة غير ملائمة.
- هـ- **المتوسط (Mean):** هو القيمة الوسطى أو المتوسطة لمجموعة درجات ملاحظة عبر عدة مرات قياس أو عينة من المفحوصين ويستخدم في تقدير الدرجة الحقيقية عند تكرار القياس. ويمكن تمثيل ذلك بالعلاقة الرياضية التالية:

$$\text{الدرجة الملاحظة} = \text{الدرجة الحقيقية} + \text{التحيز} + \text{الخطأ العشوائي}$$

في النظرية الكلاسيكية للقياس يُفترض عادةً أن متوسط الخطأ العشوائي يساوي صفرًا (أي أنه لا يميل باتجاه معين)، وبالتالي فإن التحيز هو الذي يؤدي إلى انحراف مستمر في القياس (Cohen, et al, 2013, p.175).

- **مثال توضيحي:**

لنفترض أن طالب أجرى اختبارًا للذكاء وحصل على درجة ملاحظة (Observed Score) تساوي 80.

- **الدرجة الحقيقية (True Score):** هي القدرة الحقيقية للطالب لنفترض أنها تساوي 85.

- **التحيز (Bias):** لنفترض أن هناك خطأ منهجي في الاختبار يجعل الدرجات تنخفض دائمًا بمقدار 3 نقاط، مثلاً (أسئلة صعبة بشكل غير عادل).

- **الخطأ العشوائي (Random Error):** بسبب حالة الطالب في يوم الاختبار (تعب، ضغط نفسي، ضوضاء في القاعة) انخفضت درجته 2 نقطة أخرى عشوائيًا.

- **كيف تظهر لنا الدرجة الملاحظة؟**

الدرجة الملاحظة = الدرجة الحقيقية + التحيز + الخطأ العشوائي، وبالتعويض في المعادلة نحصل على التالي: $2 - 3 - 85 = 80$

إذا الطالب حصل فعلياً على الدرجة 80 لكن قدرته الحقيقية هي الدرجة 85 والفرق يرجع إلى تحيز الخطأ المنهجي (-3) وتأثير الخطأ العشوائي (-2).

- دور المتوسط: لو كررنا الاختبار مرات كثيرة على نفس الطالب (تحت ظروف مختلفة) فإن تأثير الخطأ العشوائي يتغير (أحياناً يزيد الدرجة، أحياناً ينقصها) لذا عندما نحسب متوسط الدرجات الملاحظة عبر هذه المحاولات المتعددة، فإن تأثير الخطأ العشوائي يميل للصفر ويختفي ويبقى التحيز فقط لذلك:

متوسط الدرجات الملاحظة \approx الدرجة الحقيقية + التحيز

ولو اختفت كل التحيزات (بتصميم اختبار أفضل)، يصبح:

متوسط الدرجات الملاحظة \approx الدرجة الحقيقية

- والسؤال المطروح هو كيف يمكننا تقليل الخطأ العشوائي والتحيز (الخطأ المنتظم) عملياً في تصميم الاختبارات؟

1- تقليل الخطأ العشوائي: الخطأ العشوائي هو تقلبات غير متوقعة تؤثر على الدرجات ولتقليله يمكن اتباع عدة استراتيجيات، كالاتي:

- **زيادة عدد البنود:** كلما زاد عدد الأسئلة أو البنود في الاختبار تقل احتمالية أن يؤثر خطأ عشوائي في النتيجة النهائية لأن الأخطاء تتوزع وتتوازن فيما بينها.

- **توحيد ظروف تطبيق الاختبار:** توفير بيئة هادئة وثابتة كتقليل الضوضاء وضمان الإضاءة المناسبة وتقليل التشتيب يساعد في تقليل تأثير العوامل الخارجية على أداء الممتحن.

- **التدريب والتوجيه:** إرشاد المشاركين حول كيفية أداء الاختبار بشكل صحيح مع التأكد من فهم التعليمات جيداً.

- **إعادة إجراء عملية القياس:** إجراء اختبار مكرر في أوقات مختلفة بحيث يمكن تقدير الدرجة الحقيقية بالمتوسط وتقليل تأثير الخطأ العشوائي.

2- تقليل التحيز (الخطأ المنتظم): التحيز أو الخطأ المنتظم هو خطأ مستمر ينتج عن تصميم أو صياغة غير مناسبة ولتقليله، يجب:

- صياغة الأسئلة بعناية: التأكد من أن جميع البنود واضحة ومحايدة وغير متحيزة ثقافياً أو للنوع الاجتماعي (الجنس) أو لغوياً.
- اختبار البنود مسبقاً (اختبار تجريبي): تجريب الاختبار على عينة مماثلة للمجتمع المستهدف لتحليل أداء البنود واكتشاف البنود المتحيزة أو الصعبة بشكل غير عادل.
- استخدام خبراء متعددين: إشراك مجموعة من المختصين في صياغة البنود لمراجعة المحتوى والتأكد من توافقه مع الهدف المرجو.
- تحليل البيانات بعد التطبيق: استخدام أساليب إحصائية، مثل: التأكد من الخصائص السيكومترية لمفردات الاختبار أو استخدام نماذج نظرية الاستجابة للمفردة لاكتشاف البنود التي قد تسبب تحيزاً وتصحيحها أو حذفها (Anastasi and Urbina, 1997, p.127).

الفصل الخامس

نظرية الاستجابة للمفردة الاختبارية (IRT)

الأهداف (Objectives):

يهدف هذا الفصل إلى التعريف بنظرية الاستجابة للمفردة (IRT) من خلال تتبع تطورها التاريخي والوقوف عند الافتراضات الأساسية التي تقوم عليها النماذج الاختبارية أحادية البعد. كما يسعى إلى توضيح الأنواع المختلفة لهذه النماذج مع إبراز مفهوم الموضوعية في نموذج راش والجانب السيكولوجي الكامن فيه، إضافة إلى شرح وحدات القياس المستخدمة ضمن هذا النموذج. كما سيتعرض كذلك لأساليب تقدير بارامترات النماذج أحادية البعد وإلى الاختبارات التحصيلية التي بُنيت وفق هذه النظرية مع التطرق إلى آلية تدريج الاختبار على أساس نموذج راش. ولا يقتصر الهدف على الجانب التطبيقي فحسب بل يشمل أيضاً كيفية التحقق من مدى مطابقة البيانات للاختبارات للافتراضات الأساسية للنموذج وصولاً إلى مناقشة دالة المعلومات (Information Function) وما توفره من مؤشرات إحصائية مهمة لتقييم الاختبار.

المحتوى (Contents):

- 1- التطور التاريخي لنظرية الاستجابة للمفردة (IRT)
- 2- افتراضات نماذج نظرية الاستجابة للمفردة الاختبارية أحادية البعد
- 3- أنواع نماذج الاستجابة للمفردة أحادية البعد
- 4- معنى الموضوعية في نموذج راش
- 5- الجانب السيكولوجي في نموذج راش
- 6- وحدات القياس المستخدمة في نموذج راش
- 7- تقدير بارامترات النماذج أحادية البعد
- 8- الاختبارات التحصيلية المعتمدة في بنائها على نظرية الاستجابة للمفردة
- 9- تدريج الاختبار وفق نموذج راش
- 10- كيفية التحقق من حسن مطابقة البيانات الاختبارية الافتراضات نموذج راش
- 11- دالة المعلومات (Information Function)

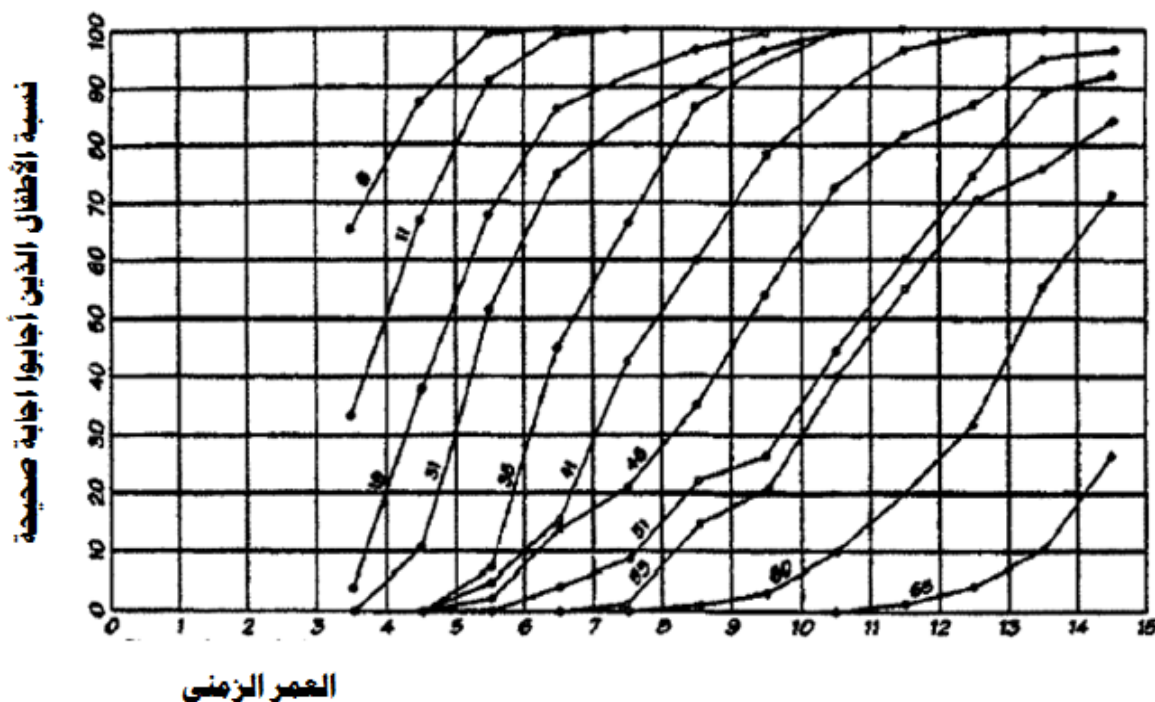
تمهيد:

ترتكز نظرية الاستجابة للمفردة (Item Response Theory – IRT) على مفاهيم ومبادئ تختلف جذرياً عن تلك التي تقوم عليها النظرية الكلاسيكية في القياس النفسي والتربوي (Classical Test Theory – CTT). إذ تعتمد منهجية القياس في هذه النظرية على نماذج رياضية احتمالية تهدف إلى ضبط العوامل المربكة التي قد تؤثر في المقارنات بين الدرجات وذلك من خلال نمذجة رياضية لبارامترات العملية الاختبارية. ويُنسب الفضل في عرض هذه المنهجية بصورة متكاملة إلى لورد ونوفيك (Lord & Novick, 1968) اللذان قدماها بالتفصيل في كتابهما المرجعي "النظريات الإحصائية لدرجات الاختبارات العقلية" (Statistical Theories of Mental Test Scores). وقد اقترحا نماذج رياضية تصف ما يحدث عند استجابة فرد معين لمفردة اختبارية بما يسمح بالتنبؤ بأدائه أو تفسيره استناداً إلى خصائص محددة تُعرف باسم السمات. تُستخدم هذه النماذج لتقدير بارامترات تخص المفردات وأخرى تخص الأفراد فيما يتعلق بهذه السمات. ومن خلال القيم التقديرية لهذه البارامترات يمكن تفسير أداء كل فرد في الاختبار. ونظراً لصعوبة ملاحظة هذه السمات أو قياسها بشكل مباشر يتم الاستدلال عليها عبر القيم التقديرية والتي يُطلق عليها عادةً السمات الكامنة (Latent Traits) أو القدرات (Abilities) (علام صلاح، 2013، ص47).

1- التطور التاريخي لنظرية الاستجابة للمفردة (IRT):

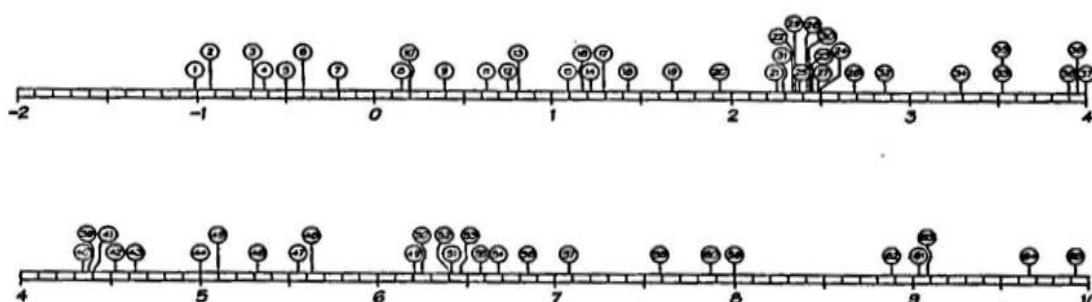
تعود الإشارات الأولى إلى نظرية الاستجابة للمفردة إلى الورقة البحثية التي نشرها "لويس ثرستون" (Louis Leon Thurstone) عام 1925 بعنوان "منهجية للمقايسة الاختبارية النفسية والتربوية" (A Method of Scaling Psychological and Educational Tests). في هذه الورقة قدم "ثرستون" حلاً لمشكلة كانت تشغل الباحثين النفسيين آنذاك وهي إيجاد أفضل طريقة لمعايرة القدرة العقلية من خلال مفردات مقياس بينيه – سيمون للأطفال على تدرج العمر الزمني. وقد قام "ثرستون" برسم العلاقة لعينة كبيرة بين نسبة الأطفال الذين نجحوا في أداء مهام اختبار "بينيه – سيمون" (على المحور الرأسي) والعمر الزمني (على المحور الأفقي). وأسفرت النتائج عن منحنيات على شكل الحرف S توضح أن نسبة الاستجابة الصحيحة في مقياس "بينيه – سيمون" تزداد كلما ارتفع العمر الزمني. وتتشابه هذه المنحنيات الموضحة في الشكل (14)، مع منحنيات الترجيح اللوغاريتمي في نموذج راش والتي تُعرف باسم "منحنيات القدرة".

شكل 14 رسم بياني لثورستون (Thurstone, 1925) يوضح نسبة الاجابة الصحيحة للأطفال على مفردات اختبار "بينيه - سيمون" للأطفال تبعاً لمتغير العمر الزمني (Wilson, 2005, p89).



قام "ثرستون" بعد ذلك بتحديد مواقع المفردات الـ (65) في اختبار الذكاء لـ "بينيه - سيمون" على مقياس العمر حيث اعتبر نقطة الأصل الاعتباطية تساوي صفر ووحدة القياس المسافية تساوي (1) انحراف معياري. ويعرض الشكل 15 التدرج العمري للمفردات الناتج عن عملية المعايرة التي أجراها حيث يعكس موضع كل نقطة في الشكل مستوى صعوبة المفردة تبعاً للعمر الزمني. وبذلك قدّم "ثرستون" لمستخدمي اختبار "بينيه - سيمون" أساساً علمياً لترتيب مفردات الاختبار وفق تزايد صعوبتها فضلاً عن تحديد العمر العقلي للفرد من خلال عدد الاستجابات الصحيحة وفقاً لما وضعه "بينيه" (ديانا، 2010، ص36).

شكل 15 المواضع المقدرة للمفردات على مقياس للعمر ذي نقطة أصل اعتباطية ووحدات قياس متساوية (Louis, 1925, p449).



يمثل الشكل 15 مقياسًا مطلقًا لمفردات اختبار "بينيه" حيث تعبّر المسافة الفترية بين الأرقام عن الانحراف المعياري لمستوى ذكاء الأطفال في عمر ثلاث سنوات ونصف على هذا الاختبار. وتشير نقطة الأصل إلى متوسط ذكاء الأطفال في نفس العمر على اختبار "بينيه". وبذلك يمكن اعتبار "ثرستون" من أوائل من قدّموا إسهامات مبكرة تتقاطع مع أفكار نظرية الاستجابة للمفردة قبل أن تُعرف باسمها الحالي وذلك نظرًا لما يشتركان فيه من نقاط أساسية، وهي:

أ- **نموذج الاستجابة الاحتمالي:** كلاهما يضع نموذجًا للاستجابة يقوم على احتمالية الإجابة الصحيحة باعتبارها دالة لمتغير مستمر يقيس سمة لدى الفرد إذ يُعرف هذا المتغير في نظرية الاستجابة للمفردة بـ "القدرة" بينما عند "ثرستون" يُقاس بـ "العمر الزمني".

ب- **بارامترات المفردة والمقياس:** يفترض كلا النموذجين وجود بارامترات تصف خصائص المفردة فموضع المفردة على متصل العمر عند "ثرستون" يقابله في نظرية الاستجابة للمفردة بارامترات المقياس المرتبطة بدالة التوزيع الطبيعي.

ج- **موضوع التحليل:** في كلا الإطارين يتم تمثيل موقع المفردة (Item Location) واستجابات الفرد (Respondent Values) كنقاط على مقياس المتغير الكمي. وتكمن أهمية ذلك في أن تصحيح مقياس "بينيه" لا يعتمد على المجموع الكلي للاستجابات الصحيحة فحسب بل على عدد الاستجابات الصحيحة للمفردات المطابقة للتدرج العمري الذي يمكن للطفل أدائه (ديانا، 2010، ص37).

2- افتراضات نماذج نظرية الاستجابة للمفردة الاختبارية أحادية البعد:

ترتكز نماذج الاستجابة للمفردة أحادية البعد على مجموعة من الافتراضات القوية التي يجب أن تتوافر في بيانات الاختبارات لضمان الحصول على نتائج موثوقة. ويُقيّم مدى ملائمة النموذج للبيانات من خلال التحقق من تحقق هذه الافتراضات إذ إن استيفاءها يسهم في تحقيق دقة وإحكام أكبر في عملية القياس. وفيما يلي عرض لأهم هذه الافتراضات الأساسية:

2-1- أحادية البعد Unidimensionality:

تفترض نماذج الاستجابة للمفردة أن هناك قدرة واحدة تفسر أداء الفرد في الاختبار. إلا أن هذا الافتراض قد يتحقق مع مجموعة من المفحوصين بينما لا يتحقق بالقدر نفسه مع مجموعة أخرى. فعلى سبيل المثال في حالة المسائل اللفظية في الرياضيات يُتوقع أن يلعب الفهم القرائي دورًا مهمًا في حل هذا النوع من المسائل. فإذا كان أفراد العينة يتمتعون بمستوى جيد من الفهم القرائي فإن الأداء في الاختبار سيكون معتمدًا بالدرجة الأولى على القدرة الرياضية فقط وبالتالي يُعد الاختبار أحادي البعد. أما إذا كانت مجموعة المفحوصين منخفضة الكفاءة في الفهم القرائي فإن الأداء في

الاختبار سيتأثر بكلٍ من القدرة الرياضية والقدرة على الفهم القرائي معًا مما يعني أن الأداء لا يمكن تفسيره من خلال بعد واحد فقط وبالتالي لا يحقق الاختبار شرط الأحادية البعد.

استخدم هامبلتون وتروب (Hambleton & Traub, 1973) أسلوب التحليل العاملي للتحقق من افتراض أحادية البعد. كما اقترح لورد ونوفيك (Lord & Novick, 1968) استخدام معاملات الارتباط الرباعية بدلاً من معاملات ارتباط Φ (Phi) في اختبار أحادية البعد وذلك بسبب المشكلات المرتبطة بمعاملات Φ ، ومنها أنها قد تكون غير ثابتة عبر العينات المختلفة. واقترح ريكاس (Reckase, 1979) طريقة أخرى للتحقق من أحادية البعد تتمثل في استخدام مصفوفة معاملات الارتباط البينية للمفردات، بحيث تُحسب القيم الجذرية الكامنة (Eigenvalues) وترتب من الأكبر إلى الأصغر ثم تمثل بيانياً. ويمكن من خلال هذه الرسوم البيانية تحديد ما إذا كان هناك عامل بارز مهيم أم لا. كما استخدم كل من ماكدونالد (McDonald, 1980) وهاتي (Hattie, 1981) أسلوب تحليل البواقي والتحليل العاملي غير الخطي للتحقق من أحادية البعد في المفردات الاختبارية. وأوضح هامبلتون وسواميناثان (Hambleton & Swaminathan, 1985) أن تحليل البواقي يكون أكثر وضوحاً عند استخدام اختبارات مميزة لاختلاف كمية التباين المفسر.

ويشير بي هيوشين (Pei-Huachen, 2005) إلى أن تحقيق افتراض أحادية البعد بشكل كامل يُعد أمراً صعباً وذلك لوجود مجموعة من العوامل الشخصية والمعرفية إضافة إلى العوامل المرتبطة بالاختبار نفسه مثل التعليمات والتوجيهات والتي تؤثر مجتمعة في أداء الأفراد حتى وإن كان تأثيرها محدوداً. وتشمل هذه العوامل: مستوى الدافعية، وقلق الاختبار، والقدرة على الإجابة بسرعة، إلى جانب المهارات المعرفية التي يمكن قياسها من خلال مفردات الاختبار (ميمي السيد، 2014، ص56).

2-2- الاستقلال الموضوعي Local Independence:

يقصد بافتراض الاستقلال الموضوعي أن احتمال إجابة الفرد إجابة صحيحة على مفردة اختبارية يكون مستقلاً عن ناتج إجابته على أي مفردة أخرى في الاختبار وذلك عند ضبط كل من القيمة التقديرية لقدرته والقيمة التقديرية لصعوبة المفردة. وبذلك يشير الاستقلال الموضوعي إلى أن مفردات الاختبار لا تكون مترابطة إحصائياً بعد ضبط هذين العاملين. وربما يبدو هذا الافتراض متعارضاً مع أحد مبادئ نظرية القياس الكلاسيكية التي تؤكد أهمية الاتساق الداخلي للاختبارات حيث يُفترض أن زيادة الارتباط بين درجات المفردات يؤدي إلى ارتفاع معامل الثبات. إلا أن هذا التعارض يزول عند النظر إلى النماذج المعاصرة أحادية البعد إذ تكون استجابات الأفراد على المفردات مشروطة

بمستوى قدرتهم وصعوبة المفردات بحيث لا تبقى أي ارتباطات بين المفردات بعد هذا الإشراف (Conditioning). وهذا بدوره يشير إلى أن مفردات الاختبار تقيس بعداً واحداً. ولتحقيق هذا الافتراض ينبغي أن يكون نمط استجابة الفرد (صحيح = 1 أو خطأ = 0) مساوياً لحاصل ضرب احتمالات حصوله على درجة معينة في كل مفردة من مفردات الاختبار. ويتحقق ذلك إذا كانت جميع المفردات تقيس سمة كامنة واحدة أو قدرة واحدة. وبالتالي فإن أحادية البعد تتطلب أن تحقق جميع أزواج المفردات افتراض الاستقلال الموضعي وهو ما يعني أن المفردات تكون مستقلة إحصائياً لأي مجتمع فرعي متجانس من الأفراد فيما يتعلق بالسمة الكامنة. وعليه فإن الاستقلال الموضعي يرتبط أيضاً بعدد السمات أو القدرات التي ينطوي عليها الأداء على المفردات ويُعد مؤشراً على تحقق أحادية البعد في حال كان النموذج المستخدم يقدّر قدرة الفرد على بعد واحد. وعلى الرغم من العلاقة الوثيقة بين الاستقلال الموضعي وأحادية البعد فإنهما ليسا مفهومًا واحدًا. فقد يكون الاختبار متعدد الأبعاد إذا وُجدت أكثر من سمة كامنة؛ فعلى سبيل المثال يمكن أن تتحقق استقلالية المفردات بالنسبة للأفراد المتجانسين في كل من هاتين السمتين على حدة. وبشكل عام يمكن القول إن عدد أبعاد الاختبار يساوي عدد السمات الكامنة اللازمة لتحقيق الاستقلال الموضعي. ومن هذا المنطلق يتضح أن تحقيق افتراض الاستقلال الموضعي قد يكون أمراً صعباً في كثير من الاختبارات ولا سيما الاختبارات العقلية (علام، 2013، ص63).

ويشير حمادنة (2011) في هذا السياق إلى عدد من المؤشرات الإحصائية المستخدمة للكشف عن الارتباط الموضعي بين المفردات ذات الاستجابة الثنائية، ومن أبرزها:

أ- **مؤشر G^2** : يُستخدم للكشف عن الفروق بين القيم المتوقعة والقيم الملاحظة لتكرارات الاستجابات على زوج من المفردات.

ب- **مؤشر فيشر (Fisher Z)**: يعتمد على تحويل الأخطاء الملاحظة إلى أخطاء معيارية ذات متوسط يساوي صفراً بحيث تُعد أي قيمة من هذا المؤشر تتجاوز ± 2 انحراف معياري مؤشراً على أن المفردة قد لا تكون مستقلة موضعياً.

ج- **المؤشر الإحصائي (Q3)**: الذي اقترحه ين (Yen) والذي يقيس درجة عدم الاستقلالية المحلية (Local Dependence - LD) في أنماط استجابات الأفراد. تقوم فكرته على حساب معامل الارتباط بين البواقي لاستجابات الأفراد لأي زوج من المفردات وذلك بعد عزل تأثير النموذج المطبق. وتتم العملية على النحو الآتي:

أ- بعد تقدير معلمات المفردات وقدرات الأفراد يتم حساب القيم المتوقعة لاستجابات كل فرد على جميع المفردات.

ب- يُحسب الباقي لكل فرد في كل مفردة وهو الفرق بين الاستجابة الفعلية والاستجابة المتوقعة.

ج- تُحسب معاملات الارتباط بين قيم البواقي لكل زوج من المفردات عبر جميع الأفراد.

وعندما يكون حجم العينة كبيراً يُتوقع أن يكون معامل الارتباط بين البواقي مساوياً للصفر عند تحقق الاستقلالية المحلية. وبالتالي فإن ارتفاع قيمة مؤشر Q3 يُعد دليلاً على وجود ارتباط موضعي. ويُعتبر مؤشر Q3 معامل ارتباط للبواقي بعد ضبط السمة المقدّرة، وقد أكد زينسكي وهامبلتون وسيريسي (2002) وكذلك أيلّا Ayala (2009، ص.133) على كفاءته العالية في الكشف عن افتراض الاستقلال الموضعي لكونه يختبر العلاقة بين البواقي لأزواج المفردات مباشرة. ولحسابه يجب أولاً تقدير قدرة (θ_i) كل مفحوص ثم استخدامها لتقدير احتمالات الإجابة الصحيحة على المفردات، ومن ثم حساب البواقي والتي يرمز له بالرمز (dij) كما يلي:

$$\text{الباقي } (dij) = \text{الأداء الملاحظ } (xij) - \text{الأداء المتوقع } (pj \widehat{(\theta_i)})$$

(التقي، 2013، ص147)

باعتبار (xij) يمثل العلامة الملاحظة للشخص (i) على المفردة (j) وهي اما واحد عند الاستجابة الصحيحة أو صفر عند الاستجابة الخاطئة. و $pj \widehat{(\theta_i)}$ تمثل احتمال اجابة المفردة (j) اجابة صحيحة من قبل المفحوص (i)، ويمثل مستوى بواقي الشخص والمفردة (k) ويعطي بالمعادلة التالية:

$$dik = xik - pk \widehat{(\theta_i)}$$

وبالأخذ القيمة المطلقة لمؤشر |Q3| تكون قيمة المؤشر ما بين الصفر والواحد الصحيح المطلقة. وكلما اقترب من الصفر فهو أفضل، وترى ين (Yen,1984) أن قيمة هذا المؤشر تميل لان تكون سالبة وخاصة في المقاييس والاختبارات القصيرة وبالتالي لو أخذنا المفردتين (j,k) فإن (Q3) يكون معامل الارتباط بين البواقي لهاتين المفردتين لجميع المفحوصين، ويمكن التعبير عنه كما يلي:

$$Q3jk = rdjdk$$

واعتبر المعيار 1- L / 1- باعتبار (L) عدد المفردات داخل الاختبار مؤشراً على تحقق افتراض الاستقلال الموضعي (Ayala, 2009, p133).

وقد أعتمد ريز (Reese,1999) في الحكم على مستويات الانتهاك بناءً على هذا الأسلوب من خلال القيم التالية: 0.30؛ 0.05؛ 0.01؛ 0.000 بوصفها مستويات للانتهاك الصفري والمتدني والمتوسط والعالي على التوالي. وأشار (Zenisky and Hambleton and Sireci, 2002) إلى أن قيم هذا الاحصائي عادة ما تأخذ قيمة صغيرة وقالوا إن أعلى القيم له التي تم ملاحظتها في الدراسات كانت تحوم حول القيمة 0.10 (النعي، 2011، ص162).

وكقاعدة عامة لتقييم مؤشر الاحصائي (Q3) للحكم على الاستقلال الموضوعي ألا يكون ارتباط أزواج الفقرات من خلال هذا المؤشر (Q3) أكبر من القيمة 0.2 أو أصغر من القيمة -0.2. (Thissen, Chen ;1997) أي اعتبار القيمة المطلقة للمؤشر Q3 أكبر من $|Q3| > 0.2$ مؤشراً للتبعية أو الاعتماد (Ayala, 2009, p133) (Francis et al, 2001, p159-174) (Chrisitine, 2010, p48).

2-3- توازي المنحنيات المميزة للمفردات (Item Characteristic Curve (ICC

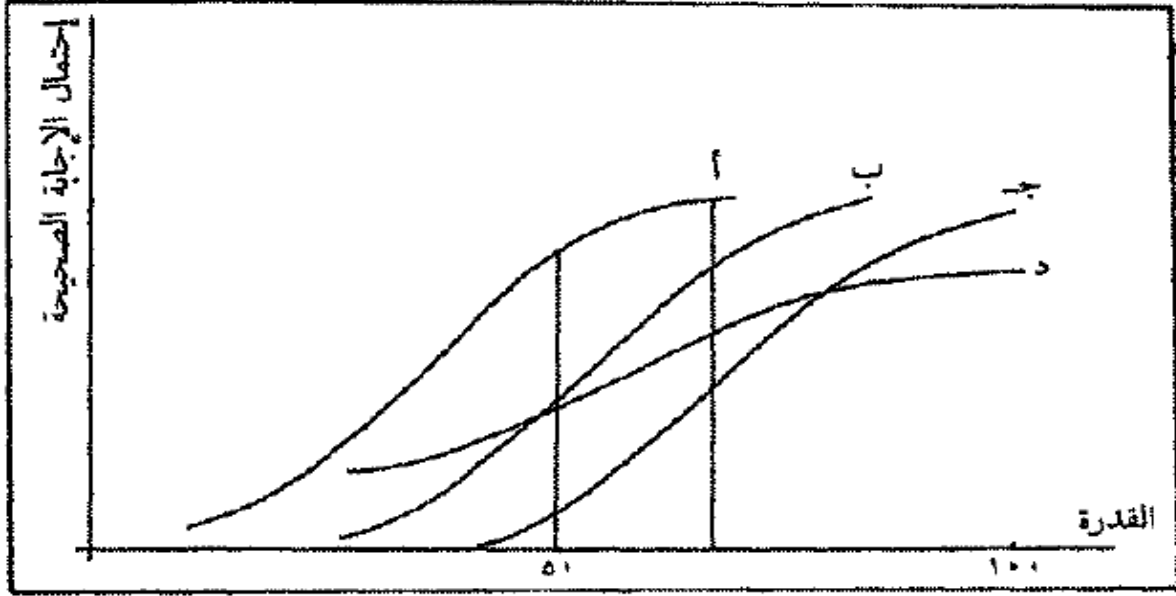
تتفق كلٌّ من النظرية الكلاسيكية والنظرية الحديثة في القياس على فرضية أساسية مفادها وجود متصل للقدرة بحيث يمكن عند معرفة موقع الفرد على هذا المتصل تقدير احتمال إجابته الصحيحة على أي مفردة اختبارية تقيس هذه القدرة. ورغم هذا الاتفاق تختلف النظريتان في شكل الدالة التي تربط موقع الفرد باحتمال الإجابة الصحيحة:

- في النظرية الكلاسيكية يُفترض أن تأخذ الدالة شكل المنحنى الاعندالي في الاختبارات ذات المرجع الجماعي أو شكلاً ملتوياً ذي التواء سلبي في الاختبارات محكية المرجع.
- أما النظرية الحديثة فتفترض أن الدالة تتخذ شكل المنحنى اللوغاريتمي.

وبناءً على ذلك تفترض نماذج نظرية الاستجابة للمفردة وجود دالة مميزة لكل مفردة في الاختبار تعتمد أساساً على موقع كلٍّ من الفرد والمفردة على متصل القدرة وتتخذ شكل المنحنى اللوغاريتمي التراكمي (وليد، 2014، ص17).

ترى أمينة كاظم (1988) أن تساوي قوة التمييز بين مفردات الاختبار يرتبط بتحقيق شرط استقلالية القياس إذ تتقارب الصعوبة النسبية للمفردات عبر مختلف مستويات القدرة. وتُعرّف 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 تبعاً لشكل هذا الانحناء. ويوضح الشكل رقم (6) أربع منحنيات مميزة لأربع مفردات (أ، ب، ج، د)، حيث يمكن ملاحظة ما يلي:

شكل 16 يوضح المنحنيات المميزة لأربع مفردات (أمانة كاظم، 1988، ص72).



أ- تتوازي تقريبا المنحنيات (أ)، (ب)، (ج)، أي أن لها شكلا وانحناء عاما واحدا. ومعنى هذا أن الزيادة المتساوية في مستوى القدرة (مستوى الدرجة الكلية) يقترن بزيادة متساوية تقريبا في احتمال الإجابة الصحيحة على أي من هذه المفردات. وبعبارة أخرى أن قوة تمييز المفردات بين قدرات الأفراد متساوية وهذا ما يختلف بالنسبة للمنحنى (د)؛

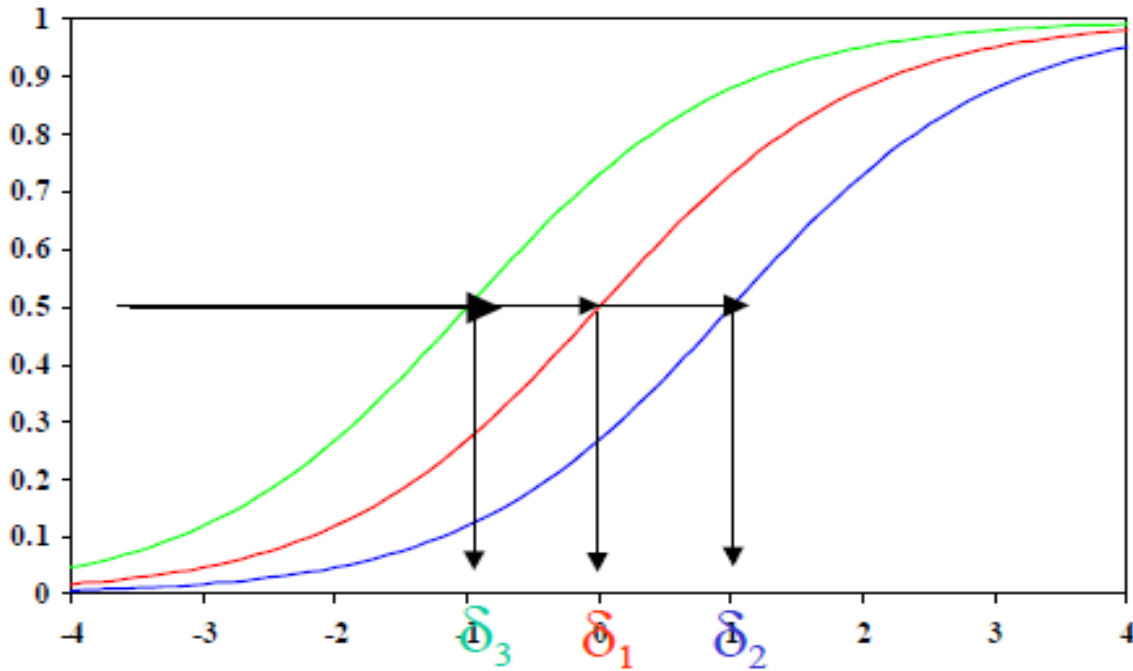
ب- إن فاعلية المفردة (أ) تمتد من المستوى المنخفض من القدرة وحتى المستوى فوق المتوسط مباشرة وليس لهذه المفردة فائدة تذكر للأفراد ذوي المستوى العالي من القدرة ولا تستطيع أن تميز بينهما على هذه القدرة حيث يمكن لهؤلاء الأفراد جميعا الإجابة الصواب على هذه المفردة. أما المفردة (ج) فتمتد فاعليتها من المستوى العالي من القدرة وليس لهذه المفردة فائدة تذكر للأفراد منخفضي القدرة. حيث يخفق هؤلاء جميعا في الإجابة 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، ج) يكون احتمال الإجابة الصواب على المفردة (أ) أكبر دائما من احتمال الإجابة صواب على المفردة (ب) وهذا يكون دائما أكبر من احتمال الإجابة الصواب على المفردة (ج) 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 (ج) 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 ملائمة هذه المفردات للنموذج؛

د- إن احتمال الإجابة الصواب على المفردة (د) يكون أكبر من احتمال الإجابة 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 أيضاً عند مقارنته بالمفردة (ج) حيث تكون المفردة (د) أسهل من المفردة (ج) 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 إخلال بفرض النموذج عن استقلال صعوبة المفردة عند قدرات العينة مما يعني عدم ملائمة هذه المفردة للنموذج (أمينة كاظم، 1988، ص72).

ويلاحظ كذلك أن المنحنيات المميزة لثلاثة مفردات في الشكل 17 التي تم تدرجها باستخدام نموذج (راش) أحادي المعلم لا تتقاطع ولكن تختلف عن بعضها في الصعوبة فقط وتتلاقى فقط عند المحور الأفقي الذي يمثل متصل القدرة أو السمة.

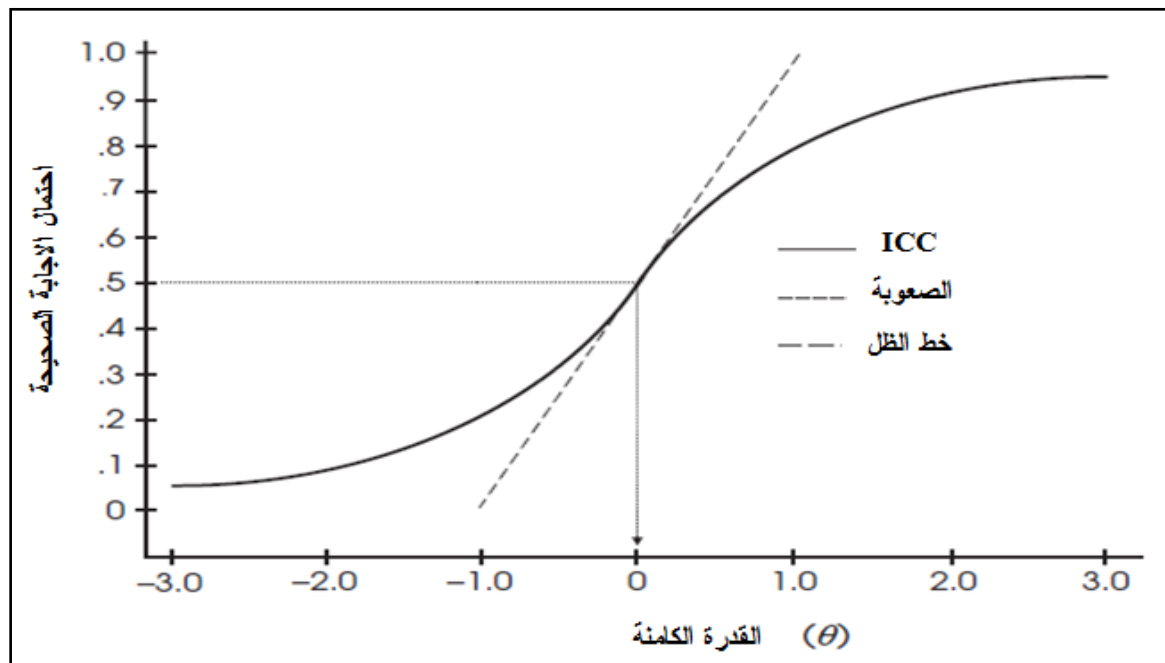
شكل 17 يوضح توازي المنحنيان المميزة للمفردات بحسب نموذج راش أحادي البارامتر (Wu, M, 2007, p31).



يمكن ملاحظة قوة التمييز لأحد المفردات من خلال ازدياد ميل منحنى خصائصه بالقرب من الوسط، أي عند النقطة على متصل القدرة التي يُتوقع أن يكون فيها احتمال الإجابة الصحيحة 0.50. فكلما كان الميل حاداً (Steep Slope) دلّ ذلك على أن أي تغيير نسبي في قدرة الفرد يؤدي إلى تغيير ملحوظ في احتمال إجابته الصحيحة (تمييز مرتفع). أما إذا كان الميل بسيطاً فإن حتى التغيرات

الكبيرة نسبياً في قدرة الفرد لا تؤثر كثيراً على احتمال الإجابة الصحيحة (تمييز منخفض). وبالتالي فإن درجة تمييز المفردة تتناسب طردياً 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 (18) التالي:

شكل 18 يوضح منحنى خصائص المفردة (ICC) والتنبؤ باحتمال الإجابة 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 حيث يُحدد ذلك من خلال ارتفاع منحنى خصائص المفردة على متصل القدرة عند أدنى مستويات القدرة. وتتراوح قيم معامل التخمين نظرياً بين (0 و 1)، بينما تتراوح عملياً بين (0 و 0.35) وما يزيد عن ذلك يُعد غير مقبول (Baker, 2001, p.28). ويمثل هذا المعامل احتمال الإجابة الصحيحة بمحض الصدفة ويتحدد من خلال الخط المقارب الأدنى (Lower Asymptote) في منحنى خصائص المفردة (أبو حطب وآخرون، 1993، ص. 285).

2-5- السرعة Speediness:

ميّز لورد ونوفيك (Lord & Novick, 1968) بين نوعين من الاختبارات: اختبارات السرعة واختبارات القوة. ففي اختبارات السرعة (Speed Tests) تعتمد الدرجة التي يحصل عليها الممتحن على قدرته في الإجابة عن أكبر عدد ممكن من المفردات خلال زمن محدد وثابت حيث تكون الدرجة انعكاساً لمعدل الاستجابة. أما اختبارات القوة (Power Tests) فتعتمد على قدرة الممتحن على الإجابة عن مفردات الاختبار دون التقيد بزمن محدد. ويعرّف فاروق عبد الفتاح موسى (1990) سرعة الأداء في الاختبار بأنها مدى اعتماد درجة الفرد على سرعة إنجازه وتقاس عادة بنسبة الأفراد الذين لم يتمكنوا من الوصول أو الإجابة عن آخر مفردة في الاختبار. وتفترض نماذج الاستجابة للمفردة أن عامل السرعة لا يؤثر في الإجابة عن مفردات الاختبار أي أن عدم تمكن بعض المختبرين من الإجابة يعود إلى انخفاض مستوى قدرتهم وليس إلى تأثير عامل الوقت. ومع أن عامل السرعة لا ينعكس مباشرة على المنحنى المميز للمفردة إلا أن تأثيره قد يكون غير مباشر حيث يدفع بعض الأشخاص إلى التخمين في الإجابة مما يجعل سرعة الأداء في بعض الحالات إحدى القدرات التي يقيسها الاختبار (ميمي، 2014، ص59).

3- أنواع نماذج الاستجابة للمفردة أحادية البعد:

تتعدد النماذج في إطار نظرية الاستجابة للمفردة تبعاً لاختلاف الافتراضات المرتبطة بالبيانات الاختبارية. فقد يُفترض أن الأداء في الاختبار يعتمد على سمة واحدة (أحادية البعد) أو على عدة سمات (متعددة الأبعاد). كما قد يُفترض أن المفردات تختلف في قدرتها على التمييز بين المستويات المختلفة للقدرة أو أن الإجابات الصحيحة في مفردات الاختبار من متعدد قد تتأثر بالتخمين العشوائي أو لا تتأثر كما هو الحال في الأسئلة المفتوحة. كذلك قد يتعلق الأمر بما إذا كانت الدرجات الممنوحة ثنائية أو متعددة الفئات وغيرها من الافتراضات. وبناءً على ذلك يمكن تصنيف هذه النماذج إلى قسمين رئيسيين، هما:

1- نماذج الاستجابة للمفردة أحادية البعد (Unidimensional Models).

2- ونماذج الاستجابة للمفردة متعددة الأبعاد (Multidimensional Models).

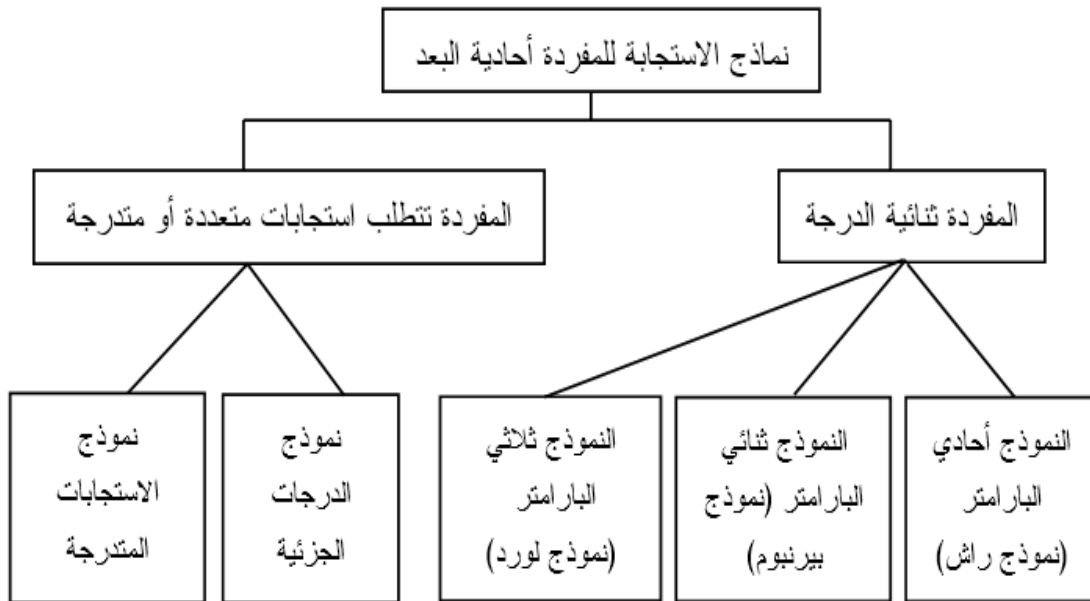
ويندرج تحت كل منهما نوعان فرعيان، هما: النماذج الاستاتيكية (Static Models) و النماذج الديناميكية (Dynamic Models) (بوعلام، 2013، ص67).

3-1- النماذج الاستاتيكية (Static Models):

يهتم هذا النوع من النماذج بقياس الأداء في لحظة زمنية واحدة مع التركيز على تحديد العمليات الكامنة وراء أداء الأفراد في الاختبارات النفسية والتربوية. وتختلف هذه النماذج من حيث الصيغ الرياضية التي تصف العلاقة بين أداء الفرد في الاختبار ومستوى قدرته وكذلك من حيث عدد الخصائص أو المعلومات المستخدمة في وصف كل مفردة من مفردات الاختبار (ميمي، 2014، ص.61).

وفي الوقت الحاضر هناك ثلاثة نماذج أساسية شائعة الاستخدام لكل منها صيغة رياضية خاصة تُجسّد في منحنيات مميزة للمفردات. تُطبق هذه النماذج على المفردات ثنائية الدرجة التي لا تتطلب سرعة في الإجابة وتقيس بعداً واحداً. كما توجد نماذج أخرى للتعامل مع الاستجابات الاسمية (Nominal Responses) والاستجابات متعددة الأقسام (Polytomous Responses)، ومنها نموذج الدرجات الجزئية (Partial Credit Model) ونموذج الاستجابات المتدرجة (Graded Response Model) وغيرها. وتُستخدم هذه المجموعة من النماذج في تحليل البيانات المستمدة من استبيانات الاتجاهات والشخصية ومقاييس التقدير والمفردات الاختبارية التي تتطلب خطوات متعددة حيث تُمنح درجات جزئية عن كل خطوة مكتملة في الإجابة عن مسألة معينة كحل المشكلات الرياضية. ويوضح الشكل (19) التخطيطي التالي هذه النماذج (بوعلام، 2013، ص.67).

شكل 19 يوضح أنواع نماذج الاستجابة للمفردة أحادية البعد.



3-2- نماذج تتعلق بالمفردة ثنائية الدرجة:

3-2-1- النموذج أحادي المعلم (One-Parameter Logistic Model (1PL):

يُعد هذا النموذج أبسط نماذج الاستجابة للمفردة أحادية البعد ويُعرف باسم نموذج راش (Rasch Model) نسبةً إلى عالم الرياضيات الدنماركي جورج راش بجامعة كوبنهاغن (Rasch, 1960). ورغم ارتباط اسمه بهذا النموذج فإن راش (1901-1980) عارض بشدة إطلاق اسمه عليه أثناء حياته مؤكداً أن الأمر لا يتعلق بنموذج واحد بل بعدة نماذج. فالنموذج القائم على الدرجة الثنائية هو واحد فقط من مجموعة نماذج إحصائية أطلق عليها راش اسم "نماذج القياس" (Measurement Models) وليس "نموذج راش" كما شاع لاحقاً. تم تطوير هذا النموذج بصورة مستقلة عن بقية نماذج الاستجابة للمفردة ويمكن اعتباره نموذجاً يُمثل فيه المنحنى المميز للمفردة بواسطة دالة ترجيح لوغاريتمي أحادية المعلمة (One-Parameter Logistic Function). في هذا النموذج يُعد المتغير التابع هو احتمال أن يجيب فرد (j) إجابة صحيحة على مفردة اختبارية (i) أما المتغيرات الكامنة المستقلة فهي: قدرة الفرد (θ) وصعوبة المفردة (b_i) حيث تُجمع هذه المتغيرات في صيغة رياضية تُطرح صعوبة المفردة من قدرة الفرد. يعتمد ارتباط هذا الفرق بالاستجابة على اختيار المتغيرات المستقلة التي تُنمذج في صيغة رياضية احتمالية متمثلة في دالة ترجيح لوغاريتمي غير خطية (Nonlinear Logistic Function) (أبو علام، 2013، ص.69).

كان "راش" قد نشر نموذج عام 1960 متناولاً 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. ويُعرف 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 $-\infty$ إلى $+\infty$ 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 تبعاً 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 يُعرف نموذج راش بأنه النموذج اللوجستي أحادي المعلمة والذي يمثل بالمعادلة الرياضية التالية (بيكر، 2001، ص.22):

$$P(x_{ji} = 1 | \theta_i, b_j) = \frac{e^{(\theta_i - b_j)}}{1 + e^{(\theta_i - b_j)}}$$

حيث إن:

x_{ji} : الاستجابة على المفردة j من خلال الشخص i .

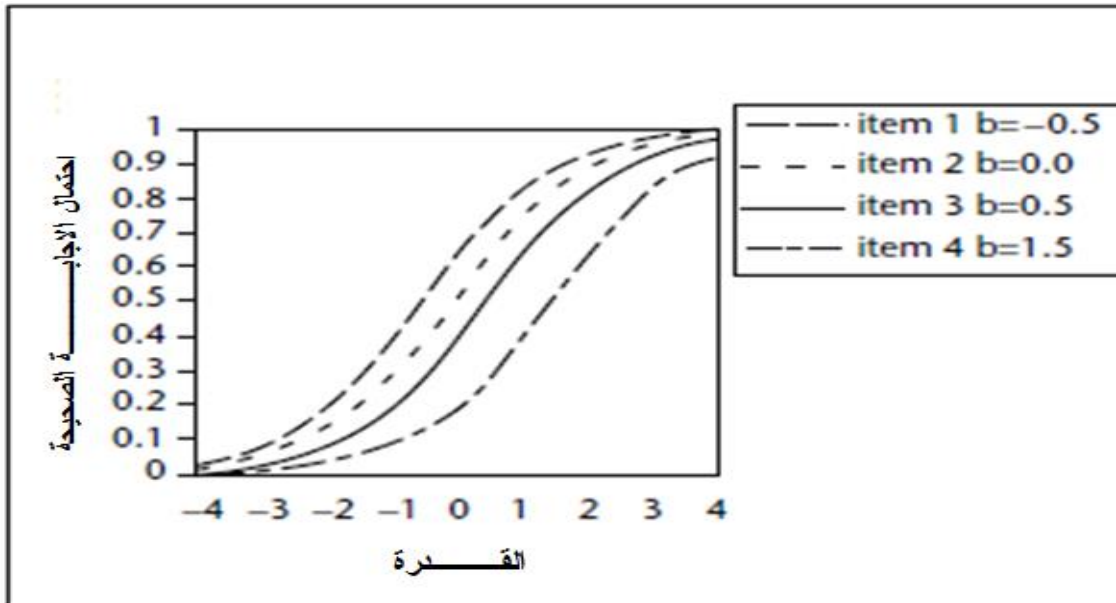
b_j : وضع (صعوبة) المفردة j .

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. وتُعرف هذه الصيغة باسم "نموذج الترجيح اللوغاريتمي أحادي البارامتر (1PL)" وذلك لأنها تعتمد على الدالة الأسية في التنبؤ بالاحتمالات وتشتمل على بارامتر (معلم) واحد فقط هو معلم صعوبة المفردة (Item Difficulty Parameter). والذي يُستخدم لتمثيل الفروق بين المفردات (Mislevy, 1990, p.195). ويمكن توضيح هذه الصيغة الرياضية من خلال المنحنى المميز لأربع مفردات تقيس سمة كامنة واحدة.

شكل 20 التمثيل البياني لصيغة نموذج راش لأربعة مفردات مختلفة الصعوبة.



يتضح من الشكل 20 أن المنحنيات الأربعة متوازية تختلف فقط في إزاحة الموقع أي في (الصعوبة) حيث أن المفردة (4) تبدو أكثر صعوبة من بقية المفردات.

مثال حسابي: لنفترض أن لدينا مستوى قدرة شخص ما θ_i تساوي (-3)، وقيمة معلم الصعوبة b_j للمفردة تساوي (1) (بيكر، 2001، ص.23).

- بداية نقوم بحساب الحد الأول في البسط من المعادلة السابقة $e^{(\theta-b)}$

$$\text{EXP} = 1 (-3-1) = 0.0183$$

- نقوم بتعويض الناتج في البسط وفي المقام مع إضافة الواحد، كالتالي:

$$P(x_{ji}) = 0.0183 / 1 + 0.0183 = 0.0179 \approx 0.02$$

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.

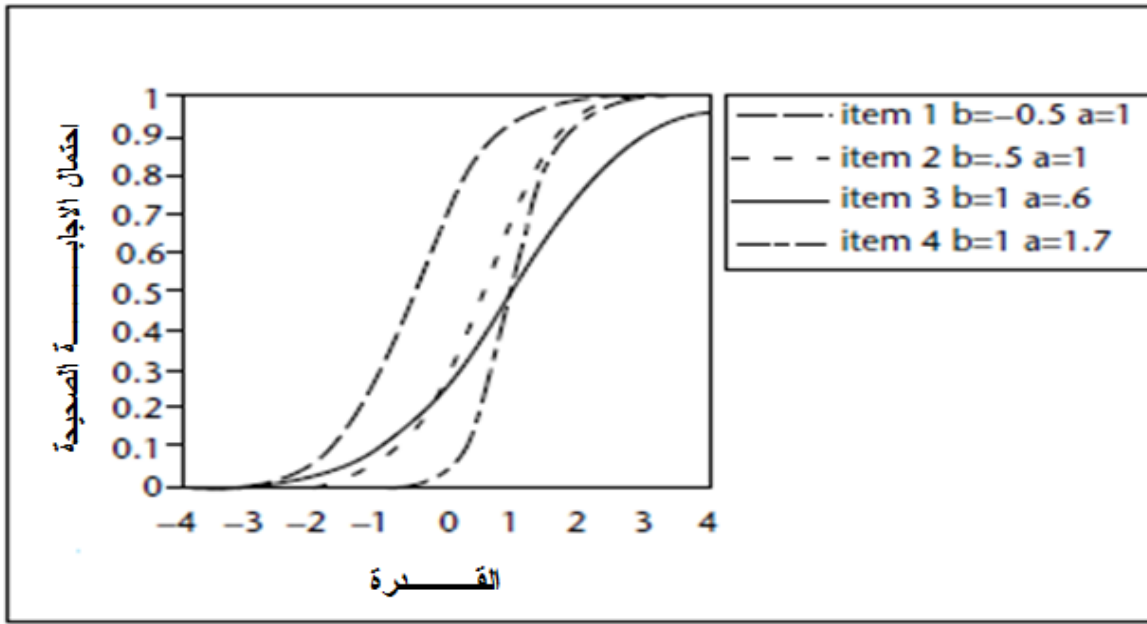
3-2-2- النموذج ثنائي المعلم (Two-Parameter Logistic Model (2PL):

اقترح عالم الإحصاء بيرنباوم (Birnbau, 1968) بالتعاون مع مجموعة من زملائه في جامعة كولومبيا الأمريكية نموذج التريجيج اللوغاريتمي ثنائي البارامتر (Two-Parameter Logistic Model - 2PL). ويختلف هذا النموذج عن نموذج راش في كونه يسمح باختلاف المفردات في كلٍّ 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:

$$P(x_j = 1 | \theta, b_j) = \frac{e^{a_j(\theta - b_j)}}{1 + e^{a_j(\theta - b_j)}}$$

يتضح من هذه الصيغة أن تمييز المفردة (a_j) يُضرب 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 معتمداً على النمط العام لإجابات الفرد الصحيحة والخاطئة وليس فقط على الدرجة الكلية. وبذلك يمكن أن نحصل على تقديرات مختلفة لمستوى السمة لدى فردين حصلوا على الدرجة الكلية نفسها إذا كان نمط الإجابات مختلفاً. 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 أوزاناً متساوية للمفردات عند تقدير مستويات السمة بخلاف نموذج راش (بيكر، 2001، ص.24). وينتج عن ذلك إمكانية تقاطع المنحنيات المميزة للمفردات في الاختبار المُعد وفق هذا النموذج، كما يوضحه الشكل (21) التالي:

شكل 21 التمثيل البياني لأربعة مفردات في النموذج ثنائي البارامتر.



يُظهر الشكل 21 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 θ_i تساوي (-3)، وقيمة معلم الصعوبة للمفردة b_j تساوي (1)، و معلم التمييز a_j يساوي (0.5) (بيكر، 2001، ص.20).

- بداية نقوم بحساب الحد الأول في البسط من المعادلة السابقة $e^{a(\theta-b)}$

$$\text{EXP} = 0.5 (-3-1)$$

$$\text{EXP} = -2 = 0.135$$

- نعوض الناتج في البسط وفي المقام مع إضافة الواحد، كآلاتي:

$$P(x_{ji}) = 0.135 / 1 + 0.135 = 0.118 \approx 0.12$$

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 جداً (-3 انحراف معياري) ما يزال يملك احتمالاً مقداره 12% للإجابة الصحيحة على المفردة، بسبب سهولة نسبية للمفردة أو هناك أثر للتخمين.

3-2-3- النموذج ثلاثي 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 أحياناً 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. انطلاقاً 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.

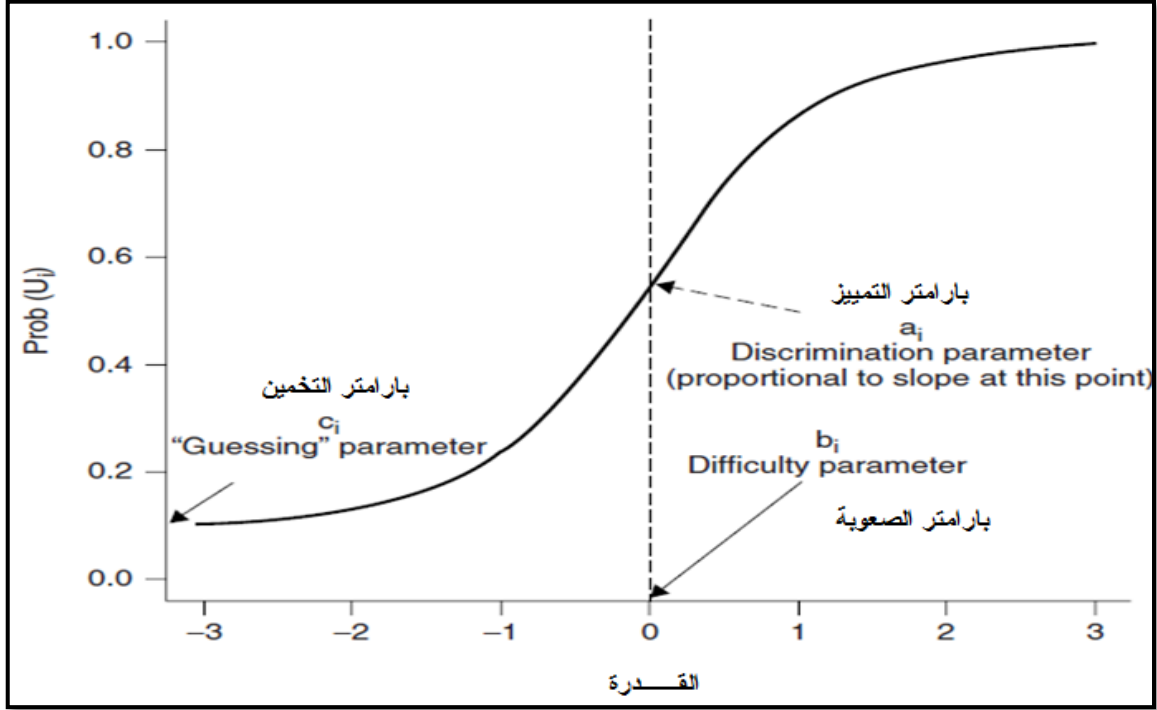
وقد صاغ لورد (Lord) الصيغة الرياضية لهذا النموذج انطلاقاً من حقيقة ثابتة في الاختبارات وهي أن بعض المفحوصين سيجيبون إجابة صحيحة على بعض المفردات نتيجة التخمين. وبالتالي فإن احتمالية الإجابة الصحيحة تشتمل على مكون صغير يُعزى إلى التخمين. ولم يُضمّن أي من النموذجين السابقين للمنحنى المميز للمفردة أثر التخمين في حساباته. لذلك قام آلان بيرنبوم (Allan Birnbaum, 1968) بتعديل النموذج اللوجستي ثنائي البارامتر ليضيف إليه معلماً ثالثاً يوضح إسهام التخمين في احتمالية الإجابة الصحيحة على المفردة (بيكر، 2001، ص.25).

ورغم أن هذا التعديل أدى إلى فقدان بعض خصائص الدالة اللوجستية إلا أن النموذج لا يزال يُعرف باسم النموذج اللوجستي ثلاثي البارامتر حتى وإن كان من الناحية الفنية لا يعد لوجستياً بالكامل، والذي يعطى بالمعادلة التالية:

$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{a_i(\theta_i - b_i)}}{1 + e^{a_i(\theta_i - b_i)}}$$

يُطلق على البارامتر C_i اسم الخط التقاربي الأدنى (Lower Asymptote) للمنحنى المميز للمفردة وهو يمثل احتمال الإجابة الصحيحة لدى الأفراد ذوي القدرة المنخفضة. ويأخذ هذا المعلم مدى نظرياً بين $0 \leq C_j \leq 1$ إلا أن القيم التي تتجاوز 0.35 تُعد 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 يُعد التخمين العشوائي أحد العوامل المؤثرة في أداء الأفراد على الاختبار (Charles et al., 2012, p. 246). ويُعتبر تغيير 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 مساوياً لـ 0.5. أما مع إدخال معلم التخمين فإن الحد الأدنى للمنحنى المميز للمفردة يصبح مساوياً لـ C_j بدلاً من الصفر مما يجعل C_j يمثل أدنى احتمال للإجابة الصحيحة على المفردة. وبناءً على ذلك يُعرّف معلم الصعوبة حالياً 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 (22) التالي:

الشكل 22 التمثيل البياني للنموذج ثلاثي البارامتر.



يتضح من الشكل (22) 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 الممتحن 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)، وقيمة معلم التمييز a_j يساوي (1.3)، وقيمة معلم التخمين C_j يساوي (0.2) (بيكر، 2001، ص.26).

- بداية نقوم بحساب الحد الأول في البسط من المعادلة السابقة $e^{a(\theta-b)}$

$$\text{EXP} = 1.3 (-3 - 1.5)$$

$$\text{EXP} = -5.85 = 0.0029$$

- نقوم بتعويض 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، كالاتي:

$$P(x_{ji}) = 0.2 + (1 - 0.2) 0.0029 / 1 + 0.0029$$

$$P(x_{ji}) = 0.2 + (0.8) (0.0029) = 0.2023$$

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.

4- معنى الموضوعية في نموذج راش:

في نموذج راش تعني الموضوعية إمكانية المقارنة بين نتيجة تفاعل قدرة فردين مع صعوبة مفردة مناسبة كما تشمل أيضًا المقارنة بين صعوبة مفردتين استجاب لهما فرد مناسب. ولا يكون التقدير الكمي لهذه المقارنة دقيقًا تمامًا بل يقع ضمن مدى معين حول القيمة المقدرة بحيث قد يقل أو يزيد عنها قليلًا، وتظهر هذه الموضوعية من جانبين رئيسيين:

أ- استقلال معلم قدرة الفرد عن المفردة المستخدمة:

على الرغم من أن المقارنة بين قدرتي فردين تعتمد على استخدام مفردة ملائمة فإن هذه المقارنة لا تتأثر بنوع المفردة المختارة من بين تلك المفردات الملائمة. والمقصود بالمفردات الملائمة هي تلك التي تتميز بالصدق في تعريفها للمتغير محل القياس والصدق في قياس الصفة وتدرجها على متصلها إضافة إلى تماثل قدرتها على التمييز بين مستويات الأفراد في هذه الصفة. وبناءً على ذلك فإن موضوعية المقارنة بين قدرتي فردين تعني أن هذه المقارنة مستقلة عن المفردة المستخدمة بحيث تؤدي استجابة الفردين لأي مفردة ملائمة إلى نفس التقدير الكمي للفارق بين قدرتهما (مع الأخذ في الاعتبار الخطأ المعياري على جانبي التقدير). ومن هنا فإن ما يتم التوصل إليه ليس القدرة المطلقة للفرد (V) على سبيل المثال وإنما المسافة بين قدرته وقدرة فرد آخر (U) بحيث تُعد قدرة الفرد (U) نقطة الأصل التي تُقاس منها قدرة الفرد (V).

ب- استقلال معلم صعوبة المفردة عن الفرد الذي يجيب عليها:

وبالمثل فإن المقارنة بين صعوبتي مفردتين تعتمد على استجابة فرد ملائم لكنها لا تتأثر بالشخص المحدد من بين هؤلاء الأفراد الملائمين. ويُعد الفرد ملائمًا عندما تكون إجابته على المفردة إجابة صادقة لا تختلف كثيرًا عن الإجابة المتوقعة من غالبية الأفراد الذين يمتلكون مستوى القدرة نفسه. وبناءً على ذلك فإن موضوعية المقارنة بين صعوبتي مفردتين تعني أن هذه المقارنة مستقلة عن هوية الفرد الذي يجيب عليهما بحيث تؤدي استجابة أي فرد ملائم إلى التقدير الكمي ذاته للفارق بين صعوبتي المفردتين (مع مراعاة الخطأ المعياري على جانبي التقدير). ومن ثم فإن ما نحصل

عليه ليس الصعوبة المطلقة للمفردة (i) على سبيل المثال وإنما المسافة بينها وبين صعوبة مفردة أخرى ولتكن (C) بحيث تُعد صعوبة المفردة (C) نقطة الأصل التي تُقاس منها صعوبة المفردة (i). وكلما تقاربت صعوبة المفردات المستخدمة مع مستوى قدرات الأفراد الذين يجيبون عليها زادت درجة تحقق الموضوعية في القياس.

5- الجانب السيكولوجي في نموذج راش:

يتجسد نموذج راش في صورة رياضية تعكس واقعًا سيكولوجيًا حيث يعبر بدقة عن ديناميكية التفاعل بين الأفراد ومفردات الاختبار. فالمعادلة التي يقوم عليها النموذج تفسر احتمال إجابة الفرد إجابة صحيحة عن مفردة اختبارية استنادًا إلى مبدئين أساسيين:

- الأول: إن ارتفاع قدرة الفرد يزيد من احتمال إجابته الصحيحة؛

- الثاني: الفرد يكون أكثر احتمالًا للإجابة الصحيحة إذا كانت المفردة سهلة بالنسبة له.

فإذا تحققت هاتان القاعدتان يصبح نموذج راش كما تمثله معادلته توصيفًا واقعيًا لهذا التفاعل والنموذج الوحيد بين النماذج المطروحة الذي يفسر أداء الأفراد في الاختبارات والمقاييس. وعليه فإن توافق البيانات المستمدة من الاختبارات مع متطلبات نموذج راش يستلزم صحة هذين المبدئين أما إذا لم تتحقق هذه المتطلبات فإن ذلك يشير إلى تعارض البيانات مع الأساس المنطقي الذي بُني عليه النموذج. فعلى سبيل المثال إذا لم تطابق مفردة ما مع النموذج فإن الافتراض القائل بأن الفرد الأعلى قدرة أكثر احتمالًا للإجابة الصحيحة مقارنة بالفرد الأقل قدرة يصبح غير صحيح ما يستدعي التساؤل حول طبيعة ما تقيسه هذه المفردة. وبالمثل إذا لم يتوافق أداء فرد معين مع النموذج فإن الافتراض القائل بأن احتمال إجابته الصحيحة على مفردة سهلة أكبر من احتمال إجابته على مفردة صعبة يكون غير قائم ما يدفع إلى التساؤل عن العوامل التي أثرت في أدائه مثل عدم قدرة المفردة على التمييز بين مستويات القدرة (علام، 2001، ص211).

ومن ثم يتضح أن أي نموذج لا يساهم في التفسير السيكولوجي للسمات المستهدفة لا يصلح أداة دقيقة للقياس النفسي والتربوي. ومع ذلك فإن لنموذج راش تطبيقات عملية مميزة فهو نموذج رياضي يعتمد أقل عدد ممكن من المعلمات وفي الوقت ذاته يستند إلى أساس سيكولوجي متين وهو ما يفسر الاهتمام الكبير الذي حظي به من قبل علماء القياس النفسي والتربوي المعاصرين (أبو علام، 2013، ص58).

6- وحدات القياس المستخدمة في نموذج راش:

تتدرج صعوبة المفردات وقدرات الأفراد على مقياس واحد بحيث تُقاس كلُّ من قدرة الفرد وصعوبة المفردة بوحدة قياس موحَّدة تُشتق مباشرة من نموذج راش وتُعرف بوحدة "اللوجيت" (Logit) أو "الترجيح اللوغاريتمي" وهي اللوغاريتم الطبيعي لاحتمال نجاح الفرد في المفردات التي تقع عند نقطة الصفر على التدرج عندما يساوي هذا الاحتمال قيمة الأساس الطبيعي (e) أي (2.718) ويكون حينها احتمال النجاح 0.73 (كاظم، 1988، ص 57).

ولتفادي ظهور تقديرات سالبة لكلِّ من القدرة والصعوبة يمكن تعديل نقطة الأصل في التدرج باعتبار الصفر قيمة اعتبارية كما يمكن تغيير حجم وحدة القياس أو سعتها لتجنُّب القيم الكسرية. ويعتمد ذلك على حقيقة أن الفرق بين القدرة والصعوبة هو الذي يحدد احتمال الإجابة الصحيحة وفق نموذج راش وبالتالي فإن إضافة أو طرح قيمة ثابتة من تقديرات القدرة والصعوبة لا يؤثر على هذا الاحتمال. ومن هنا يمكن اختيار نقطة أصل تتوافق مع هدف القياس. يُجرى تحويل وحدة القياس باستخدام معادلة التحويل الخطي التالية:

$$y = a + \beta x$$

حيث أن:

y = هو التدرج الجديد.

a = تمثل الإزاحة اللازمة لتحديد نقطة الأصل الجديدة.

β = هو عامل المسافة الذي يحدد سعة وحدة القياس الجديدة.

x = وتدرج اللوجيت هو القياس الأصلي.

وتختلف تدرجات المفردات وتتنوع تبعاً لأهداف القياس المختلفة، وفيما يلي بعض الانواع المختلفة من التدرجات:

1. وحدات التدرج الجماعية (نيت) (Normative Scaling Units Nits).

2. وحدات التدرج المعتمدة على محك مستقل (سيت) (Substantive Scaling Units (sits)).

3. وحدات التدرج الخاصة باحتمال الاستجابة الصحيحة (شيب) (Response Probability Scaling Units (chips)).

4. وحدة التدرج المستخدمة في المقاييس البريطانية (BAS) (اسماعيل، 2007، ص 47).

5. **وحدة المنف:** إذ تحول تدريج اللوجيت الى تدريج له خصائص تدريج الواط نفسها (تدريج مؤوي نقطة أصل تساوي 50) إلا أن طول أو سعة وحدة القياس (منف) يختلف عن سعة وحدة الواط. وتحصل على تدريج المنف ($\alpha=50$) ($\beta=5$) في معادلة التحويل الخطي السابقة.

6. **وحدة (الواط):** إذ يتم تحويل تدريج اللوجيت الى تدريج يمتد من صفر الى 100 ومتوسط صعوبة المفردات يساوي (50) ويمتاز هذا التدريج المؤوي بسهولة التفسير لتقديرات قدرة الافراد (Masters, 1988, p140).

7- تقدير بارامترات النماذج أحادية البعد:

في نماذج الاستجابة للمفردة أحادية البعد يعتمد احتمال استجابة الفرد لإجابة صحيحة على مفردة اختبارية على قدرته (θ) والبارامترات الخاصة بالمفردة. غير أن هذه القيم – سواء الاحتمالية أو البارامترات – تكون غير معروفة مسبقاً بينما المتاح لدينا هو استجابات الأفراد على المفردات. ومن ثم تهدف أساليب تقدير البارامترات إلى تحديد قيمة (θ) لكل فرد مختبر بالإضافة إلى تقدير قيم بارامترات المفردة بالاعتماد على إجابات الأفراد عليها.

7-1- تقديرات الأرجحية القصوى (Maximum Likelihood Estimation):

حظيت تقديرات الأرجحية القصوى باهتمام كبير من خبراء القياس الحديث الأمر الذي دفع العديد من العلماء إلى ابتكار أساليب ومنهجيات متنوعة لتسهيل عملية تقدير بارامترات نماذج الاستجابة للمفردة. وتتباين هذه الأساليب فيما بينها من حيث قدرتها على تحقيق المحكات التي ينبغي اعتمادها عند تقييم جودة أي أسلوب إحصائي استدلالي بحيث تمكّننا من الوصول إلى قيم تقديرية دقيقة للبارامترات في المجتمع المستهدف (أبو علام، 2013، ص93). ومن أهم هذه المحكات:

(أ) **الاتساق الداخلي Consistency** ويعني أن قيمة هذا الاحصاء تقترب تدريجياً من قيمة معلم المجتمع كلما زاد حجم العينة؛

(ب) **عدم التحيز Unbiasedness** وهذا يعني أن القيم التقديرية المتوقعة لقياس احصائي بعيد أي قيمة متوسط جميع العينات العشوائية الممكنة ذات حجم معين ينبغي أن تساوي قيمة معلم المجتمع؛

(ج) **الكفاية Sufficiency** وهي تعني أن توزيع العينة لا يعتمد على المعلم المراد تقديره لذلك فإن معرفة الأسلوب الاحصائي وتوزيع معانياته يكون كافياً لتزويدنا بهذه المعلومات؛

(د) **الفعالية النسبية Relative Efficiency** وتعني أنه إذا توافر مقياسان احصائيان غير متحيزين لتقدير معلم المجتمع فإن أفضلهما هو الذي يكون أكثر فعالية بالنسبة للآخر أي يكون الخطأ المعياري لتوزيع معانياته أقل (ميمي، 2014، ص87).

7-2- أسلوب الأرجحية القصوى المشتركة (JML): (Joint Maximum Likelihood Procedure):

نظراً لأن بارامترات الأفراد (θ) وبارامترات المفردات تكون غير معلومة في معظم الأحيان فإن تقديرها يمثل أحد المشكلات الرئيسية في نماذج الاستجابة للمفردة. وعند استخدام أسلوب الأرجحية القصوى المشتركة (JML) يمكن إجراء ذلك في خطوتين كالتالي:

أ- يتم اختيار قيم مبدئية لبارامترات الأفراد ويفضل أن تحسب هذه القيم المبدئية بقسمة لوغاريتم النسبة بين عدد الاجابات الصحيحة وعدد الاجابات الخاطئة لكل من فرد مختبر. وتحول القيم الناتجة الى قيم معيارية للتخلص من مشكلة عدم التحديد (Indeterminacy) والتعامل مع قيم بارامترات الأفراد (بارامترات القدرة) على أنها معلومة.

ب- يتم التعامل مع بارامترات المفردات على أنها معلومة، ويجري تقدير بارامترات الأفراد. وتتكرر هاتان الخطوتان حتى نجد أن القيم التقديرية للبارامترات لا تتغير بعد الانتهاء من الخطوتين في مرتين متتاليتين (Hambleton et al, 1989).

7-3- أسلوب الأرجحية القصوى الهامشية (MML): (Marginal Maximum Likelihood):

طوّر بوك وأيتكن (Bock & Aitken, 1981) هذا الأسلوب للتعامل مع حالات مستويات القدرة غير المعلومة وذلك عبر تطبيق مرحلتين متعاقبتين للتقدير ويُعرف هذا الأسلوب باسم خوارزمية التوقع-التعظيم (EM-Algorithm).

1- مرحلة التوقع (Expectation): يتم خلالها حساب العدد المتوقع من الأفراد الذين يستجيبون إجابة صحيحة لكل مفردة ثم استخدام هذه القيم المتوقعة في معادلات التقدير.

2- مرحلة التعظيم (Maximization): يُعاد في هذه المرحلة تقدير معالم المفردة (بارامترات) استناداً إلى التوقعات المعدلة بهدف الحصول على تقديرات أكثر دقة.

يتميز هذا الأسلوب بعدة مزايا من أبرزها إمكانية استخدامه في تقدير معالم جميع النماذج أحادية البعد وكذلك النماذج متعددة الأبعاد إضافةً إلى فعاليته العالية سواء كان الاختبار قصيراً بعدد محدود من المفردات أو طويلاً بعدد كبير منها. ورغم هذه المزايا إلا أن لهذا الأسلوب بعض القيود والسلبيات، من أهمها ما يلي:

- (أ) لا يؤدي في بعض الأحيان الى نتائج مرضية حتى بعد عمليات التكرار المتعددة؛
- (ب) يحتاج الى توافر بيانات اختبارية لعينة كبيرة الحجم نسبياً من أجل الاقتراب من توزيع مستويات القدرة والحصول على دالة الأرجحية الهامشية لمعالم المفردات (ميمي، 2014، ص89).

7-4- أسلوب الأرجحية القصوى المشروطة (Conditional Maximum Likelihood (CML):

حتى وقت قريب كان تطبيق هذا الأسلوب لتقدير بارامترات مفردات الاختبارات ممكناً فقط عندما لا يتجاوز عدد المفردات 30 إلى 40 مفردة. إلا أن التطورات الحسابية والبرمجية الحديثة جعلت من الممكن عملياً استخدامه مع اختبارات يصل عدد مفرداتها إلى 80 أو حتى 100 مفردة. تنبع أهمية هذا الأسلوب من أن التقديرات الناتجة عن بارامترات المفردات تكون متسقة (Consistent Estimates) وهو ما يعزز دقتها على المدى الطويل. ومع ذلك فإن نطاق استخدامه يقتصر على النموذج أحادي البارامتر والنماذج المتفرعة عنه حيث يتم تقييد دالة الأرجحية بعدد الإجابات الصحيحة التي يحققها الأفراد على المفردات الاختبارية (Wainer, et al, 1980).

ومن الجدير بالذكر أن هذا الأسلوب لا يصلح في حالة النموذجين ثنائي وثلاثي البارامتر حيث تختلف قيم بارامترات تمييز المفردات. وبذلك تعتمد الدرجات الكلية للأفراد في الاختبار على بارامترات تمييز غير معلومة وغير منظورة. فهذه الدرجات الكلية لا تعد احصاء كافياً لتقدير مستويات قدرة الأفراد حيث أن هذه التقديرات سوف تعتمد على أى المفردات أجاب عليها الفرد اجابة صحيحة ويتميز هذا الأسلوب بعدة مميزات، من أهمها ما يلي:

- (أ) لا يتطلب افتراض توزيع لمستويات القدرة كما في الأسلوب الثاني فبارامترات القدرة غير المعلومة لا تؤخذ بعين الاعتبار في معادلات تقدير صعوبة المفردات؛
- (ب) التقديرات الناتجة تتميز بخصائص مطلوبة في ظروف متعددة، مثل: الاتساق، والتوزيع الاعتدالي، والكفاية.

غير أن هذا الأسلوب يشوبه أيضاً بعض العيوب من أهمها مايلي:

- (أ) اقتصار استخدامه على النموذج أحادي البارامتر؛
- (ب) يؤدي استخدامه الى فقدان بعض المعلومات، حيث أنه لا يستطيع التوصل الى تقديرات للمفردات والأفراد الذين أجابوا على جميع المفردات اجابة صحيحة، أو اجابة خطأ؛
- (ج) توجد صعوبات في التحليل العددي عند استخدام اختبارات عدد مفرداتها كبير مما يؤدي الى فشل برنامج الحاسوب في استكمال اجراء التقديرات (Baker,1987 ; Embertson et al, 2000).

7-5- أساليب أخرى تعتمد على التقدير البايزي (Bayesian Estimation):

من المهم الإشارة إلى وجود أساليب أخرى لتقدير بارامترات نماذج الاستجابة للمفردة تستند إلى نظرية بيز والتي تقوم على الربط بين الاحتمالات الشرطية (Conditional Probabilities) والاحتمالات القبلية (Prior Probabilities). وتتطلب هذه الأساليب تحديد احتمالات قبلية للبارامترات

استنادًا إلى اعتبارات نظرية أو أدلة تجريبية بالإضافة إلى استخدام إجراءات إحصائية متقدمة لتحقيق التقدير (أبوعلام، 2013، ص 101).

8- الاختبارات التحصيلية المعتمدة في بنائها على نظرية الاستجابة للمفردة:

يُعد هذا الاختبار التحصيلي مثالًا على القياس الموضوعي حيث تتوافق درجات الطلبة فيه مع خصائص بنود الاختبار نظرًا لاعتماده في البناء على نظرية السمات (Trait Theory). وتمكّن هذه الاختبارات من قياس القدرات الحقيقية للطلبة بدقة أكبر.

وتفترض نظرية السمات أو ما يُعرف بنظرية الخصائص الكامنة أنه يمكن التنبؤ بالأداء التحصيلي للطلّاب أو تفسير نتائجه على الاختبار في ضوء خصائص مميزة لهذا الأداء تُسمّى السمات. تسعى هذه النظرية إلى تقدير درجات الأفراد على هذه السمات استنادًا إلى أدائهم في مجموعة من مفردات الاختبار (أمين، 2009، ص 236).

➤ صدق الاختبار وثباته في ضوء نموذج راش اللوغاريتمي الاحتمالي:

8-1- صدق الاختبار (Test Validity):

في ضوء نموذج راش يعني صدق الاختبار أن مفرداته تقيس متغيرًا واحدًا أي أن هذه المفردات تتدرج من حيث الصعوبة بطريقة تعكس قياس متغير واحد مشترك. كما يشير ذلك إلى أن قدرات الأفراد على هذا المتغير يمكن ترتيبها وفق تقديرات أدائهم على المقياس نفسه بحيث تتدرج كلٌّ من صعوبة المفردات وقدرات الأفراد على متصل واحد يمثل ذلك المتغير (كاظم، 1988، ص 98).

8-2- ثبات الاختبار (Test Reliability):

في ضوء نموذج راش يتحقق ثبات الاختبار من خلال استيفاء متطلبات الموضوعية في القياس بحيث يجب أن يتوافر في الأداة المبنية وفق هذا النموذج على ما يلي:

أ- استقلال القياس عن المقياس المستخدم.

ب- استقلال القياس عن مجموعة الأفراد التي طُبّق عليها المقياس.

ويعني ذلك أن الاستقلالية التي يوفّرها نموذج راش تضمن ثبات القياس بحيث لا تتأثر تقديرات قدرة الفرد أو صعوبة المفردة باختلاف عينة التدرّج أو باختلاف الأداة المستخدمة لقياس القدرة نفسها (الكلوت، 1991، ص 131).

كما يتميز نموذج راش بأنه لا يقتصر على توفير معامل ثبات للمقياس ككل بل يقدّم أيضًا معامل ثبات لكل فرد ولكل مفردة على حدة (El-Korashy, 1995, p. 760).

9- تدريج الاختبار وفق نموذج راش:

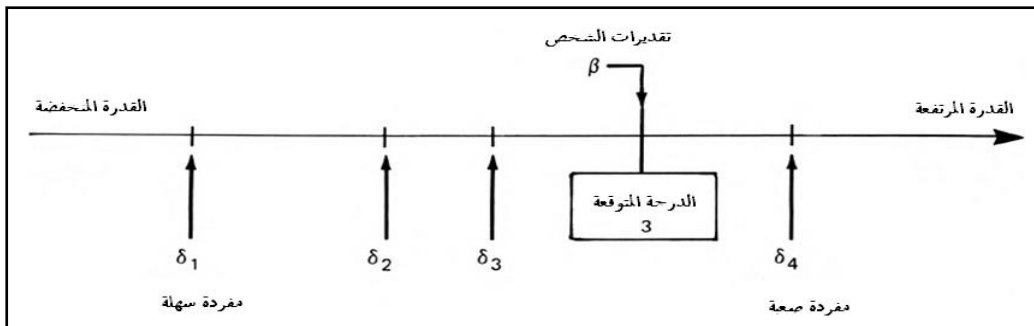
إن التطرق إلى مفهوم تدريج الاختبار يستلزم الإلمام بعدة مفاهيم مترابطة، وهي: منحى خصائص المفردة، ومنحنى خصائص الاختبار، واقتران المعلومات لكل من المفردة والاختبار، بالإضافة إلى مواصفات متصل القدرة. ويُفترض في متصل القدرة أن تتوسطه القيمة صفر وأن تكون وحدة القياس فيه هي الواحد الصحيح ويمتد ضمن المدى النظري ($\pm\infty$). وتقع قيم تقديرات معلمات كلٍّ من المفردات والمفحوصين على هذا المتصل ضمن المدى المشار إليه (محاسنة، 2013، ص 271).

ويقصد بـ تدريج الاختبار إنشاء ميزان لقياس الصفة السلوكية يحدد مستوى الأفراد على هذه الصفة بوحدة قياس معروفة. ويتجسد ذلك في انتظام مواقع المفردات على متصل صعوبتها حول نقطة الصفر بحيث يعكس موقع كل مفردة درجة صعوبتها (إسماعيل، 2007، ص 53).

تتم عملية تدريج الاختبار من خلال تطبيقه على مجموعة من المفحوصين بحيث تأخذ استجاباتهم على مفردات الاختبار شكلاً ثنائياً (1 - صفر). ثم تُجرى مجموعة من المعالجات والعمليات الرياضية على البيانات المستخلصة من هذه الاستجابات باستخدام البرامج الإحصائية بالحاسوب بهدف الحصول على متصل القدرة بشكل موحد لكل من المفردات والمفحوصين. بعد ذلك يتم تقدير معلمات المفردات والمفحوصين على هذا المتصل حيث يُعدّ تدريج الاختبار (القدرة، الصعوبة، التمييز) خطوة أساسية في إعداد الاختبار ووضعه قيد الاستخدام مما يدعم تفسير وتوضيح البناء النظري لنظرية القياس الحديثة.

وفي ظل أنموذج راش (Rasch) تكون وحدة القياس لم متصل السمة مساوية لـ (1) مع نقطة منتصف عند القيمة صفر. ووفق هذا النموذج يحصل المفحوصون الذين يحققون نفس العلامة الخام على التقدير نفسه للقدرة على خلاف ما يحدث في ظل النموذجين ثنائي وثلاثي المعلمات حيث قد تختلف تقديرات القدرة رغم تساوي العلامة الخام (محاسنة، 2013، ص 281).

شكل 23 يوضح تعريف متغير القياس من خلال عملية تدريج أو معايرة مفردات الاختبار.



يوضح الشكل متصل القدرة الذي يمتد من اليسار (قدرة منخفضة) إلى اليمين (قدرة مرتفعة) حيث تُوضع صعوبات مفردات الاختبار ($\delta_1 \dots \delta_4$) وقدرة المفحوص (β) على نفس المقياس. تمثل δ_1 مفردة سهلة، بينما δ_4 مفردة صعبة، وتحدد قدرة المفحوص موقعه على هذا المتصل استناداً إلى درجته الخام مثل الدرجة المتوقعة (3). في نموذج راش إذا تجاوزت قدرة المفحوص صعوبة مفردة ما، يكون احتمال إجابته الصحيحة عليها مرتفعاً مما يسمح بمقارنة مباشرة بين القدرات والصعوبات على مقياس واحد (Wright and Stone, 1979, p02).

10- كيفية التحقق من حسن مطابقة البيانات الاختبارية الافتراضات نموذج راش:

سبق أن أوضحنا أن نماذج الاستجابة للمفردة تقوم على مجموعة من الافتراضات التي تحدد طبيعة العلاقة بين استجابات الأفراد للمفردات الاختبارية وبين السمة أو السمات الكامنة التي تفسر هذه الاستجابات، ومن أهم هذه الافتراضات ما يلي:

أ- افتراض أحادية البعد (Unidimensionality): يفترض أن هناك سمة كامنة واحدة تكفي لتفسير التباين المشترك بين استجابات الأفراد على جميع المفردات.

ب- افتراض الاستقلال الإحصائي الموضعي (Local Independence): يعني أن الاستجابات للمفردات الاختبارية مستقلة عن بعضها البعض بعد ضبط مستوى أو مستويات قدرة الأفراد. وبمعنى آخر فإن احتمال استجابة الفرد لمفردة معينة يتحدد فقط بمستوى قدرته ولا يتأثر باستجاباته لمفردات أخرى. ويمثل هذا الافتراض في إطار نماذج القياس الكلاسيكية افتراض "عدم ارتباط الأخطاء المشروطة بالدرجات الحقيقية للفرد".

ومن الجدير بالذكر أن افتراض أحادية البعد والاستقلال الموضعي يرتبطان ارتباطاً وثيقاً رغم تمايزهما المفاهيمي إذ يمكن القول إن البيانات الاختبارية تكون أحادية البعد إذا كانت استجابات المفردات مستقلة إحصائياً عند تفسيرها في ضوء سمة كامنة واحدة.

وقد قام كما جاء في أبوعلام (2013، ص104) كل من (Hambleton and Rogers, 1988)، وهاتي (Hattie, 1984-1985) بإجراء مراجعة شاملة للمداخل والأساليب المستخدمة في تقييم أبعاد مصفوفة بيانات معينة، وخلص إلى الخطوات التالية:

- 1- التحقق من اتساق أنماط استجابات الأفراد للمفردات الاختبارية؛
- 2- تحليل المكونات الأساسية لمصفوفة البيانات الاختبارية وتحديد عدد قيم الجذور الكامنة الأكبر من الواحد الصحيح أو عمل رسم بياني لهذه القيم بعد ترتيبها من الأكبر إلى الأصغر وتمثيل هذا

الترتيب على المحور الأفقي وقيم الجذور الكامنة على المحور الرأسي وذلك لمصفوفة معاملات الارتباط الرباعي (Tetrachoric Correlation) ودراسة هذه القيم لتحديد ما اذا كان هناك عامل أول أكثر تشبعا أي يفسر أكبر قدر من التباين المشترك في المصفوفة؛

3. التحليل العامل غير الخطي (Nonlinear Factor Analysis) وهذا يتطلب اجراء مطابقة لنموذج التحليل العامل غير الخطي لمصفوفة معاملات الارتباط الداخلي بين المفردات (Ineritem Correlation Matrix)، وفحص بواقي تحليل التباين (Residual Covariance) بعد اجراء مطابقة لنموذج معين من نماذج التحليل العامل غير الخطي؛

4. التحقق من المفردات التي يبدو احتمال مخالفتها لافتراضات نموذج راش وذلك لمعرفة ما اذا كانت هذه المفردات تسلك بطريقة مختلفة عن المفردات الأخرى في المجموعة. ويمكن اجراء ذلك بالتعبير أو التدريج (Calibration) لهذه المفردات كمجموعة فرعية ومقارنة نتائج التعبير من خلال بارامترات الصعوبة بنتائج تعبير المجموعة الكلية للمفردات.

ب- تقييم الاستقلال الموضعي (Local Independence):

إن عدم تحقق خاصية الاستقلال الموضعي في البيانات الاختبارية يشكّل عائقاً أمام استخدام نماذج الاستجابة للمفردة أحادية البعد. فقد أظهرت الدراسات الإمبريقية لكل من ين (Yen, 1993) وستاينبرج وثيسن (Steinberg & Thissen, 1996) أن غياب هذه الخاصية يؤثر على تقدير معاملات نماذج الاستجابة للمفردة وقد ينتج عنه تقديرات أكبر من قيمها الحقيقية إضافة إلى التسبب في أخطاء عند تحديد عدد الأبعاد الكامنة في مصفوفة البيانات الاختبارية. وتوجد عدة ظروف في المواقف الاختبارية قد تؤدي إلى عدم تحقق خاصية الاستقلال الموضعي كتداخل المفردات بحيث تتأثر الاستجابة لإحداها بالاستجابة للمفردات الأخرى أو توقيت الاختبار أو اختلاف الأفراد في مدى تدريبهم على الإجابة عن مفردات الاختبار (أبوعلام، 2013، ص105).

ج- أساليب تقييم حسن مطابقة البيانات لنموذج راش:

بعد التأكد من تحقق خاصيتي البعد الأحادي والاستقلال الموضعي في مصفوفة البيانات الاختبارية والحصول على القيم التقديرية لمعاملات المفردات والأفراد باستخدام أحد النماذج الملائمة يصبح من الضروري تقييم مدى حسن مطابقة النموذج للبيانات. وقد أوضح العديد من خبراء القياس الحديث أنه لا يوجد أسلوب واحد كافٍ لهذا الغرض بل ينبغي الاعتماد على أدلة متعددة لإجراء مراجعة شاملة لمدى ملائمة النماذج والحكم على صلاحية النموذج المختار في تطبيق محدد (أبوعلام، 2013، ص107). ويتم هذا التقييم على مستويين:

- مستوى الاختبار كوحدة واحدة: للتحقق من مدى مطابقة مجموعة المفردات المكوّنة للاختبار للنموذج المستخدم.

- مستوى المفردة الفردية: للتحقق من مطابقة كل مفردة على حدة للنموذج خاصةً إذا كان الاختبار يتضمن أنواعًا مختلفة من المفردات تتطلب نماذج مختلفة لتقدير معلوماتها.

ويقصد بمطابقة المفردات للنموذج قدرة النموذج على تفسير كيفية استجابات الأفراد لإحدى المفردات أو التنبؤ بهذه الاستجابات. وقد قدّم خبراء القياس العديد من الأساليب والمقاييس والمؤشرات التي يمكن الاستعانة بها لتقييم مدى مطابقة المفردات للنموذج، ومن أبرزها:

أ- أسلوب اختبار مربع كاي (χ^2 Chi-Square):

يستخدم هذا الأسلوب لقياس مدى ملاءمة المفردة للنموذج المستخدم في العديد من البرامج الإحصائية، مثل: RUMM2020 و BILOG-MG و PARSCALE ومع ذلك يجب التعامل معه بحذر إذ إن أسلوب مربع كاي (χ^2) الإحصائي يتميز بحساسية عالية تجاه حجم العينة كما أنه يُشترط ألا يقل عدد مفردات الاختبار عن 20 مفردة. ولهذا السبب ينصح خبراء القياس بعدم الاعتماد عليه كمؤشر وحيد لحذف المفردات بل ينبغي للباحث أن يجمع بين نتائجه وبين الفحص التحليلي والنظرة المتفحصة للمفردات (Embretson & Reise, 2000, pp. 230–238).

ويشير فيكميروفيتش (Vukmirovic, 2009، ص15) إلى أنه عند التعامل مع العينات الكبيرة لا ينبغي التركيز على دلالة قيمة مربع كاي (χ^2) نفسها بقدر التركيز على اتساق القيم المختلفة داخل الاختبار؛ إذ إن القيم الكبيرة الشاذة غالبًا ما تعكس وجود مشكلة في المفردة وعلى الباحث أن يتوخى الحذر قبل اتخاذ قرار الحذف.

ومن الجدير بالذكر أن هامبلتون وسواميناثان (Hambleton & Swaminathan, 1985) أوضح أن كثيرًا من قيم مربع كاي ذات الدلالة الإحصائية تتأثر بشكل كبير بكم حجم العينة مما قد يثير الشك حول تفسيرها كدليل على عدم مطابقة البيانات للنموذج. ولهذا اقترح الاعتماد على القيم النسبية لمربع كاي عند تقييم حسن مطابقة المفردات بدلًا من الاكتفاء بالدلالة الإحصائية. كما اقترح ميسلفاي وبوك (Mislevy & Bock, 1996) استخدام نسبة أرجحية مربع كاي (Likelihood Ratio Chi-Square) في حالة الاختبارات التي تضم أكثر من 20 مفردة.

ب- أسلوب حساب الملاءمة عن طريق مقياسي الملاءمة التقاربي والتباعد:

يستخدم هذان الأسلوبان في برنامجي WINSTEPS و MICROSCALE لتحديد جودة المفردات من خلال مدى ملاءمتها للنموذج وذلك بالاعتماد على نوعين من إحصاءات الملاءمة:

- إحصاء الملاءمة التقريبية (Infit): يعتمد على أداء الأفراد الذين تقترب قدراتهم من مستوى صعوبة مفردات المقياس.

- إحصاء الملاءمة التباعدية (Outfit): يعتمد على أداء الأفراد الذين تختلف قدراتهم بشكل كبير (زيادة أو نقصان) عن مستوى صعوبة المفردات.

تُعد المفردة ضعيفة الملاءمة (Misfit) إذا تجاوزت القيمة الموجبة لإحصاءات الملاءمة – سواء كانت تقريبية أو تباعدية – الحد (2) أي تخطت حدود دلالة توزيع T عند مستوى (0.05). ويشير ذلك غالبًا إلى عيب في صياغة المفردة أو ضعف صدقها في قياس ما تقيسه بقية المفردات. أما إذا انخفضت إحصاءات الملاءمة السالبة عن القيمة (-2) فإن ذلك يُعد تجاوزًا لحدود الملاءمة (Overfit) مما يعني أن المفردة شديدة التشابه أو الاعتماد على بعض مفردات المقياس أي أنها ليست مستقلة عنها وبالتالي ينبغي حذفها لعدم تحقيقها افتراضات النموذج. وبالنسبة لاختبار مدى ملاءمة الأفراد للنموذج:

- إذا تجاوزت القيم الموجبة لإحصاء الملاءمة التقريبية (Infit) الحد (2)، فهذا يشير إلى أن الأفراد يجيبون خطأً عن مفردات تتقارب في صعوبتها مع مستوى قدراتهم.

- إذا تجاوزت القيم الموجبة لإحصاء الملاءمة التباعدية (Outfit) الحد (2)، فهذا يعني أن الأفراد إما يجيبون بشكل صحيح عن مفردات أصعب من مستوى قدراتهم أو يفشلون في الإجابة عن مفردات أسهل من قدراتهم.

- وإذا انخفضت القيم عن (-2)، فقد يشير ذلك إلى وجود تشابه كبير في إجابات الأفراد (وليد، 2014، ص30).

وفي هذا السياق تشير أمينة كاظم إلى أن رايت ولينكر (Wright & Linacre) أكدا عدم وجود محك مطلق للحكم على الملاءمة بل يجب أن يتوافق هذا الحكم على أهداف الموقف البحثي. وقد اقترحا اعتماد حد دلالة قدره (2) على أن يقوم الباحث أولاً بحذف الأفراد غير الصادقين ثم إعادة التحليل وبعد ذلك فحص المفردات غير الملائمة وحذفها ليُعاد التحليل مرة ثالثة للتأكد من جودة المطابقة للنموذج (1994، ص169).

وخلاصة القول إن تقييم مدى حسن مطابقة البيانات الاختبارية لأحد نماذج الاستجابة للمفردة يتطلب أولاً التحقق من افتراضات النموذج وبشكل خاص افتراض البعد الأحادي والاستقلال الموضوعي، ثم فحص تنبؤات النموذج من خلال تحليل البواقي المعيارية لتحديد مدى مطابقة البيانات للنموذج وتحديد طبيعة عدم المطابقة (Misfit).

وعلى الرغم من تنوع أساليب هذا التقييم، فإنه يمكن – وخصوصاً في نماذج راش – تصنيفها في ثلاث فئات رئيسية:

أ- حساب قيمة مربع كاي (χ^2)، من خلال مقارنة تكرارات الاستجابات الملاحظة بالاستجابات المتوقعة لمجموعات فرعية من الأفراد المختبرين محددة مسبقاً.

ب- حساب القيم المعيارية (Z)، بالاعتماد على دالة الأرجحية للاستجابات أو لأنماط المفردات.

ج- حساب بواقي الدرجات (Score Residuals)، بالاستناد إلى متوسطات انحرافات الدرجات الملاحظة عن الدرجات المتوقعة لاستجابات الأفراد المختبرين (أبوعلام، 2013، ص112).

11- دالة المعلومات (Information Function):

يعد مفهوم دالة المعلومات من المفاهيم الأساسية في نظرية الاستجابة للمفردة سواء في حالة المفردات ثنائية الدرجة أو المفردات التي تتطلب استجابات متدرجة متعددة. وتكمن أهمية هذه الدالة في مساعدتها على بناء أدوات قياس تحقق أعلى قدر ممكن من الدقة أو كمية المعلومات المطلوبة. وعلى الرغم من أن معظم تطبيقات دوال المعلومات انحصرت في نماذج الاستجابة للمفردة ثنائية الدرجة (إجابة صحيحة أو خطأ) فإن الباحثين في الآونة الأخيرة بدأوا بدراسة خصائص هذه الدوال عند استخدام نماذج الاستجابة متعددة الأقسام وكذلك نماذج الاستجابة متعددة الأبعاد.

ويمثل مفهوم دالة المعلومات إطاراً سيكومترياً عامّاً يرتبط مباشرة بتباين تقديرات الأرجحية القصوى؛ فعند الحصول على تقدير الأرجحية القصوى لمعامل القدرة يكون تباين هذا التقدير هو مقلوب دالة المعلومات المقابلة له. وبالمثل عند الحصول على تقديرات الأرجحية القصوى لمعاملات المفردات فإن مصفوفة التباين-التغاير (Variance-Covariance Matrix) الخاصة بهذه التقديرات تكون هي معكوس مصفوفة المعلومات (Information Matrix) المرتبطة بتقديرات معاملات المفردات.

ويرى بيرنبوم (Birnbau, 1968) أن مفهوم "المعلومات" (Information) يُعبّر عنه كمقدار يتناسب عكسياً مع مربع سعة فترة الثقة حول تقدير قدرة فرد معين. ويُحسب الخطأ المعياري لهذا التقدير وفق العلاقة التالية:

$$Standard Error = \frac{1}{\sqrt{Information}}$$

فكلما كان مقدار المعلومات عند مستوى قدرة معين كبيراً كانت فترة الثقة (Confidence Band)

حول تقدير القدرة أضيق مما يعكس دقة أعلى في التقدير. وعلى العكس تزداد سعة فترة الثقة كلما انخفض مقدار المعلومات عند نفس مستوى القدرة. ونظرًا لأن دالة المعلومات تتغير تبعًا لمستوى القدرة فقد اقترح خبراء القياس أن منحنيات معلومات الاختبار يمكن أن تكون بديلًا لكل من تقديرات الثبات المستخدمة في النماذج الكلاسيكية للقياس والأخطاء المعيارية للقياس المستخدمة في تفسير درجات الاختبارات (علام، 2013، ص113).

11-1- دالة معلومات المفردات (Item Information Functions):

يمكن لأي منحنى مميز لمفردة اختبارية – (ICC) سواء كان تابعاً لأحد نماذج الاستجابة للمفردة ثنائية الدرجة (Dichotomous Item) أو متعددة الأقسام (Polytomous Item) أن يخضع لتحويل ليصبح منحنى معلومات المفردة (IIC). ويعرض هذا المنحنى مقدار المعلومات السيكومترية التي توفرها المفردة الاختبارية عند مختلف النقاط على متصل السمة الكامنة. أما الصيغة الرياضية العامة لمنحنى معلومات المفردة (IIC) فتعطى كما يلي:

$$I(\theta) = \frac{[P'i(\theta)]^2}{[Pi(\theta)][1 - Pi(\theta)]}$$

يشير الرمز $P_i'(\theta)$ إلى المشتقة الأولى للدالة المميزة للمفردة عند مستوى محدد من القدرة أو السمة θ . وتمثل قيمة هذه المشتقة عند مستوى معين من القدرة θ ميل المماس للمنحنى المميز للمفردة عند النقطة $P(\theta)$. أما الرمز $P_i(\theta)$ فيشير إلى الاحتمال المشروط لإجابة المفردة i إجابة صحيحة، كما يُستمد من المنحنى المميز للمفردة (ICC) (ميمي السيد، 2014، ص93).

11-2- دالة معلومات الاختبار (Test Information Function):

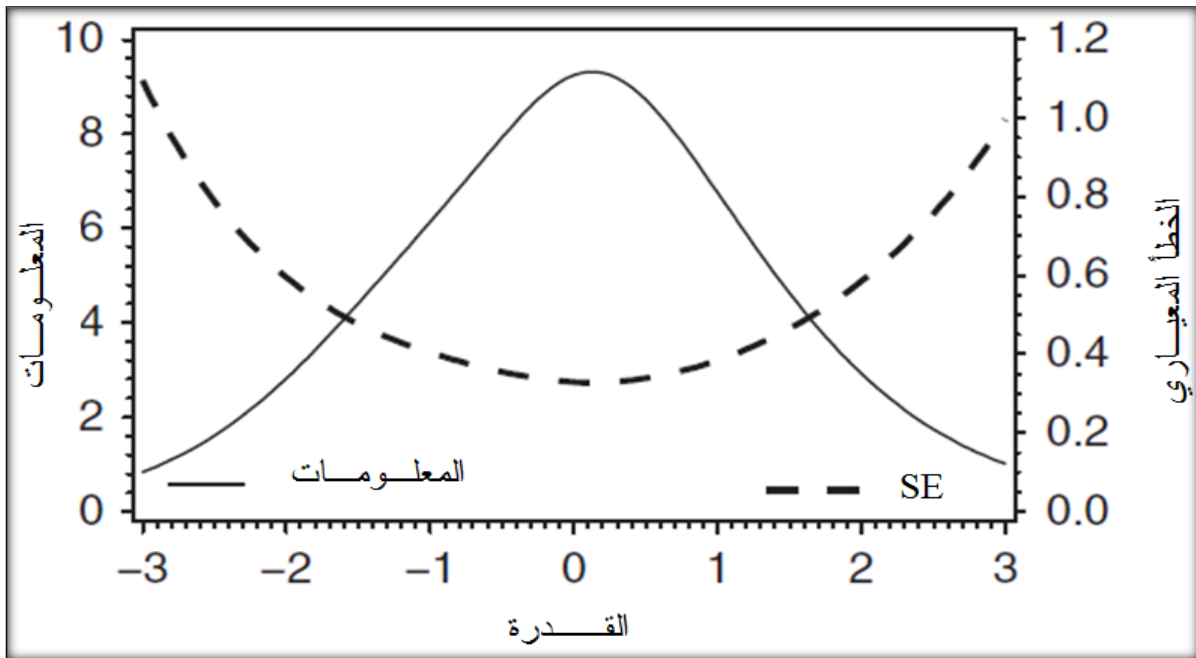
يعرف لوجيمان (Logemanm, 2006) دالة معلومات الاختبار بأنها مجموع دوال معلومات المفردات المكوّنة له، حيث تشكّل دوال معلومات المفردات الأساس الذي تُبنى عليه دالة معلومات الاختبار، ويتم تحديدها باستخدام المعادلة الآتية:

$$I(\theta) = \sum_{i=1}^n \frac{[P'i(\theta)]^2}{[Pi(\theta)][1 - Pi(\theta)]}$$

ويلاحظ أنه تم إضافة علامة (Σ) الى الطرف الأيمن من معادلة دالة معلومات المفردة وذلك لأن دالة معلومات الاختبار تعتمد على تجميع اسهامات كل مفردة من مفردات الاختبار من المعلومات عند مستوى قدرة معين (θ) (علام، 2013، ص117).

توفّر دالة معلومات الاختبار تقديراً لمدى كفاءة جميع مفردات الاختبار ضمن إطار دالة الاختبار على مقياس القدرة (θ). وكلما ارتفعت قيمة دالة المعلومات للاختبار عند درجة قدرة محددة (θ) زادت قدرة الاختبار على التمييز عند ذلك المستوى وانخفض خطأ القياس المصاحب له (ميمي السيد، 2014، ص93)، ويُوضّح الشكل (24) الآتي مثلاً على دالة معلومات الاختبار.

شكل 24 العلاقة بين معلومات الاختبار والخطأ المعياري لتقدير الكفاءة وهو الموضع الذي يكون فيه أكبر معلومات يكون فيه الخطأ المعياري للقياس أقل (Christine, 2010, p7).



يوضح الشكل منحنى المعلومات ومنحنى الخطأ المعياري في إطار نظرية الاستجابة للمفردة. يظهر على المحور الأفقي مستوى القدرة (θ) الذي يتراوح بين -3 و+3، بينما يمثل المحور الرأسي الأيسر مقدار المعلومات التي يوفرها البند أو الاختبار عند مستويات القدرة المختلفة، ويمثل المحور الرأسي الأيمن مقدار الخطأ المعياري المرتبط بتقدير هذه القدرة.

يبين المنحنى المتصل بالبيانات أن الاختبار يقدم أعلى مقدار من المعلومات عند مستوى القدرة المتوسط ($\theta = 0$) مما يشير إلى أن دقة التقدير تكون أكبر عند هذه المنطقة. ومع

الابتعاد عن المتوسط في الاتجاهين (القدرات المنخفضة أو المرتفعة) يقل مقدار المعلومات تدريجيًا وهو ما يعني أن الاختبار يصبح أقل فاعلية في التقدير عند الأطراف.

أما المنحنى المتقطع الذي يمثل الخطأ المعياري فيظهر العلاقة العكسية مع المعلومات؛ حيث يكون الخطأ المعياري في أدنى مستوياته عند المتوسط ($\theta = 0$) وهو ما يعكس أعلى درجات الدقة في التقدير. في المقابل يزداد الخطأ المعياري تدريجيًا كلما اتجهنا نحو القدرات المنخفضة أو العالية مما يشير إلى انخفاض الدقة في تلك المناطق.

وباختصار يؤكد الشكل أن العلاقة بين المعلومات والخطأ المعياري علاقة عكسية؛ فزيادة المعلومات تؤدي إلى انخفاض الخطأ المعياري والعكس صحيح وهو ما يعد من المبادئ الأساسية في دقة القياس وفق نظرية الاستجابة للمفردة.

الفصل السادس

الخصائص السيكومترية للاختبارات والمقاييس النفسية والتربوية

الأهداف (Objectives):

يهدف هذا الفصل إلى إلقاء الضوء على خصائص الاختبارات والمقاييس الجيدة مع بيان الشروط العامة التي ينبغي أن تتوفر 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 وفعالة في القياس النفسي والتربوي.

المحتوى (Contents):

- 1- خصائص الاختبارات والمقاييس الجيدة
- 2- الشروط العامة للاختبار النفسي والتربوي
- 3- الشروط التجريبية أو السيكومترية (Psychometric Conditions)
 - 1-3- الصدق (Validity)
 - 2-3- العوامل المؤثرة على الصدق
 - 2-3- الثبات (Reliability)
 - 3-3- العوامل المؤثرة على ثبات الاختبار

تمهيد:

تشكّل أدوات القياس والتقويم التربوي والنفسي أحد الأعمدة الجوهرية في البحث العلمي، إذ تمثل 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 مشوّهة 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 يساهم 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 وفعاليتها 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:

تُعدّ 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 الجوهرى يظل قائماً: أيّ 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؟

ووفقاً 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 للاختبار النفسي والتربوي:

هذه الشروط ترتبط بالتصميم الشكلي والإجرائي للأداة وتشمل:

أ- الشمول (Comprehensiveness): تمثيل جميع مكونات السمة أو القدرة المستهدفة بحيث يغطي الاختبار مجال المحتوى بشكل متوازن يضمن دقة القياس.

ب- التقيت (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 التجريبية أو السيكومترية (Psychometric Conditions):

وهي الخصائص الإحصائية التي تحدد دقة القياس وفاعليته:

أ- الثبات (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 البعد أو الظاهرة التي وُضع 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):

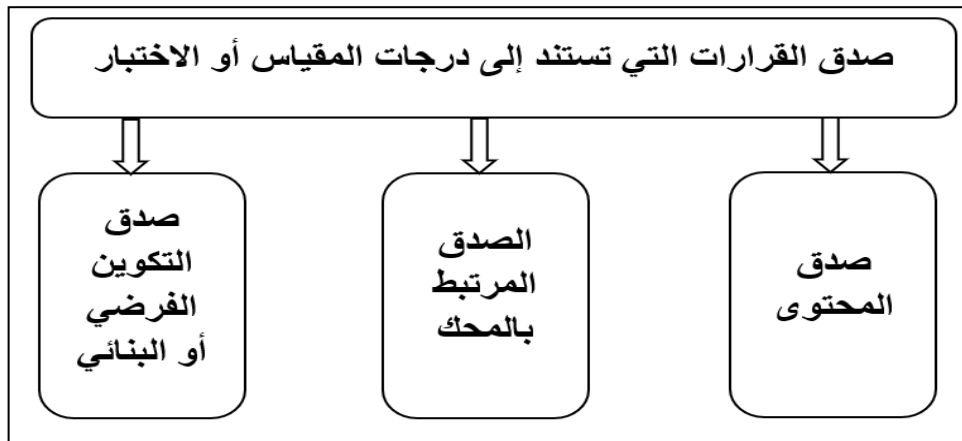
وتتعلق بالجوانب التطبيقية، مثل: وضوح التعليمات، سهولة التطبيق، تكلفة الوقت والموارد، ملائمة الاختبار للثقافة المحلية، وإمكانية استخدامه مع فئات مختلفة من المفحوصين.

3- الشروط التجريبية أو السيكومترية (Psychometric Conditions):

1-1-3- الصدق (Validity):

يُعد الصدق من أهم الخصائص السيكومترية لأي اختبار نفسي أو تربوي إذ يمثل الدرجة التي يتمكن فيها الاختبار من قياس السمة أو القدرة أو البُعد الذي وُضع لقياسه بدقة دون أن يتأثر بعوامل خارجية أو غير ذات صلة. ووفقاً لمعايير (APA) وإرشادات اللجنة الدولية للاختبارات (ITC) يُعد الصدق جوهر جودة القياس حيث يحدد مدى صلاحية التفسيرات والقرارات المستندة إلى نتائج الاختبار. ويُقاس صدق المفردة على وجه الخصوص بمدى ارتباطها المباشر بالهدف الذي صُممت لقياسه بحيث تعكس كل مفردة جانباً محدداً من المحتوى أو السمة المستهدفة. ومن الضروري التحقق من صدق الاختبار في صورته النهائية قبل اعتماده وذلك من خلال تطبيقه تجريبياً على عينات متنوعة وفي مواقف متعددة وجمع بيانات كافية حول الخصائص التي يقيسها أو النتائج التي ينتجها. وللصدق عدة أنواع إلا أنه يمكن تصنيفها في ثلاثة أنواع رئيسية:

شكل 25 يوضح أنواع الصدق الثلاثة.



3-1-1- صدق المحكمين أو الخبراء (Judgmental Validity):

يُعد 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. ووفقاً لمعايير APA Standards for Educational and Psychological Testing وإرشادات اللجنة الدولية للاختبارات (ITC Guidelines) تُعدّ عملية تحكيم الخبراء خطوة محورية

في مراحل البناء الأولية للاختبار وتسهم في تعزيز صدق المحتوى وضمان ملائمته 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 صُمم لقياسه. ويتمثل ذلك في المظهر الخارجي للاختبار من حيث نوع المفردات وصياغتها ووضوحها ودقة تعليماته ومدى موضوعيتها بالإضافة إلى ملائمة زمن الاختبار للاختبارات الموقوتة (التي تعتمد على السرعة) أو ملائمة 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 يُعدّ ذا صدق ظاهري جيد. ومع ذلك فإن هذا النوع من الصدق يُعتبر أضعف أشكال الصدق من الناحية العلمية لأنه لا يتجاوز المظهر السطحي للأداة ولا يوفر دليلاً إحصائياً على جودة القياس.

3-1-3- صدق المحتوى (Content Validity):

صدق المحتوى يُعنى بدرجة تمثيل مفردات الاختبار لكامل المجال أو البُعد المستهدف بالقياس بحيث يمكن اعتبار الاختبار عينة ممثلة لمحتوى المادة أو المهارة. فعلى سبيل المثال تُستخدم 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.

تُعد هذه العملية من أكثر الإجراءات أهمية في ضمان أن تكون نتائج الاختبار انعكاساً صادقاً للسمة أو القدرة المستهدفة وتسهم في تعزيز موثوقية القرارات المستندة إليها.

3-1-4- الصدق البيئي (Ecological Validity):

الصدق البيئي يُعد شكلاً خاصاً من صدق المحتوى ويختلف عن المغالطة البيئية أو مغالطة الاستدلال البيئي (Ecological Fallacy)¹ التي تُناقش 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 يُظهر 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:

- 1- إظهار أن العوامل البيئية لا تؤثر على النتائج.

- 2- إجراء القياسات داخل البيئة الطبيعية مع مراعاة صعوبة التحكم في المتغيرات الخارجية.

- 3- محاكاة العوامل البيئية في بيئة القياس كاستخدام أسئلة تمهيدية في المقابيس.

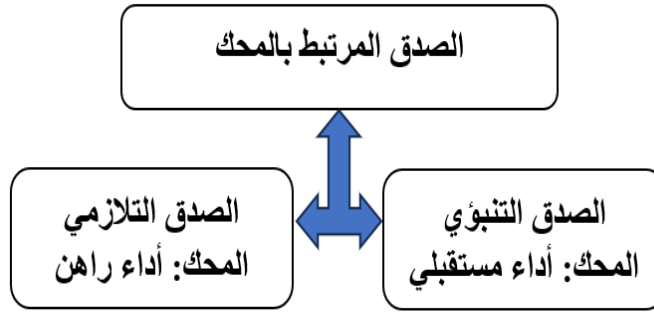
- 4- الاعتراف بأن السلوك في المختبر لا يعكس بالضرورة السلوك في البيئة الطبيعية مع تبرير استخدام هذا الأسلوب لفهم السلوك بعيداً عن المؤثرات البيئية (Garson, 2013, p.23).

3-1-5- الصدق المرتبط بالمحك (Criterion-Related Validity):

الصدق المرتبط بالمحك يُشير إلى درجة ارتباط نتائج الاختبار بمحك خارجي (External Criterion) يُمثل السلوك أو الأداء الذي يُراد التنبؤ به أو التحقق منه. ويُستخدم هذا النوع من الصدق للتأكد من قدرة الأداة على التنبؤ بأداء مستقبلي أو تفسير أداء حالي في مجال معين. ويعتمد في تقديره غالباً على حساب معامل الارتباط بيرسون (Pearson Correlation Coefficient) بين نتائج الاختبار ونتائج المحك المعتمد. ووفقاً للمعايير الدولية في القياس النفسي (APA Standards) و (ITC Guidelines) يُعد 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.

¹ - إذا كانت دراسة تظهر أن الدول التي لديها معدل تعليم عالٍ لديها أيضاً معدل دخل مرتفع، لا يعني ذلك بالضرورة أن كل فرد متعلم في هذه الدول لديه دخل مرتفع.

شكل 26 يوضح أنواع الصدق المرتبط بالمحك



أ- الصدق التلازمي (Concurrent Validity):

الصدق التلازمي يُستخدم 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 مرتفعاً 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 يُعتبر المحك أكثر أهمية من علامة الاختبار نفسها إذ تُستخدم نتيجة الاختبار فقط باعتبارها مؤشراً على الأداء المتوقع في المحك.

3-1-6- الصدق البنائي أو الفرضي (المفهوم) (Construct Validity):

يهدف الصدق البنائي إلى الإجابة عن سؤالين جوهريين:

- ما السمة أو الصفة التي يقيسها الاختبار تحديداً؟

- إلى أي مدى يتمكّن الاختبار من قياس هذه السمة بدقة وشمول؟

يمثل الصدق البنائي حجر الأساس في التأكد من أن الأداة تقيس فعلياً البنية أو السمة النفسية الكامنة التي وُضعت لقياسها كالميول والاتجاهات، التفكير الإبداعي، التفكير الناقد، الذكاء، 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 داخلياً.

• 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 البنائي:

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 نظرياً أن تختلف في السمة المقاسة.

مثال تطبيقي: إذا صُمم 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 وأخيراً فحص أدائه لدى مجموعتين: طلبة برامج جامعية متقدمة مقابل طلبة في بداية الدراسة حيث يُتوقع نظرياً وجود فروق لصالح المجموعة الأولى.

3-1-7- التحليل العاملي (Factor Analysis):

يُعد 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):

يُستخدم التحليل العاملي الاستكشافي (EFA) في المراحل الأولى من تطوير الاختبار عندما لا تكون البنية العاملية محددة مسبقاً حيث يهدف إلى استكشاف عدد العوامل الكامنة وطبيعة الارتباط بين المفردات دون فرض نموذج محدد مسبقاً. يقوم الباحث بحساب معاملات الارتباط بين كل مفردة وبقية المفردات مما ينتج عنه مصفوفة ارتباطات تُظهر مدى تجانس المفردات. تُقسم 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):

يُستخدم 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 10 لمقارنة بين الفروق الأساسية بين التحليل العاملي الاستكشافي والتحليل العاملي التوكيدي

البند	التحليل العاملي الاستكشافي (EFA)	التحليل العاملي التوكيدي (CFA)
الهدف	استكشاف البنية الكامنة للاختبار دون وجود نموذج محدد مسبقاً.	التحقق من مدى مطابقة البيانات لنموذج نظري محدد مسبقاً.
مرحلة الاستخدام	المراحل الأولى من تطوير الاختبار أو عند غياب معرفة سابقة بالبنية العاملية.	بعد تحديد نموذج نظري للبنية العاملية للاختبار.

الافتراضات المسبقة	لا توجد فرضيات محددة حول عدد العوامل أو طبيعة ارتباط المفردات.	توجد فرضيات واضحة حول عدد العوامل والمفردات المرتبطة بكل عامل.
طريقة 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 النمذجة بالمعادلات البنائية (SEM) وفحص مؤشرات المطابقة.
النتيجة	تحديد عدد العوامل والمفردات التي تنتمي لكل عامل، وتوليد نموذج أولي للبنية العاملية.	تحديد مدى ملاءمة النموذج النظري للبيانات وتقدير جودة المطابقة (Fit Indices).
المؤشرات الإحصائية	معاملات تشبعات العوامل (Factor Loadings)، ونسبة التباين المفسر.	مؤشرات المطابقة، مثل: CFI، TLI، χ^2 ، RMSEA.
المرونة	مرن في اكتشاف العلاقات، لكن نتائجه استكشافية وقد تختلف بين عينات مختلفة.	أقل مرونة لأنه يختبر نموذجًا محددًا، لكن نتائجه أدق في التحقق من البنية النظرية.

3-1-8- دراسات الارتباط (Correlation Studies):

تُعد 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 البنية 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 نظراً لاعتبارها إحدى السمات الجوهرية للفرد الطموح. فإذا حصل على معامل ارتباط مرتفع ودال إحصائياً بين الاختبارين فهذا يُعد دليلاً 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 معياراً صادقاً لقياس الظاهرة المستهدفة وبالتالي فإن مفردات الاختبار الجيدة يجب أن تكون مرتبطة ارتباطاً وثيقاً 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 منخفضاً باعتبار أنها لا تقيس البنية أو السمة النفسية التي يقيسها الاختبار ككل.

9-1-3- الصدق التمايزي (Differential Validity):

لا يقتصر الصدق البنائي على أن يظهر الاختبار ارتباطاً مرتفعاً مع مقاييس تشبهه بل يتطلب أيضاً ألا يرتبط بقوة مع مقاييس يفترض أن تكون مختلفة عنه. فمثلاً إذا حقق اختبار التفكير الرياضي ارتباطاً قدره (0.60) مع اختبار للقدرة العددية ولكنه أظهر ارتباطاً أعلى من (0.70) مع اختبار للفهم القرائي فإن ذلك يثير الشك حول صدقه. بمعنى آخر ليس مجرد وجود علاقة بين الاختبار ومحك معين دليلاً كافياً 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).

وغالباً ما يُفحص 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):

تُعد طريقة المجموعات المعروفة أحد أكثر أشكال الأدلة شيوعاً لدعم الصدق البنائي (Construct Validity) وتستخدم عندما يكون الاختبار قادراً على تمييز الأفراد المعروف أنهم يمثلون صفة معينة عن أولئك الذين لا يمثلونها. باستخدام طريقة المجموعات المعروفة يتم اختيار محك يمكنه تحديد وجود أو غياب صفة معينة ويُستخدم السياق النظري وراء البنية (Construct) لتوقع كيفية تصرف المجموعات المختلفة. وبالتالي يُدعم صدق الاختبار إذا أظهرت نتائج هذه الاختلافات المعروفة بين المجموعات. وبالمثل يمكن دراسة المجموعات المعروفة باستخدام مجموعات من الأفراد الذين يختلفون في مستويات أو شدة الصفة نفسها. تُقِيم 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):

نظراً 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. وانطلاقاً 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 دعماً متفاوتاً للفرضيات حيث برزت علاقة قوية بين درجات الاستقلالية الوظيفية (FIM) ووجهة خروج المرضى وشدة إصابات الحبل الشوكي والسكتة الدماغية. ويُعد هذا التحليل دليلاً واضحاً 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 النزهي:** 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 الفيزيائية:** 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- **صياغة التعليمات:** التعليمات الغامضة أو المضللة قد تؤدي إلى تباين في فهم المطلوب من المفحوصين.

د- **طريقة المراقبة:** الرقابة الصارمة أو العشوائية المفرطة قد تخلق ضغوطًا أو تمنح فرصًا للغش.

3-5- عوامل تتعلق بالاختبار نفسه:

ترتبط بخصائص الأداة وملاءمتها لقياس الظاهرة المستهدفة، وتشمل:

أ- **مستوى صعوبة المفردات:** المفردات السهلة جدًا أو الصعبة جدًا تقلل من القدرة على التمييز بين مستويات القدرة المختلفة.

- ب- **وضوح الصياغة:** المفردات الغامضة أو متعددة المعنى قد تربك المفحوصين.
- ج- **طول الاختبار:** الاختبارات الطويلة جداً قد تسبب الملل أو الإرهاق، بينما القصيرة جداً قد لا تمثل السلوك المستهدف تمثيلاً كافياً.
- 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 جزءاً من صدقه.

ح- **تجانس عينة التقنين:** العينة غير الممثلة للمجتمع المستهدف تقلل من إمكانية تعميم النتائج.

2-3- الثبات (Reliability):

يشير مفهوم الثبات في الاختبارات إلى مدى خلو الدرجات من الأخطاء غير المنتظمة (العشوائية) التي قد تشوب عملية القياس أي مدى قدرة الاختبار على قياس المقدار الحقيقي للسمة المستهدفة بشكل متسق. ويُعد الاختبار ثابِتاً (Reliable) إذا كانت نتائجه متشابهة نسبياً عند إعادة تطبيقه في ظروف مختلفة أو عند إعادة صياغة بنوده بما يعكس نفس البنية النظرية مع بقاء مستوى القياس دقيقاً 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 والدرجات الحقيقية:

وفق النظرية الكلاسيكية للقياس (Classical Test Theory) تتكون الدرجة التي يحصل عليها الفرد في الاختبار من عنصرين رئيسيين:

أ- **الدرجة الحقيقية (True Score):** وهي الدرجة التي تمثل المستوى الفعلي للفرد في السمة التي يقيسها الاختبار خالية من أي خطأ عشوائي.

ب- **درجة الخطأ (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:

$$\begin{aligned} \text{الدرجة الملاحظة (OS)} &= \text{الدرجة الحقيقية (TS)} + \text{درجة الأخطاء العشوائية (ES)} \\ \text{الدرجة الحقيقية (TS)} &= \text{الدرجة الملاحظة (OS)} - \text{درجة الأخطاء العشوائية (ES)} \end{aligned}$$

وإذا تمكنا من تقدير درجة الخطأ وإزالتها من الدرجة الملاحظة (Observed Score) فإننا نقرب من معرفة الدرجة الحقيقية للفرد. وكلما كان حجم الخطأ أصغر زاد ثبات الاختبار والعكس صحيح.

مثال 01: حصل طالب في اختبار 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:

$$T = \bar{X} = \frac{\sum X}{n} = \frac{840}{12} = 70$$

- ونكمل الحل كما في الجدول التالي:

جدول 11 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:

$$r^2 = \frac{1}{n} [\sum E^2]$$

$$= \frac{242}{12} = 20.17$$

مثال 02: كانت العلامات الحقيقية والعلامات الملاحظة (المشاهدة) لخمس طلبة كما في الجدول التالي (سامي، 2011، ص 255).

المطلوب: أوجد أخطاء القياس والخطأ المعياري للقياس ثم أوجد معامل الثبات؟

جدول 12 معطيات لحساب أخطاء القياس والخطأ المعياري للقياس.

رقم الطالب	العلامة الملاحظة X	العلامة الحقيقية T	أخطاء القياس 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 الدرجات الملاحظة والدرجات الحقيقية أي:

$$\bar{T} = \frac{144}{5} = 28.8$$

$$\bar{x} = \frac{144}{5} = 28.8$$

4- ثم نحسب التباينات باستخدام المعادلة التالية:

$$S^2 = \frac{1}{n} [X^2 - n \bar{X}^2]$$

$$s_T^2 = \frac{1}{5} [(32^2 + 15^2 + 30^2 + 43^2 + 24^2) - 5 \times 28.8^2] = 85.34 \approx 85$$

$$s_x^2 = \frac{1}{5} [(36^2 + 20^2 + 24^2 + 44^2 + 20^2) - 5 \times 28.8^2] = 92.16 \approx 92$$

- بنفس الطريقة نحسب تباين أخطاء القياس:

$$s_E^2 = \frac{1}{5} [((+4)^2 + (+5)^2 + (-6)^2 + (+1)^2 + (-4)^2) - 5 \times 0] = 18.8$$

$$\approx 19$$

- ومعامل الثبات r يساوي:

$$r_{xx} = \frac{S_T^2}{S_X^2} = \frac{85}{92} = 0.92$$

$$S_E = S_x \sqrt{(1 - r)}$$

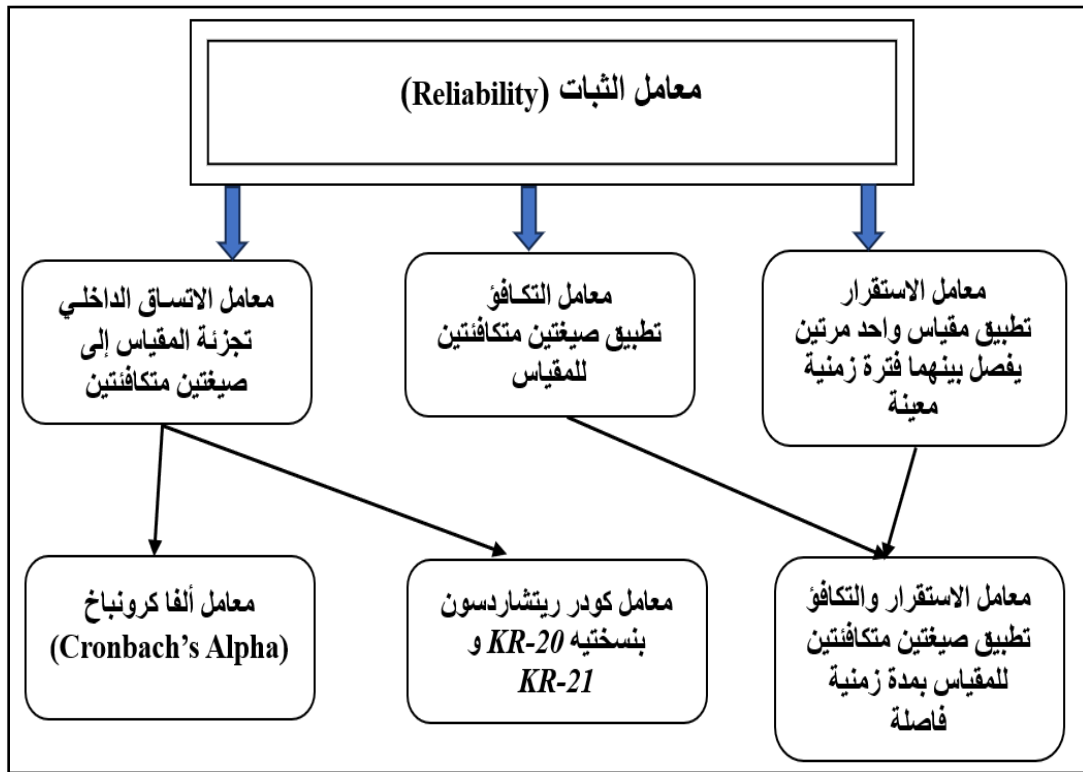
- إذن الخطأ المعياري للقياس يساوي:

- حيث: S_x = الانحراف المعياري.

$$S_E = \sqrt{92.16} \sqrt{1 - 0.92} = 2.72$$

- ويمكن حساب ثبات الاختبار بطرق عديد يمكن تلخيصها في الشكل التالي:

شكل 27 يوضح أنواع معامل الثبات



3-2-2- طريقة الاختبار وإعادة الاختبار (Test- Retest Method):

تُعد 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 تساعد 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 استخدامهما مع اختبارات الذاكرة أو الاختبارات التي تتأثر بدرجة كبيرة بالتعلم والممارسة.

3-2-3- طريقة الصور المتكافئة (Equivalent or Parallel Forms Method):

تُعد 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 بفواصل زمنية قصيرة 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 الأداء 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 تماماً: إعداد اختبارين متوازنين يتطلب جهداً كبيراً 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 قصيراً.

ج- تكلفة زمنية ومادية مرتفعة: تحتاج هذه الطريقة إلى موارد إضافية لتطوير صورتين متكافئتين بدلاً من اختبار واحد.

د- احتمالية عدم التطابق الكامل: حتى مع الحرص الشديد قد توجد فروق دقيقة بين الصورتين تؤثر على عدالة القياس ودقة معامل الثبات.

- ملاحظة: 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 كثيراً 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):

تُعد هذه الطريقة من أقدم وأكثر طرق تقدير الثبات شيوعاً حيث يُقسّم الاختبار إلى قسمين متكافئين من حيث: عدد المفردات، مستوى الصعوبة، طبيعة الأهداف، وزمن الإجابة. وغالباً ما يُجرى هذا التقسيم إما بطريقة الفردي-الزوجي (Odd-Even) بحيث تُجمع إجابات البنود الفردية في جزء والزوجية في جزء آخر أو بطريقة تقسيم الاختبار إلى نصفين متساويين مع مراعاة التوازن. بعد ذلك يُحسب معامل الارتباط بين أداء الأفراد على الجزأين ويُعتبر هذا المعامل مقياساً لثبات نصف الاختبار فقط. ولأن الهدف هو تقدير ثبات الاختبار كاملاً تُستخدم 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):

تُستخدم هذه المعادلة 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 يُعرف 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 المعادلة:

- شائعة جداً في اختبارات التحصيل الدراسي واختبارات القدرات.
- تساعد الباحث على تقدير الثبات المتوقع إذا زاد أو نقص طول الاختبار.
- تُستخدم أيضاً للتنبؤ بمستوى الثبات عند تعديل عدد المفردات (زيادة البنود عادة ترفع الثبات).

- حدود المعادلة:

- تفترض أن نصفي الاختبار متكافئان في جميع الجوانب (الصعوبة، الأهداف، عدد البنود).
- إذا لم يكن الانقسام متكافئاً، قد تُعطي نتائج مضللة.
- لا تُستخدم 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- معادلة رولون (Rulon Formula):

تُعد 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}).

$$S_t^2 = \frac{\sum X_t^2 - \frac{(X_t)^2}{n}}{n}$$

2- تباين الفروق بين درجات الأفراد في النصف الأول والنصف الثاني (S^2_{diff}).

$$S_d^2 = \frac{\sum X_d^2 - \frac{(X_d)^2}{n}}{n}$$

ثم يُطبَّق القانون التالي للحصول على معامل ثبات الاختبار:

$$r_{11} = 1 - \frac{S_d^2}{S_t^2}$$

- مزايا معادلة رولون:

- لا تحتاج إلى حساب معامل الارتباط بل تعتمد فقط على التباين.
- تُعتبر بديلاً جيداً في حالة الرغبة في تجنب الانحياز الناتج عن اختلاف طريقة تقسيم البنود.
- تُبرز 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 (Guttman):

تُعد 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 فعلياً في كثير من الاختبارات حيث لا يكون النصفان متكافئين تماماً من حيث الصعوبة أو التباين. وتكتسب هذه الطريقة أهميتها من كونها من أكثر أساليب تقدير الثبات شيوعاً إذ يتم تطبيق الاختبار مرة واحدة فقط على الأفراد مما يوفر الوقت والجهد مقارنةً بالطرق التي تتطلب إعادة تطبيق الاختبار. وبعد ذلك يُجزأ الاختبار إلى نصفين (أو أحياناً أكثر) بطرائق متعددة كالفصل بين البنود الزوجية والفردية أو تقسيم الاختبار إلى جزأين متساويين في الطول والمحتوى. ويُعتبر هذا الإجراء عملياً وبسيطاً كما أنه يعكس بدرجة معقولة مدى اتساق أداء الأفراد على أقسام مختلفة من نفس الاختبار مما يمنح الباحثين مؤشراً واضحاً على ثبات الاختبار وموثوقيته، وتعطى بالصيغة الرياضية التالية:

$$r_{tt} = \frac{S_x^2 - (S_1^2 + S_2^2)}{S_x^2}$$

حيث أن:

r_{tt} = معامل الثبات الكلي للاختبار.

S_x^2 = التباين الكلي لدرجات الاختبار.

S_1^2 = تباين درجات النصف الأول من الاختبار.

S_2^2 = تباين درجات النصف الثاني من الاختبار.

وتُظهر هذه المعادلة أن الثبات يزداد كلما كان التباين الكلي أكبر مقارنة بمجموع تباينات نصفي الاختبار أي أنها 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 الاختبار 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 البُعد أو السمة النفسية ذاتها وهو ما يعكس مستوى الصدق والثبات الداخلي للأداة.

أ- معادلة كودر-ريتشاردسون (Kuder-Richardson Formula - KR-20):

يُعد من أشهر مؤشرات ثبات الاتساق الداخلي للاختبارات 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:

$$KR-20 = \left(\frac{K}{K-1} \right) \left[1 - \frac{\sum_{k=1}^p (p_k q_k)}{S_{Total}^2} \right]$$

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^2_{Total} = 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 اختباراً مكوناً من خمس مفردات (أسئلة)، وقد جرى تطبيقه على عينة صغيرة مكونة من ستة طلاب. يجيب كل طالب عن كل مفردة باستخدام بديلين فقط، إما (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 13 معطيات لمثال لحساب معامل كودر- ريتشاردسون (KR-20)

المجموع	المفردة 5	المفردة 4	المفردة 3	المفردة 2	المفردة 1	
الطالب 1	1	1	1	0	1	4
الطالب 2	1	1	1	1	1	5
الطالب 3	0	0	1	0	1	2
الطالب 4	0	1	0	0	0	1
الطالب 5	1	1	1	1	1	5
الطالب 6	1	1	0	1	1	4
$S^2 = 2.70$	0.6667	0.8333	0.6667	0.5	0.8333	p_k
	0.3333	0.1667	0.3333	0.5	0.1667	q_k
	0.2222	0.1389	0.2222	0.25	0.1389	$p_k \times q_k$

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:

$$x_i - \bar{x}$$

$$0.25 = (0.5)^2 \quad \text{- الطالب 1: } 0.5 = 3.5 - 4$$

$$2.25 = (1.5)^2 \quad \text{- الطالب 2: } 1.5 = 3.5 - 5$$

$$2.25 = (-1.5)^2 \quad \text{- الطالب 3: } 1.5 = 3.5 - 2$$

$$6.25 = (-2.5)^2 \quad \text{- الطالب 4: } 2.5 = 3.5 - 1$$

$$2.25 = (1.5)^2 \quad \text{- الطالب 5: } 1.5 = 3.5 - 5$$

$$0.25 = (0.5)^2 \quad \text{- الطالب 6: } 0.5 = 3.5 - 4$$

$$\Sigma = (x_i - \bar{x})^2 = 13.5 \quad \text{- المجموع:}$$

4- حساب التباين الكلي للدرجات، بالمعادلة التالية:

$$\sigma^2 = \frac{\Sigma = (x_i - \bar{x})^2}{N - 1} = \frac{13.5}{5} = 2.7$$

- وبتعويض النتائج في معادلة كودر-ريتشاردسون نتحصل على التالي:

$$\sum pk x qk = 0.1389 + 0.25 + 0.2222 + 0.1389 + 0.2222$$

$$\sum pk x qk = 0.972$$

$$KR-20 = \frac{5}{4} \left(1 - \frac{0.972}{2.70} \right)$$

$$= 1.25 (1 - 0.36)$$

$$= 1.25 (0.64) = 0.80$$

ب- ألفا كرونباخ (Cronbach Alpha):

يُعتبر ألفا كرونباخ من أكثر المؤشرات الإحصائية شيوعاً لتقييم الاتساق الداخلي للمقاييس ويُعد حالة خاصة من معادلة كيودر-ريتشاردسون (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 مريضاً 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 و 1، 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 بعداً مختلفاً 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 تجانساً وبالتالي يُستحسن استبعاد هذا البند.

وتشير الأدبيات إلى أن المقياس ذو الاتساق الداخلي الجيد يجب أن تظهر بنوده ترابطاً متوسطاً بين 0.70 و 0.90 فالقيم المنخفضة تشير إلى أن البنود قد تقيس سمات مختلفة والقيم العالية جداً قد تعني تكرار البنود وتقل صدق المقياس الكلي (Portney, et al, 2008, p.616). ويعطى ألفا كرونباخ بالمعادلة التالية:

$$\text{Coefficient Alpha} = \left(\frac{K}{K-1} \right) \left[1 - \frac{\sum_{k=1}^p S_k^2}{S_{\text{Total}}^2} \right]$$

حيث:

K = العدد الكلي لمفردات المقياس.

Sk^2 = تباين المفردة K .

S^2 = التباين الكلي للمقياس.

مثال تطبيقي:

نفترض أن لدينا اختباراً يتكوّن من خمس مفردات وقد طُبّق على ستة طلاب. يجيب كل طالب عن كل مفردة على مقياس ليكرت خماسي يتراوح من 1 إلى 5 (حيث تمثّل القيم الأعلى درجة أكبر من السمة المقاسة)، سجّلت استجابات الطلاب كما في الجدول الآتي:

جدول 14 معطيات لمثال لحساب معامل ألفا كرونباخ

المجموع	المفردة 5	المفردة 4	المفردة 3	المفردة 2	المفردة 1	
21	5	5	4	3	4	الطالب 1
14	3	3	2	3	3	الطالب 2
10	1	2	2	3	2	الطالب 3
20	4	3	5	4	4	الطالب 4
14	3	2	4	3	2	الطالب 5
10	3	1	2	2	2	الطالب 6
$S^2 = 22.57$	1.77	1.87	1.77	0.40	0.97	Sk^2

- الخطوة 1: التذكير بصيغة تباين العينة (Sample Variance):

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{N - 1}$$

الخطوة 2: إيجاد التباين لكل مفردة:

❖ حساب تباين المفردة 1 (القيم: 4؛ 3؛ 2؛ 4؛ 2؛ 2).

- المتوسط: $2.833 = 6 / 17 = 6 / (2 + 2 + 4 + 2 + 3 + 4)$

- الانحرافات عن المتوسط:

$1.167 = 4 - 2.833$ ؛ $0.833 = 2 - 2.833$ ؛ $0.167 = 3 - 2.833$ ؛ $1.167 = 4 - 2.833$

$0.833 = 2 - 2.833$ ؛ $0.833 = 2 - 2.833$

- مربعاتها: 1.361؛ 0.028؛ 0.694؛ 1.361؛ 0.694؛ 0.694

- مجموع المربعات: 4.833

- تباين المفردة $1 = 4.833 / (1-6) = 0.967$.

❖ وبنفس الطريقة نقوم بحساب بقية تباينات المفردات المتبقية والنتيجة في الجدول أعلاه.

الخطوة 3: إيجاد التباين الكلي لدرجات الطلبة:

❖ لحساب تباين المجموع الكلي للدرجات، نقوم بالتالي:

- المتوسط الكلي للمجاميع: $14.83 = 6 / 89 = 6 / (10 + 14 + 20 + 10 + 14 + 21)$

- حساب الانحرافات عن المتوسط ومربعاتها:

- الطالب 1: $38.07 = (21 - 14.83)^2 = (6.17)^2$

- الطالب 2: $0.69 = (14 - 14.83)^2 = (0.83)^2$

- الطالب 3: $23.33 = (10 - 14.83)^2 = (4.83)^2$

- الطالب 4: $26.73 = (20 - 14.83)^2 = (5.17)^2$

- الطالب 5: $0.69 = (14 - 14.83)^2 = (6.17)^2$

- الطالب 6: $23.33 = (10 - 14.83)^2 = (6.17)^2$

- حساب مجموع المربعات وقسمتها على (N-1): $0.69 + 26.73 + 23.33 + 0.69 + 38.07$

$22.57 = 5 / 112.84 = (1-6) / (23.33 +$

الخطوة 4: وبتطبيق معادلة ألفا كرونباخ نتحصل على النتائج التالية:

$$\alpha = \frac{5}{4} \left(1 - \frac{0.97 + 0.40 + 1.77 + 1.87 + 1.77}{22.57} \right)$$

$$\alpha = 1.25 \left(1 - \frac{6.78}{22.57} \right)$$

$$\alpha = 1.25 (1 - 0.30)$$

$$\alpha = 1.25 (0.70) = \mathbf{0.875}$$

- التفسير: يلاحظ أن قيمة ألفا كرونباخ تقريبا تساوي 0.88 وهذا يدل على ثبات داخلي جيد أي أن

القيمة المرتفعة تُشير إلى تماسك واتساق جيد بين مفردات الاختبار.

ويلاحظ أنه يتوجب على الشخص التأكد من صدق الاختبار أولاً ثم من ثباته لأن الاختبار

الصادق دائماً 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 ثبات الاختبار:

يتأثر معامل ثبات الاختبار بعدة عوامل أساسية تؤثر على دقة وموثوقية القياس، ومن أبرزها:

أ- **طول الاختبار:** 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 درجات الأفراد:** يرتبط معامل الثبات بتنوع الدرجات داخل عينة المفحوصين، كلما كانت المجموعة أكثر تجانساً 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 القدرة في مجموعة الاختبار:** ترتبط دقة الاختبار بمستوى القدرة لدى المفحوصين إذ يختلف أداء الاختبار حسب مستوى الصعوبة للمفردات. لذلك يُنصح بأن يحتوي الاختبار على أغلبية من المفردات متوسطة الصعوبة لتمييز المستويات المتوسطة للقدرة بدقة إلى جانب عدد محدود من المفردات السهلة جداً والصعبة جداً 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.

الفصل السابع

الخصائص السيكمترية لفقرات الاختبار

الأهداف (Objectives):

بعد دراسة موضوع الخصائص السيكمترية 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 وعدالة.

المحتوى (Contents):

- 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

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.

يساعد التحليل الإحصائي لفقرات الاختبار على تحقيق عدة غايات، منها: قياس مدى تحقق الأهداف السلوكية استناداً إلى الفقرات السهلة والصعبة من خلال حساب معامل الصعوبة ومعامل السهولة، وتحديد القوة التمييزية لكل فقرة عن طريق مقارنة أداء الطلبة ذوي التحصيل المرتفع بأداء ذوي التحصيل المنخفض من خلال معامل التمييز. كما تساهم هذه المؤشرات في بناء خطة تعليمية شاملة للصف الدراسي أو خطط علاجية موجهة لبعض الطلبة الذين يحتاجون إلى رعاية خاصة. وبعد تصحيح الاختبار وتحديد درجة كل طالب على كل فقرة، يمكن التعمق في تحليل دلالات العلامات للكشف عن مستويات الصعوبة والسهولة، وقوة التمييز بين الفقرات، وكذلك فعالية المموهات في أسئلة الاختيار من متعدد، فضلاً عن تقدير معامل الثبات. وتُطبق هذه التحليلات بشكل أدق في الاختبارات معيارية المرجع باستخدام أساليب إحصائية مناسبة.

1- معامل صعوبة الفقرة (Item Difficulty Index):

يُستخدم 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 ارتباطاً عكسياً، 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:

$$\text{معامل صعوبة الفقرة} = \frac{\text{عدد الاجابات الخاطئة على الفقرة}}{\text{عدد الاجابات الكلية على الفقرة}} \times 100\%$$

مثال: في اختبار مادة الرياضيات لطلبة السنة الأولى متوسط، تبين أن 20 طالباً قد أجابوا إجابة خاطئة عن الفقرة الأولى من الاختبار، بينما بلغ إجمالي عدد الطلبة المفحوصين 30 طالباً.
- المطلوب هو حساب معامل صعوبة هذه الفقرة؟

$$\text{معامل صعوبة الفقرة} = \frac{20}{30} \times 100\% = 67\%$$

التفسير:

- معامل الصعوبة 0.67 = تقريباً.

- هذا يشير إلى أن الفقرة صعبة نسبياً، لأن حوالي 67% من الطلبة أخطأوا في الإجابة عنها.

- معامل السهولة المقابل:

$$\text{معامل السهولة} = 1 - 0.6667 = 0.3333$$

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 السيكمترية 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 يُفضّل اختيار فقرات يكون مدى صعوبتها معتدلاً بحيث يكون متوسط صعوبة جميع الفقرات قريباً من (0.50) (صلاح علام، 2022).
نيتكو (Nitko, 1983) يرى أن مدى معاملات الصعوبة الذي يقع بين القيم 0.20 و 0.80 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 أيضاً 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.

$$P = \frac{Ru + RL}{T} \times 100$$

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 15 معطيات لمثال لحساب مؤشر صعوبة الفقرة

البدائل	المجموعة العليا	المجموعة الدنيا
أ	0	0
ب*	8	4
ج	3	5
د	4	6
مجموع الطلبة	15	15

وتطبيق المعادلة السابقة:

$$P = \frac{8 + 4}{30} \times 100 = 0.4 = 40\%$$

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

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

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

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

حيث أن:

P = مؤشر صعوبة الفقرة الانشائية

Ru = مجموع علامات الطلاب في الفئة العليا على الفقرة.

RL = مجموع علامات الطلاب في الفئة الدنيا على الفقرة.

T = عدد أفراد المجموعتين.

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

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

- 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، ويسهم في بناء اختبار عادل وفعال.

- 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، ويُعبّر عن النسبة المئوية للطلبة الذين تمكنوا من الإجابة عن الفقرة إجابة صحيحة. يساعد هذا المؤشر المعلم أو الباحث على تحديد الفقرات السهلة جداً والفقرات الصعبة، وبالتالي ضبط مستوى الاختبار بما يتناسب مع قدرات الطلاب، ويمكن استخراج معامل سهولة الفقرة باستخدام المعادلة التالية:

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

وبالنظر إلى المثال السابق فإن معامل السهولة للفقرة الأولى يساوي:

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

يُلاحظ أن مجموع معاملي السهولة والصعوبة لأي فقرة في الاختبار يساوي دائماً 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 عُرف أحد المعاملين يمكن بسهولة معرفة الآخر ولا يحتاج الحكم على السؤال إلى كلا المعاملين معاً 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 تحدياً في تحقيق مستوى الصعوبة المرغوب فيه لكل فقرة.

وكقاعدة عامة يُفضل أن يتراوح معامل صعوبة الفقرات في الاختبار المرغوب فيه بين 0.50 و 0.75 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 عدالة 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:

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

$$\bullet \text{ أفضل معامل الصعوبة لفقرة اختيار من بديلين} = \frac{1 + \left(\frac{1}{2}\right)}{2} = 0.75$$

$$\bullet \text{ أفضل معامل الصعوبة لفقرة اختيار من ثلاثة بدائل} = \frac{1 + \left(\frac{1}{3}\right)}{2} = 0.665$$

- أفضل معامل الصعوبة لفقرة اختيار من خمسة بدائل $0.6 = \frac{(\frac{1}{5})+1}{2}$
- أفضل معامل الصعوبة في حال مراعاة التخمين للفقرة الانشائية أو المقالية هو 0.5.
- الفقرة الانشائية أو المقالية لا يوجد لها معامل صعوبة مصحح.

ولتجنب هذا الخلل الناتج عن التخمين اقترح جيلفورد (Guilford, 1936) قانوناً لتصحيح صعوبة الفقرة من أثر التخمين. ووفق هذا الاقتراح تُعرّف صعوبة الفقرة إحصائياً 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 للإجابات 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:

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

-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 بعيداً 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 اختباراً مكوناً 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 جزءاً 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 تقديراً أدق لمستوى تحصيل الطلاب الحقيقي ويجعل الاختبار أكثر عدالة 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 مؤشراً 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 مناسباً من هذه الناحية. لكن قبل الحكم على جودة السؤال من الضروري تحليل نوعية الطلاب الذين أجابوا عنه، فإذا كانت نسبة الإجابات الصحيحة متساوية تقريباً بين فئة الممتازين والمتوسطين والضعفاء (15% لكل فئة) فهذا يدل على أن السؤال لم يقيم بوظيفته الأساسية وهي التمييز بين الطلاب ذوي التحصيل الجيد وغيرهم. وبالتالي لضمان أن كل سؤال يقي 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.

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 ترتيباً تنازلياً، 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 طالباً. أما إذا كان العدد فردياً مثل 41 طالباً فيتم استبعاد الطالب الوسيط (المرتبة 21) من التحليل لضمان توازن المجموعتين. وعند وجود عدد كبير من الطلاب يكفي أحياناً أخذ ربع الأوراق (25%) أو 27% أو 33% من كل طرف من توزيع العلامات بعد ترتيبها تنازلياً أو تصاعدياً 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 بدلين:

لنأخذ مثالاً عملياً: إذا كان عدد الطلاب في كل مجموعة متطرفة هو 12 طالباً 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 17 معطيات لحساب معامل تمييز لفقرة اختيار من بديلين.

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

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 تميّز 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 جداً 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- حالة حساب معامل التمييز لفقرة اختيار من متعدد مع أربع بدائل:

لنفترض أن لدينا فقرة اختيار من متعدد تحتوي على أربعة بدائل للإجابة والبديل الصحيح هو (ب). 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 18 معطيات لحساب معامل تمييز لفقرة اختيار من متعدد مع أربع بدائل.

البدائل	المجموعة العليا	المجموعة الدنيا
أ	0	0
ب*	10	3
ج	1	3
د	1	6
مجموع الطلبة	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 يُحسب 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 تُظهر قدرة عالية على التمييز بين مستويات الطلاب المختلفة مما يجعلها مناسبة للاختبارات التي تهدف إلى تقييم الأداء بشكل دقيق.

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

عند التعامل مع الفقرات الإنشائية أو المقالية يختلف حساب معامل التمييز قليلاً عن فقرات الاختيار من متعدد نظراً 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 أيضاً 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 تحصيلاً والأقل تحصيلاً 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.

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

S = العلامة المخصصة للفقرة أو السؤال المقالية أو الانشائية.

لنفترض أن لدينا سؤالاً مقالياً مخصصاً له 5 علامات وقد تم جمع درجات الطلاب في المجموعتين المتطرفتين (العليا والدنيا) كما هو موضح في الجدول.

جدول رقم 18 معطيات لحساب معامل تمييز لفقرة مقالية.

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

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

$$D = \frac{Ru - RL}{T \times x}$$

حيث أن:

D = معامل تمييز الفقرة أو السؤال.

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

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

x = العلامة المخصصة للفقرة أو السؤال.

$T = \text{عدد الطلبة الكلي.}$

وبالتطبيق في المعادلة نجد التالي:

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

التفسير:

- قيمة معامل التمييز يساوي $D=0.22$ تشير إلى أن الفقرة تميّز 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 التمييز باستخدام معامل الارتباط الثنائي النقطي 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.

5-2- 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 تقييمية.

وقد قدم أيبيل (Ebel, 1965) محكات احصائية لتحديد مستويات قوة المفردات التمييزية على أساس قيم التمييز الخاصة بها، وتظهر هذه القيم والتسميات المقابلة في الجدول رقم (19) الآتي:

جدول (19) 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-	الفترة ضعيفة وينصح بحذفها

(Crocker & Algina, 2008, 315)

توفر هذه الطريقة طريقة كمية دقيقة لتقييم الفقرات بما يضمن اختيار أسئلة تساهم فعلياً في تمييز مستويات الطلاب المختلفة مما يزيد من موثوقية وجودة الاختبار كأداة تقييمية.

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 وال خاطئة. لذلك يُعتبر تحليل فاعلية المموهات خطوة ضرورية في بناء الاختبار وتقييم 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 مشنت 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 طُبِّقَ على عينة مكوّنة من 30 طالباً.

ولإجراء التحليل يتم إعداد جدول يضم البيانات الأساسية لكل بديل من بدائل الفقرة، بحيث يُظهر:

- 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 20 معطيات لحساب معامل فاعلية المموهات (المشتتات).

البدائل	المجموعة العليا	المجموعة الدنيا	الإجابات الخاطئة
أ*	9	3	-
ب	3	4	7
ج	2	5	7
د	1	3	4
المجموع	15	15	18

وبتطبيق معادلة حساب فاعلية المموهات السابقة، نجد أن المموهات الثلاثة:

$$\bullet \text{ م ت (ب)} = \frac{4-3}{15} = 0.07$$

$$\bullet \text{ م ت (ج)} = \frac{5-2}{15} = 0.20$$

• م ت (د) = $\frac{3-1}{15} = 0.13$

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 يُفترض 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 يُفترض أن يتجهوا نحو الإجابة الصحيحة.

هذا الوضع يدل على أن البديل الخاطئ غير فعال بل قد يُضعف من صدق الفقرة ويمس بجودتها، مما يستلزم إعادة صياغته أو استبداله بمموه أكثر ملائمة وواقعية بحيث يجذب الطلبة ضعاف التحصيل بدلاً من المتميزين.

❖ معايير صياغة المموهات 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 مهمة (غير وظيفية): يجب التأكد بعد التجريب أن كل مموه يُختار من قبل نسبة مقبولة من الطلبة ($\leq 5\%$)، وإلا فإنه يُعتبر غير وظيفي ويضعف من جودة الفقرة.

7- معامل ثبات الفقرة (Item Reliability):

يُتصد 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 دلّ 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.

الفصل الثامن

الاختبارات المدرسية أنواعها وأغراضها

الأهداف (Objectives):

يهدف هذا الفصل إلى توضيح مفهوم الاختبار بصفة عامة، مع التركيز على الاختبار التحصيلي (Achievement Tests) وأهميته في العملية التعليمية. كما يسعى إلى بيان كيفية بناء الاختبارات التحصيلية في إطار النظرية الكلاسيكية للقياس من خلال التمييز بين الاختبارات الموضوعية المرجعية للمعيار (NRT) والاختبارات المرجعية للمحك (CRT). ويتناول الفصل أيضاً موضوع العدالة في التقييم التربوي، والعدالة الاختبارية (Test Fairness)، مع التعمق في العلاقة بين العدالة والإنصاف (Fairness and Equity). إضافة إلى ذلك يركز الفصل على معاني العدالة في الاختبارات وكيفية رصد الأداء التفاضلي للمفردة (DIF) باعتباره مؤشراً على تحقيق العدالة الاختبارية. وبذلك يسعى الفصل إلى تزويد الطالب برؤية شاملة تجمع بين البعد النظري والتطبيقي لمفهوم العدالة في بناء وتفسير الاختبارات التربوية.

المحتوى (Contents):

- 1- مفهوم الاختبار (Test)
- 2- الاختبار التحصيلي (Achievement tests)
- 3- بناء الاختبارات التحصيلية بناءً على النظرية الكلاسيكية في القياس
- 3-1- الاختبارات التحصيلية الموضوعية مرجعية المعيار (Norm-Referenced Tests: NRT)
- 3-2- الاختبار التحصيلية مرجعية المحك (Criterion-Referenced Tests (CRT)
- 4- العدالة في التقييم التربوي
- 5- العدالة الاختبارية (Test Fairness)
- 6- العدالة والإنصاف (Fairness and Equity)
- 7- معاني العدالة في الاختبارات
- 8- الأداء التفاضلي للمفردة والعدالة الاختبارية

تمهيد:

تُعَدّ الاختبارات التحصيلية من أبرز الأدوات التي يعتمد عليها التربويون والباحثون في تقويم نواتج التعلم وتحديد مدى تحقق الأهداف التعليمية. فهي ليست مجرد وسيلة لقياس مستوى الطالب فحسب بل أداة علمية تتيح فهماً أدق للفروق الفردية وتوجيهاً أفضل للعملية التعليمية. وانطلاقاً من ذلك يهدف هذا الفصل إلى تقديم رؤية شاملة حول مفهوم الاختبار عامة والاختبارات التحصيلية خاصة مع توضيح أسس بنائها في ضوء النظرية الكلاسيكية للقياس والتمييز بين الاختبارات المرجعية للمعيار (NRT) والاختبارات المرجعية للمحك (CRT). كما يركز الفصل على قضية محورية في مجال القياس النفسي والتربوي وهي العدالة سواء في التقييم التربوي أو في الاختبار ذاته باعتبارها شرطاً أساسياً لضمان مصداقية النتائج وموضوعيتها. ويتناول الفصل كذلك معاني العدالة في الاختبارات وعلاقتها بمفهومَي الإنصاف والمساواة إضافة إلى إبراز أهمية تحليل الأداء التفاضلي للمفردة (DIF) كأداة لتشخيص مواطن الانحياز والتمييز غير العادل. ومن خلال هذه المعالجة يُمكن هذا الفصل القارئ من إدراك العلاقة الوثيقة بين البناء العلمي للاختبار ومتطلبات العدالة والإنصاف في التقييم التربوي.

1- مفهوم الاختبار (Test):

- يرى كرونباخ (Cronbach) أن الاختبار: هو "أسلوب منظم لمقارنة سلوك شخص ما بسلوك شخص آخر" ويُقصد بالسلوك في علم النفس كل استجابة تصدر عن الفرد نتيجة وجود منبه أو مثير. أما التنظيم فهو سمة أساسية من سمات العلم ما يجعل الاختبار في نظر كرونباخ وسيلة علمية لجمع بيانات كمية حول جوانب محددة من شخصية الفرد.

- أما ثورندايك (Thorndike): فيعرف الاختبار بأنه "طريقة لقياس مقدار شيء ما، انطلاقاً من أن كل ما يوجد يمكن قياسه بكمية معينة". والمقصود بالشيء هنا هو الجوانب المعرفية كالأفكار والمعلومات التي يكتسبها الفرد بوعي أو من دون وعي.

- بينما يرى بين (Bean): أن الاختبار هو "مجموعة من المثيرات التي تُستخدم لقياس بعض العمليات العقلية أو السمات النفسية بصورة كمية أو كيفية"، أي أن الاختبار ينبغي أن يقيس أو يُقَيِّم جانباً معيناً لدى الفرد.

وبصورة عامة يمكن القول إن الاختبار هو "عملية منظمة لقياس عينة من السلوك" (سلمي،

2001، ص12).

2- الاختبار التحصيلي (Achievement tests):

يُعد الاختبار من أبرز أدوات القياس والتقويم المستخدمة لتقدير تحصيل الطلبة وأكثرها شيوعاً. ويُقصد به في مجال القياس والتقويم أسلوب منظم يهدف إلى تحديد مدى امتلاك الفرد لسمة أو معرفة معينة من خلال استجاباته لعينة من المثيرات التي تمثل تلك السمة (عودة، 1985).

- ويُعرف الاختبار التحصيلي: بأنه إجراء منظم يقيس مقدار ما تعلمه الطلبة في موضوع محدد وفقاً للأهداف الموضوعية الأمر الذي يجعله وسيلة مهمة لتحسين أساليب التعلم والمساهمة في جودة التخطيط ودقة التنفيذ وتقويم النتائج (Gronlund, 1977).

كما يرى سعادة (1984) أن الاختبار التحصيلي عملية منظمة يتم من خلالها ملاحظة سلوك الطلبة والتأكد من مدى تحقيقهم للأهداف التعليمية وذلك عبر طرح مجموعة من الفقرات أو الأسئلة والاستناد إلى مقاييس متنوعة لتوصيف استجاباتهم.

وبصورة عامة فإن الاختبار التحصيلي يرتبط دائماً بمادة دراسية محددة سبق تدريسها فعلياً مما يعني أنه لا جدوى من إعداد اختبارات تحصيلية لمواد لم يتم تدريسها بعد. ومن هنا تتجلى أهمية الاختبار التحصيلي كأداة للحكم على ما تم إنجازه بالفعل من عملية التدريس (القومي، 2008، ص05).

3- بناء الاختبارات التحصيلية بناءً على النظرية الكلاسيكية في القياس:

لقد أولى علماء القياس والتقويم التربوي اهتماماً كبيراً بابتكار أساليب متنوعة لقياس مدى تحقق الأهداف التعليمية وفاعلية البرامج التربوية المنفذة، إضافة إلى قياس مستوى التحصيل الدراسي للمتعلمين والحكم عليه وفق معايير موضوعية محددة. ومن هذا الاهتمام نشأت أنماط متعددة لبناء الاختبارات التحصيلية تبعاً للغرض من استخدامها، سواء كان الغرض تشخيصياً للتعرف على جوانب القوة والضعف لدى المتعلمين أو تكوينياً لمتابعة تقدمهم خلال العملية التعليمية أو تجميعياً للحكم على نتائج تعلمهم النهائية.

وتستند هذه الاختبارات في إطار النظرية الكلاسيكية للقياس إلى مبدأ أساسي يتمثل في أن الدرجة التي يحصل عليها الطالب تمثل مجموعاً لإجاباته الصحيحة مع افتراض أن هذه الدرجة تعكس بشكل مباشر مستوى تحصيله في المادة الدراسية. ومن هنا ركزت الجهود البحثية على كيفية صياغة الفقرات الاختبارية بصورة دقيقة وضبط مستوى الصعوبة والتمييز فيها والتحقق من صدقها وثباتها باعتبارها مؤشرات جوهرية لجودة الاختبار.

كما أن بناء الاختبارات التحصيلية وفق هذا المنظور لا يقتصر على الجانب الكمي فقط بل يمتد ليشمل الجوانب التربوية والعملية كإسهام الاختبار في تحسين طرائق التدريس وتوجيه المعلم والمتعلم نحو تحقيق أهداف المنهج وتوفير بيانات علمية تدعم عملية اتخاذ القرار التربوي. وبهذا يصبح الاختبار أداة لا تقتصر على القياس فحسب بل وسيلة لتقويم شامل ومتكامل للعملية التعليمية.

3-1- الاختبارات التحصيلية الموضوعية مرجعية المعيار (Norm-Referenced Tests: NRT):

تُعدّ الاختبارات الموضوعية مرجعية المعيار من أكثر الأدوات شيوعاً في مجال القياس التربوي إذ تهدف إلى مقارنة أداء الفرد بأداء أقرانه ضمن مجموعة معيارية محددة مما يسمح بالحكم على مستوى تحصيله النسبي لا المطلق. وتستند هذه الاختبارات إلى أسس علمية دقيقة في بنائها وصياغة فقراتها الأمر الذي يعزز من صدقها وثباتها.

3-1-1- خطوات إعداد اختبار تحصيلي موضوعي مرجعي المعيار (NRT):

أشار كل من مراد وسليمان (2005، ص.143) وأمين (2009، ص.237) إلى مجموعة من الخطوات المنهجية لبناء هذا النوع من الاختبارات، والتي يمكن تلخيصها فيما يلي:

1- تحديد الأهداف التعليمية/ التدريسية وتقدير وزنها النسبي:

عقب الانتهاء من دراسة مقرر معين يتم تحديد الأهداف التي سبق أن حظيت بالاهتمام والتدريب ضمن الجوانب المعرفية والانفعالية والنفس حركية. وبعد ذلك، تُقدّر الأهمية النسبية لكل هدف على حدة ويمكن الاستعانة بأراء الخبراء في الميدان لتحقيق ذلك. ويقصد بالخبراء هنا المعلم المتميز ذو خبرة تزيد عن خمس سنوات في التدريس أو المعلم الأول أو موجه المادة أو أحد المتخصصين في مراكز البحوث التربوية ذات الصلة.

1-2- تحليل محتوى المقرر الدراسي السابق وتحديد الوزن النسبي لعناصره:

من المؤكد أن تحليل مكونات المقرر الدراسي أو مكونات العمل (المهارة) المكتسبة مسبقاً أو تحليل المهمة وتحديد موضوعاتها الرئيسة والفرعية التي ركّز عليها المعلم أثناء التدريس مما يساهم في تسهيل إعداد اختبار يتمتع بالصدق. ويتم تحديد الأهمية النسبية لعناصر المحتوى الرئيسة والفرعية أو لمكونات المهارة أو المهمة المدروسة سابقاً استناداً إلى واحد أو أكثر من المحكات التالية:

أ- عدد الصفحات التي يخصصها الكتاب لكل موضوع مقارنة بالعدد الكلي للصفحات (وهو محك ضعيف).

ب- الزمن الذي استغرقه تدريس كل موضوع على حدة.

ج- آراء الخبراء والمتخصصين في المادة العلمية حول كل موضوع ويُعد هذا المحك الأكثر دقة وموثوقية (أمين، 2009، ص.238).

1-3- إعداد جدول مواصفات الاختبار التحصيلي:

يُعد جدول المواصفات مخططاً تفصيلياً ثنائي البعد يتمثل أحد أبعاده في قائمة الأهداف (نواتج التعلم المرغوب تحقيقها)، بينما يمثل البعد الآخر عناصر المحتوى التي يتضمنها الاختبار، وتتمثل أبرز فوائد إعداد جدول المواصفات فيما يلي:

أ- يساعد واضع الاختبار على توزيع الأسئلة بشكل عادل على مختلف أجزاء المقرر الدراسي بحيث يحصل كل جزء على وزنه النسبي أو أهميته الحقيقية عند صياغة الأسئلة.

ب- يضمن عدم اقتصار الأسئلة على مستوى واحد داخل المجال (معرفي أو نفسي-حركي) بل يشمل جميع المستويات المطلوب تحقيقها.

ج- يمنح التلميذ الذي اجتهد في المذاكرة شعوراً بالاطمئنان والعدالة حيث لا يتضرر من طبيعة الأسئلة.

د- يحقق شمولية في تغطية المحتوى عند إعداد أسئلة الاختبار مما يعزز صدق المحتوى (مراد وسليمان، 2005، ص.146).

1-4- صياغة مفردات (أسئلة) الاختبار التحريري الموضوعي:

تُعد صياغة الأسئلة الجيدة عملية طويلة ومعقدة شأنها شأن صياغة الأهداف التعليمية والتدريسية والتقويمية وأهداف القياس الأمر الذي يفرض على المعلمين وواضعي الاختبارات في مراكز البحوث التربوية عدم التقليل من أهميتها سواء كانت الأسئلة موضوعية أو غير موضوعية بل العمل على تحسينها من حيث الشكل والمستوى. ويتحقق ذلك من خلال التمكن من مادة التخصص والمراجعة المستمرة للأسئلة المعدة مسبقاً وتقبل الملاحظات من زملاء التخصص والاستفادة من إجابات التلاميذ في كشف أوجه القصور بالإضافة إلى الاطلاع على كتب القياس للتعرف على الأشكال المختلفة للأسئلة وشروط صياغتها والتدريب على استخدام بعض الأساليب الإحصائية لتحليل مفردات الاختبار واكتشاف عيوبها والعمل على تعديلها. وفيما يلي سيتم عرض أنواع الأسئلة التحريرية الموضوعية وأسس تصنيفها:

1- التصنيف حسب درجة تعقدها:

يمكن تصنيف الأسئلة التحريرية الموضوعية وفقاً لدرجة تعقيدها إلى عدة أنواع، حيث تشمل: أسئلة الاختيار من بديلين مثل الصورة البسيطة، والتصحيحية، والعنقودية؛ وأسئلة المزاجية التي

تنقسم إلى الصورة البسيطة والتصنيفية؛ وأسئلة الإجابات القصيرة التي تتضمن أسئلة التكملة والتعيين والمماثلة؛ بالإضافة إلى أسئلة إعادة الترتيب؛ ثم الأسئلة التفسيرية أو أسئلة حل المشكلات؛ وأخيراً أسئلة الاختيار من متعدد التي تأخذ عدة صور مثل: اختيار الإجابة الصحيحة، أو الخاطئة، أو أفضل إجابة، أو الترتيب الصحيح، أو الترتيب الخاطئ.

جدول رقم 21 يوضح المقارنة بين صور وأشكال الأسئلة.

نوع السؤال	الصور / الأشكال الفرعية
أسئلة الاختيار من بديلين	-الصورة البسيطة – التصحيحية – العنقودية
أسئلة المزاجية	-الصورة البسيطة – التصنيفية
أسئلة الإجابات القصيرة	-أسئلة التكملة -أسئلة التعيين - أسئلة المماثلة
إعادة الترتيب	—
الأسئلة التفسيرية / حل المشكلات	—
الاختيار من متعدد	-اختيار الإجابة الصحيحة -اختيار الإجابة الخاطئة - أفضل إجابة - الترتيب الصحيح - الترتيب الخاطئ

2- التصنيف حسب العمليات العقلية المتضمنة أثناء الإجابة:

أ- أسئلة انتقاء الاستجابة: وتشمل أسئلة الاختيار من بديلين، أسئلة المزاجية، أسئلة إعادة الترتيب، وأيضاً أسئلة الاختيار من متعدد.

ب- أسئلة إنتاج الاستجابة: وتشمل أسئلة الإجابات القصيرة مثل أسئلة التكملة، التعيين، والمماثلة (أمين، 2009، ص.246).

1-5- إخراج الورقة الامتحانية:

يتضمن إعداد الورقة الامتحانية عدة جوانب أساسية:

أ- شكل الورقة الامتحانية (نظام عرض الأسئلة):

يُستحسن ترتيب الأسئلة وفقاً لدرجة تعقّد النشاط العقلي المطلوب للإجابة مع تجميع الأسئلة المتشابهة في صياغتها تحت تعليمات خاصة بها وذلك لضمان بقاء التلميذ في حالة من التهيؤ العقلي أثناء الحل.

ب- صياغة الأسئلة:

يجب أن تتميز الأسئلة بالدقة والوضوح مع تحديد جيد للمفردة أو المطلوب وكتابتها بخط واضح خالٍ من الأخطاء الإملائية أو العلمية.

ج- تعليمات الاختبار:

تهدف إلى توضيح طريقة الإجابة وتسجيلها مع بيان الزمن المخصص للاختبار إضافة إلى تنبيه التلاميذ بعدم اللجوء إلى التخمين في حال كان للتخمين أثر في التصحيح (مراد وسليمان، 2005، ص.205).

1-6- تحليل مفردات الاختبار (التجريب والمعالجة الإحصائية):

بعد الانتهاء من إعداد الورقة الامتحانية (مع إمكانية إعداد نماذج متكافئة) يتم إخضاعها للتجريب من خلال تطبيقها على عينة ممثلة للمجتمع المستهدف وذلك بهدف التحقق من صدق الاختبار (أي قياسه لما أُعد لقياسه) وثباته (إعطاؤه نفس النتائج عند إعادة التطبيق على العينة ذاتها وتحت نفس الظروف) إضافة إلى الكشف عن خصائص كل مفردة من مفردات الاختبار، وتتمثل المعالجة الإحصائية لهذه المفردات في:

أ- تحديد مستوى السهولة والصعوبة لكل مفردة.

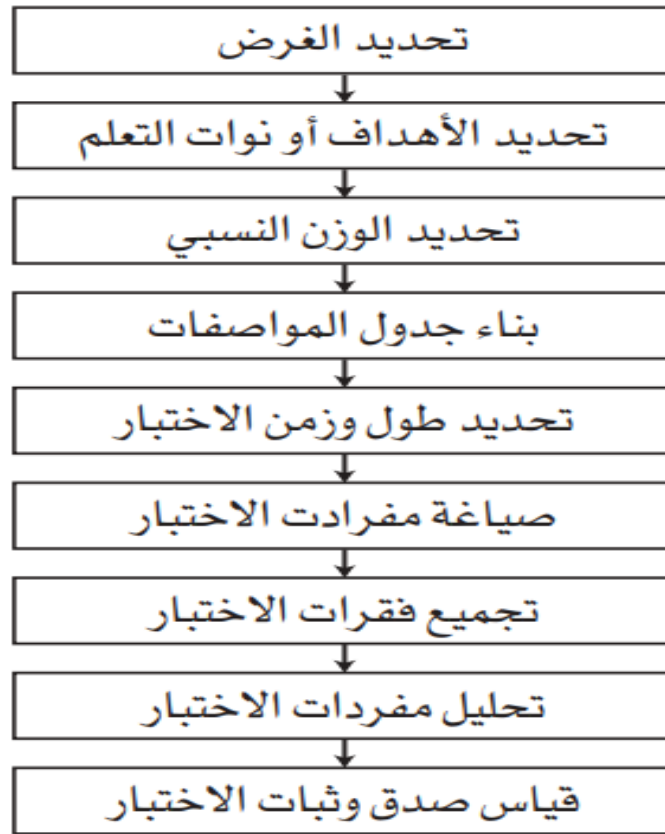
ب- حساب معامل التمييز لكل مفردة.

ج- فحص فعالية المشتتات في حال الاختبارات التحصيلية الموضوعية (أمين، 2009، ص.312).

1-7- الاختبار في صورته النهائية:

بعد الانتهاء من المعالجات الإحصائية لمفردات الاختبار يتم استبعاد المفردات شديدة السهولة أو شديدة الصعوبة مع الإبقاء على المفردات التي تقع معامل سهولتها بين (0.30 – 0.70) أي ما يعادل (30% – 70%). كما تُستبعد المفردات التي تفتقر إلى التمييز بين التلاميذ إذا كان معامل تمييزها أقل من (0.20) بينما يتم الاحتفاظ بالمفردات التي تبلغ قيمة معامل تمييزها (0.30) فأكثر. إضافة إلى ذلك يتم الحفاظ على المفردات ذات المشتتات الفعالة أي التي تستطيع جذب أفراد المجموعة الدنيا أكثر من المجموعة العليا. وفي حال توافر عدد كبير من المفردات الصالحة يمكن إعداد اختبارين متكافئين بحيث يصبح كل منهما جاهزاً للتطبيق النهائي (مراد وسليمان، 2005، ص. 211-221).

شكل (28) مخطط تفصيلي لخطوات بناء الاختبار التحصيلي (هيئة التأطير، ص32).



2- أنواع مفردات (أسئلة) الاختبارات التحصيلية:

تنقسم مفردات الاختبارات التحصيلية إلى قسمين رئيسيين:

أولاً: مفردات الاختبارات المقالية:

تُعدّ من أقدم أشكال الأسئلة وأكثرها شيوعاً في الممارسات التربوية وتُستعمل عادةً كأداة رئيسية لتقدير مستوى تحصيل التلميذ، وتمتاز هذه المفردات بعدة إيجابيات لكنها في المقابل تتضمن بعض السلبيات مما يستلزم مراعاة شروط محددة عند استخدامها لضمان فاعليتها في قياس التحصيل الدراسي للتلاميذ.

1- قواعد صياغة الأسئلة المقالية:

تتطلب صياغة الأسئلة المقالية الالتزام بمجموعة من القواعد الأساسية لضمان فاعليتها، ومن أبرزها:

- مراعاة التدرّج في مستوى الصعوبة بحيث تبدأ الأسئلة بالبسيط ثم المتوسط وصولاً إلى الأكثر تعقيداً.

- تصميم الأسئلة بشمولية تضمن العدالة في تغطية مختلف أجزاء المقرر المراد تقييم التلامذة فيه.

- ارتباط كل سؤال بأهداف المقرر التعليمية.
- تحديد المشكلة المطلوب من التلميذ معالجتها بدقة.
- صياغة التعليمات الخاصة بالسؤال بوضوح تام.
- البدء في صياغة أسئلة المقال بأفعال أو عبارات مثل *تقارن، انقد، مّيز، بيّن، حلّ*.
- تجنّب استعمال كلمات مثل: *من، متى، ماذا*، لأنها تقتصر غالبًا على استدعاء المعلومات فقط، خاصة إذا كان المتعلم قد درس الموضوع مسبقًا.
- التأكد من أن المطلوب في السؤال يعكس الأداء الحقيقي المراد قياسه عند التلاميذ.
- ملائمة طول الإجابة ودرجة تعقيدها مع مستوى نضج التلاميذ.
- عدم اللجوء للأسئلة الاختيارية إلا إذا كان الهدف التعليمي يفرض ذلك.
- وضوح الصياغة والابتعاد عن الألفاظ الغامضة (ميمي، 2014، ص.30)؛ (سوسن، 2014، ص.245).

2- المزايا الإيجابية للاختبارات التحريرية (المقالية):

- تتميز الاختبارات المقالية بعدد من الخصائص الإيجابية، من أبرزها:
 - (أ) تمنح التلميذ فرصة للتأمل وتنظيم أفكاره بحرية دون الشعور بالحرص أو التوتر الذي قد يرافق الامتحانات الشفهية، كما تتيح له وقتًا أطول للتعديل أو التصحيح أو الإضافة.
 - (ب) تسهم في تنمية قدرات التلميذ الفكرية خاصة في مجال التعبير وتنظيم الأفكار وفهم الموضوعات.
 - (ج) إعدادها يتطلب وقتًا وجهدًا أقل مقارنة بغيرها من الاختبارات.
 - (د) احتمالات الغش فيها أقل بكثير مما هو عليه الحال في الاختبارات الموضوعية.
 - (هـ) تكشف عن قدرات التلامذة في النقد والتقويم (سوسن، 2014، ص.243).

3- سلبيات وعيوب الاختبارات التحريرية (المقالية):

- رغم ما تتميز به الاختبارات المقالية من جوانب إيجابية إلا أنها لا تخلو من مجموعة من العيوب، يمكن تلخيصها فيما يلي:
 - (أ) غالبًا ما تقتصر إلى صدق المحتوى وثبات الفكرة وقد تركز على جانب محدد من المنهج وتهمل جوانب أخرى.
 - (ب) عملية تصحيحها تستغرق وقتًا طويلاً، كما تتطلب توحيد الخبرة في تقدير الدرجات الأمر الذي قد ينعكس سلبًا على مصداقية التقويم.

(ج) قد تسهم في إبراز عامل الحظ أو التوقع في إجابات التلاميذ.

(د) قد تؤدي إلى التباس في فهم المطلوب من السؤال، مما يفتح المجال لتعدد الإجابات.

(هـ) تتأثر الإجابات بعوامل فردية كمستوى المهارة اللغوية وحسن التعبير مما يؤدي إلى تفاوت في النتائج.

(و) قد تحدث حالة من الإرهاق أو الضجر لدى التلميذ وهو ما يؤثر على جودة إجاباته.

(ز) تعاني من ضعف في ثبات وموضوعية التصحيح حيث يمكن أن تتأثر بعوامل ذاتية مرتبطة بالحالة النفسية للمصحح أو بعوامل شكلية تتعلق بورقة الإجابة كوضوح الخط وتنظيمه.

(ح) قد يكون صدقها ضعيفاً إذ من المحتمل أن تقتصر على قياس معارف قائمة على الذاكرة فقط، في حين أن الهدف من التقويم قد يكون الكشف عن عمليات عقلية أعمق (سوسن، 2014، ص.243) (ميمي السيد، 2014، ص.30).

ثانياً: المفردات الموضوعية:

تتكون الاختبارات الموضوعية من مجموعة من المفردات (الأسئلة) التي تُبنى بحيث يكون لكل سؤال منها إجابة واحدة صحيحة فقط، ويُطلب من المفحوص إما استدعاؤها من الذاكرة أو اختيارها من بين عدة بدائل مقترحة. ويُعدّ النوع المبني على نظرية الاستجابة للمفردة أحد أهم أشكال هذه الاختبارات إذ يُحقق متطلبات الموضوعية في القياس ويضمن توافق درجات التلامذة مع خصائص المفردات المستخدمة، وذلك بفضل الأسس التي تقوم عليها هذه النظرية مما يتيح قياس القدرات الحقيقية للتلامذة بدقة أكبر.

أما الهدف الرئيس من الإلمام بمختلف أنواع المفردات فيتمثل في منح التلميذ والمعلم ومُعدّ الأسئلة فرصة التعرف عليها واكتساب خبرة في التعامل معها بحيث يمتلك واضع المفردات القدرة على صياغتها بكفاءة والتدرب على إعداد نماذج مشابهة لها إلى جانب معرفة الخصائص الفنية المميزة لكل نوع. وبهذا يتمكن من إنشاء بنك أسئلة متقن الصياغة يمكن الاستعانة به لاحقاً عند إعداد أوراق الامتحان بما يتوافق مع الغايات التي يحددها المعلم أو المؤسسة التعليمية أو صانع القرار التربوي (ميمي السيد، 2014، ص.30).

1- أنواع أسئلة (مفردات) الاختبارات التحصيلية الموضوعية:

1-1- أسئلة الاختيار من بديلين:

تُعرف أسئلة الاختيار من بديلين بعدة مسميات كأسئلة نعم/لا، صح/خطأ، أو موافق/غير موافق. ويُقاس من خلالها مستوى التذكر سواء باسترجاع المصطلحات، الحقائق، المفاهيم، القوانين،

النظريات، التواريخ أو الأماكن، وكذلك طرق وأساليب التحقق من صحة الآراء وتكرار التعميمات. كما تُستخدم لقياس مستوى الفهم من خلال الترجمة بين صور مختلفة أو الانتقال بين المستويات إضافةً إلى التفسير، التنبؤ، وإدراك العلاقات، وتأخذ هذه الأسئلة ثلاثة أشكال رئيسية:

أ- **الصورة البسيطة:** وهي أكثر الأشكال شيوعاً حيث يُقدّم للتلميذ عبارة أو جملة ويُطلب منه تحديد ما إذا كانت صحيحة أم خاطئة أو الاختيار بين بديلين متقابلين، مثل: (نعم/لا).

ب- **الصورة التصحيحية:** وفيها يُقدّم للتلميذ جملة خاطئة ويُطلب منه تصحيح الخطأ أو تحديد الجزء غير الصحيح فيها مما يختبر دقة الفهم والقدرة على تعديل المعلومات.

ج- **الصورة العنقودية (التجميعية):** وتتضمن مجموعة من العبارات المرتبطة بموضوع واحد، بحيث يُطلب من التلميذ الحكم على كل عبارة من حيث صحتها أو خطئها الأمر الذي يساعد على قياس مدى شمولية المعرفة ودقتها في مجال معين.

1-2- أسئلة (مفردات) المزاوجة / المقابلة - المطابقة:

يطلق على أسئلة المزاوجة أسماء عدة بينها أسئلة المقابلة وأسئلة المطابقة. وتعد من الأدوات الفعّالة في قياس مستوى المعرفة (التذكر) سواء تعلق الأمر بتذكر الخصوصيات، مثل: (المصطلحات، الأحداث، التواريخ، الأماكن) أو تذكر طرق ووسائل تناول الخصوصيات، مثل: (المحكات التي تستخدم للحكم على صحة الآراء) أو حتى تذكر التعميمات مثل أسس التعلم ومبادئ اكتساب اللغة. وتمتاز هذه الأسئلة بتنوع صورها مما يسمح للمعلم باستخدامها حسب طبيعة الهدف المطلوب قياسه:

(أ) الصورة البسيطة.

(ب) الصورة التصنيفية.

(ج) أسئلة إعادة الترتيب.

(د) أسئلة الإجابات القصيرة.

1-3- الأسئلة التفسيرية (أسئلة حل المشكلات):

يقوم التفسير على قدرة الفرد في الشرح أو التلخيص أو الترجمة وهو يُعد من العمليات العقلية الأساسية التي لا غنى عنها لحدوث التفكير بمختلف أنواعه (الابتكاري، الناقد، التقاربي، الاستدلالي). وتبرز قيمة هذا النوع من الأسئلة في مجال الرياضيات والعلوم بشكل خاص، حيث تُسهم في تنمية الفهم العميق وتدريب المتعلم على ربط المعطيات واستنتاج العلاقات وحل المشكلات على نحو منهجي.

1-4- أسئلة (مفردات) الاختيار من متعدد:

تعد أسئلة الاختيار من متعدد من أكثر أشكال الأسئلة الموضوعية شيوعاً وفاعلية، إذ تجمع في بنيتها خصائص معظم الأنواع الأخرى مثل: أسئلة البديلين، المزاوجة، إعادة الترتيب، الإجابة القصيرة، وحتى الأسئلة التفسيرية. وتمتاز بقدرتها على قياس مختلف المستويات العقلية من التذكر والفهم إلى التحليل والتقويم إذا صيغت بعناية، ولهذا فهي تُستخدم بكثرة في الاختبارات التحصيلية والمقاييس المعيارية والدولية نظراً لدقتها في التمييز بين المتعلمين وسهولة تصحيحها الموضوعي ومرونتها في تغطية نطاق واسع من المحتوى الدراسي (أمين، 2009، ص. 245-279).

أ- مميزات وعيوب أسئلة (مفردات) الاختيار من متعدد:

جدول رقم 22 يوضح ملخص لمميزات وعيوب أسئلة الاختيار من متعدد.

المميزات	العيوب
1. يمكن أن تقيس جميع المستويات العقلية المعرفية التي اقترحها بلوم إذا أحسن إعدادها.	1. لا تصلح لقياس مخرجات التعلم التي تتعلق بالتعبير الكتابي والمتمثلة في: التأليف، والتنظيم وتوليد الأفكار (الابتكار) وهذه المخرجات يصلح لقياسها أسئلة المقال بنوعها.
2. سهولة في تصحيحها ويمكن التحكم في مستوى سهولة وصعوبة الأسئلة عن طريق زيادة التشابه بين البدائل.	2. صعوبة إعدادها وذلك لصعوبة الحصول على مشتتات على درجة عالية من الجودة.
3. يقل فيها التخمين إلى أقل حد ممكن.	3. تحتاج إلى وقت في إعدادها وكذلك وقت الإجابة عليها وخاصة إذا كانت الأسئلة تتطلب تمييزاً دقيقاً بين البدائل.
4. تغطي جزء كبير من المحتوى المقرر الدراسي.	4. تكلفة مادياً نتيجة استهلاك كمية كبيرة من أوراق والطباعة وكذلك أجور الطباعة.
5. توفر للمعلم وسيلة جيدة للتشخيص في مجال التحصيل الدراسي، أي تحديد نقاط القوة والضعف لدى التلامذة.	5. الغش فيها أسهل بالقياس إلى أسئلة المقال.
6. درجة الصدق والثبات مرتفعة بالمقارنة مع باقي أنواع الأسئلة الموضوعية.	6. لا تقيس قدرات التفكير الابتكاري.
7. تساعد في تنمية قدرة التلميذ على حل المشكلات من خلال تدريبه على التمييز بين الحلول المقترحة، ثم اتخاذ القرار المناسب.	

ب- شروط صياغة أسئلة (مفردات) الاختيار من متعدد:

1- الشروط التي تتعلق برأس السؤال (المتن):

يشكل متن السؤال (أو رأسه) العنصر المحوري في صياغة أسئلة الاختيار من متعدد، حيث يحدد بدقة طبيعة المشكلة أو المهمة المطلوب من التلميذ إنجازها. ولضمان جودة هذا الجزء لا بد من مراعاة مجموعة من الشروط الأساسية، نوجزها فيما يلي:

- ارتباط السؤال بالمخرجات التعليمية الهامة: ينبغي أن يتمحور السؤال حول ناتج تعليمي أساسي وهاذف وليس حول جزئيات هامشية أو ثانوية لا تضيف قيمة حقيقية لقياس مستوى التحصيل.

- وضوح المشكلة ودقتها: يجب أن يتضمن رأس السؤال مشكلة واضحة ومحددة بشكل جيد بحيث يتمكن التلميذ من فهم المطلوب دون الحاجة إلى الاطلاع على البدائل لتفسير المتن.

- الاقتصاد في التكرار: إذا كان متن السؤال على شكل عبارة ناقصة يفضل وضع الكلمات المتكررة في متن السؤال بدلاً من إعادة تكرارها في كل بديل مما يسهل عملية القراءة ويجعل البدائل أكثر تركيزاً.

- تجنب التلميحات غير المقصودة: من المهم عدم تضمين أي إشارات أو ارتباطات لفظية بين متن السؤال والإجابة الصحيحة، لأن مثل هذه التلميحات قد تقود التلميذ إلى الحل الصحيح دون أن يمتلك المعرفة الحقيقية.

- الحذر من صياغة النفي: يستحسن الابتعاد عن استخدام النفي في صياغة المتن، وإذا استُخدم النفي لضرورة معينة، يجب توضيحه بشكل جلي، مثلاً: (وضع خط تحت أداة النفي) حتى لا يقع التلميذ في الالتباس، وتجدر الإشارة إلى أن صياغة المفردات بصيغة موجبة أكثر ملاءمة لقياس النواتج التعليمية بدقة.

توضيح المصطلحات: إذا كان السؤال يقيس معرفة التلميذ بتعريف مصطلح معين، فمن الأفضل إدراج المصطلح في متن السؤال نفسه وتضمين البدائل لتعريفات محتملة مما يسهل عملية التمييز بين الصحيح وغير الصحيح.

ومن خلال هذه الشروط يتضح أن جودة صياغة متن السؤال تمثل خطوة أساسية لضمان موضوعية السؤال وفاعليته في قياس التحصيل (مراد وسليمان، 2005، ص.196) (أمين، 2009، ص.304).

1- الشروط التي تتعلق بالبدائل:

تُعد البدائل من العناصر الجوهرية في صياغة أسئلة الاختيار من متعدد فهي التي تُمكن من قياس مدى قدرة التلميذ على التمييز بين الصحيح والخاطئ واختيار الأنسب من بين مجموعة من الخيارات، ومن أهم هذه الشروط:

أ- أن يكون هناك بديل واحد صحيح أو على الأقل أفضل الإجابات مما يُحافظ على صدق السؤال ويُجنّب التلميذ الحيرة غير المنهجية.

ب- ضرورة أن تكون البدائل متجانسة في محتواها ومرتبطة مباشرة بمجال المشكلة، بحيث لا يسهل استبعاد أي بديل دون الرجوع إلى التفكير المنطقي مع الإشارة إلى أن صعوبة السؤال تزداد كلما زاد التجانس بين البدائل.

ج- توزيع الإجابة الصحيحة توزيعاً عشوائياً بين البدائل، وذلك لتجنب أن يكتشف التلميذ أي نمط ثابت في الترتيب يساعده على الوصول إلى الحل من دون جهد معرفي حقيقي.

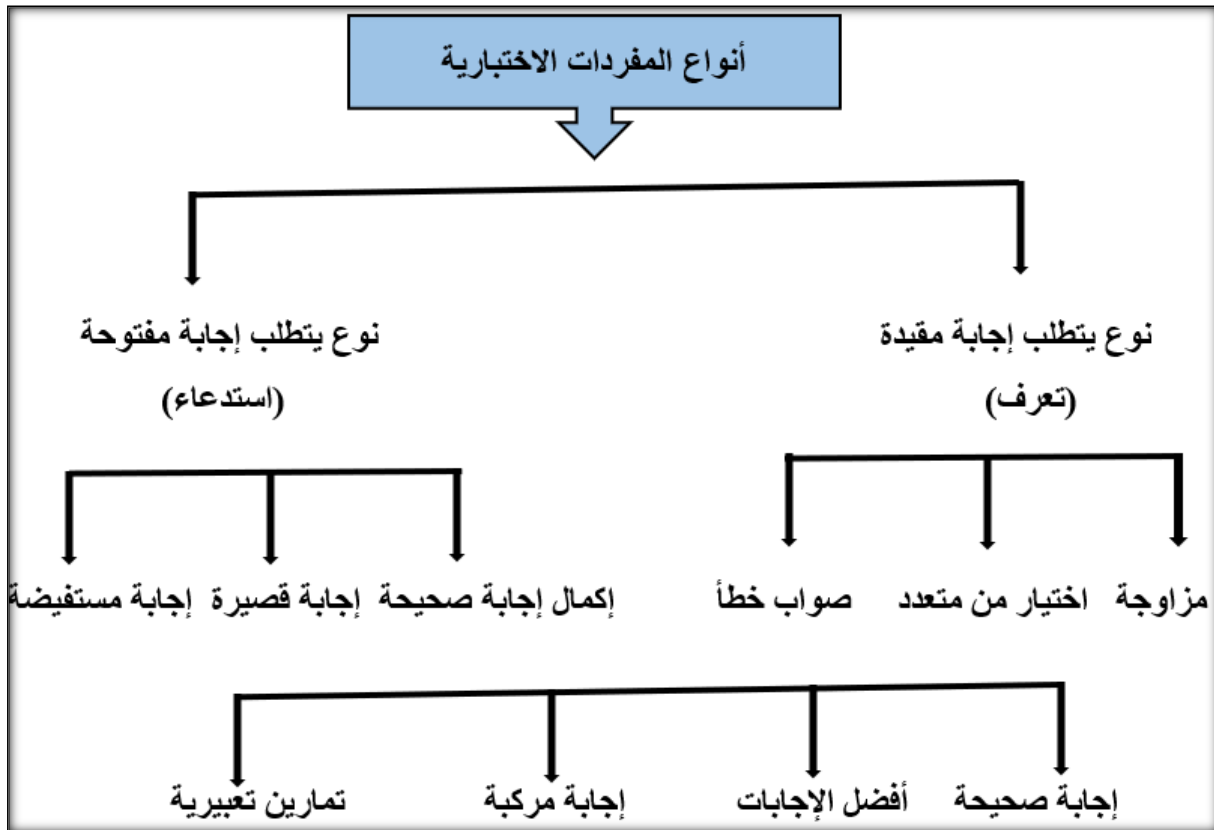
د- تقارب البدائل في الطول قدر الإمكان، لأن البدائل الصحيحة غالباً ما تكون أطول مما يجعلها مميزة ويسهل على التلميذ اكتشافها دون فهم السؤال.

هـ- أن يتراوح عدد البدائل بين (3-5) تبعاً لطبيعة المرحلة العمرية ومحتوى المادة الدراسية، إذ يساهم ذلك في تقليل احتمالية التخمين. فكلما زاد عدد البدائل تضاءلت فرص التخمين والعكس صحيح.

و- تجنّب استخدام المحددات الشائعة، مثل: جميع الإجابات صحيحة، جميع الإجابات خاطئة، لا شيء مما سبق، دائماً، أبداً... إلا في الحالات الضرورية؛ كأن يصعب توفير بدائل كافية أو عند الحاجة لقياس دقة عالية في تحديد مدى صحة الإجابة أو خطئها.

إن الالتزام بهذه الشروط عند صياغة البدائل لا يُسهم فقط في رفع جودة السؤال وموضوعيته، بل يُعزز أيضاً من دقة القياس ويُقلل من أثر العوامل العارضة مثل التخمين أو الاستدلال الشكلي مما يجعل الاختبار أكثر عدالة وفعالية (مراد وسليمان، 2005، ص. 197-203).

شكل (28) مخطط يوضح أنواع المفردات الاختبارية (هيئة التأطير، ص28).



3-2- الاختبار التحصيلية مرجعية المحك (CRT): Criterion-Referenced Tests

الاختبارات التحصيلية مرجعية المحك تعد من الأدوات التقييمية الدقيقة التي تهدف إلى قياس أداء الأفراد في ضوء مستوى التمكن المطلوب مسبقاً، أي أنها لا تكتفي بالمقارنة بين أداء الفرد وأداء المجموعة (كما هو الحال في الاختبارات مرجعية الجماعة) بل تركز على درجة إتقان الفرد لمجموعة محددة من الأهداف السلوكية المرتبطة بعينة من السلوك ضمن محتوى تعليمي معين. وبذلك فإنها تقدم صورة واضحة حول مدى تحقق نواتج التعلم المستهدفة لدى المتعلم مما يجعلها أداة مركزية في مجالات التربية والتدريب.

تتبع أهمية هذه الاختبارات من كونها تمكن المعلم أو المقيم من معرفة ما إذا كان المتعلم قد بلغ مستوى الأداء المتوقع ومن ثم تحديد نقاط القوة والضعف بدقة وهو ما يعزز من فاعلية العملية التعليمية. وقد أشار مارتن (Martin, 1975) إلى أن هذه الاختبارات تقوم على موضوعية الأداء، حيث يتم تصميمها بطريقة تسمح بالحكم بشكل قاطع على مدى إتقان الهدف من قبل المتعلم.

وترى ميمي السيد (2014، ص.9) أن الاختبارات مرجعية المحك تمثل وسيلة فعّالة للحصول على معلومات تفصيلية عن مستوى أداء الأفراد الأمر الذي يجعلها أكثر دقة وملاءمة في الحكم على نواتج التعلم مقارنة بالاختبارات التقليدية.

تُستخدم الاختبارات المحكية أو اختبارات الإتقان بشكل رئيسي في برامج التعليم الإفرادي أو في إطار نموذج إتقان التعلم (Bloom, 1968)، حيث تُبنى هذه البرامج من وحدات تعليمية أو قياسية غالباً ما تكون مرتبة بشكل هرمي ويُخصص لكل وحدة هدف سلوكي واحد أو أكثر. ويُطلب من التلميذ العمل على الوحدة التعليمية إلى أن يحقق مستوى معيناً من التحصيل الأدنى المطلوب وعندما يبلغ هذا المستوى يُعد التلميذ قد أتقن الوحدة التعليمية وسيطر على محتواها.

وفي هذه البرامج لا تُبنى القرارات التعليمية الخاصة بالتلميذ على مقارنة أدائه مع أقرانه بل تُستند فقط إلى مدى دقته في تحقيق الأهداف التعليمية المحددة. فإذا أظهر أداءً متقناً يُسمح له بالانتقال إلى الوحدة التالية، أما إذا لم يحقق المطلوب فعليه إعادة دراسة المادة التي شملها الاختبار حتى يصل إلى مستوى الأداء الدقيق الذي يعكس إتقانه للمحتوى.

كما يتطلب اختبار الإتقان من القائم على عملية القياس تحديد درجة فاصلة (قطع) تمثل الحد الأدنى للتمكن وذلك استناداً إلى أساس منطقي ومنهج علمي دقيق لاختيار هذه النقطة (النبهان، 2013، ص.459).

1- مميزات الاختبار التحصيلي مرجعي المحك:

يشير لوموس (Loumos, 1980) إلى أن الاختبارات التحصيلية مرجعية المحك تتميز بعدة خصائص تجعلها أداة فعّالة في قياس مدى تحقق الأهداف التعليمية، ومن أبرزها:

- إمداد بسجل موضوعي عن توافق التلميذ: حيث توفر هذه الاختبارات بيانات دقيقة يمكن الاستفادة منها في متابعة مستوى التلميذ وتقدير مدى توافقه مع متطلبات الأهداف المحددة.

- تركيز مباشر على أهداف الأداء: فهي تتطلب من التلميذ أن يظهر سلوكاً أو أداءً يتطابق مع ما هو منصوص عليه في الأهداف السلوكية، مما يعزز الصلة بين التعليم والتقويم.

- تصميم خاص ومحدد: فهي ليست اختبارات عامة، بل تصمم وفق معايير دقيقة لقياس هدف أو مجموعة أهداف بعينها، مما قد يجعلها في بعض الأحيان قصيرة جداً أو بسيطة أو شديدة التحديد تبعاً لطبيعة الهدف.

- توجيه القرارات التعليمية: تساعد نتائج هذه الاختبارات في اتخاذ قرارات مهمة كانتقال التلميذ إلى وحدة دراسية لاحقة أو بقاءه في نفس المرحلة حتى يتقن ما هو مطلوب منه.

- **تنوع في الشكل:** يمكن أن تُبنى هذه الاختبارات في صورة اختبارات موضوعية، مثل: الاختيار من متعدد أو اختبارات أداء عملية، بحسب طبيعة الهدف التعليمي.
- **تقديم معلومات شاملة:** فهي لا تقتصر على إفادة القائمين على الاختبار فقط بل تقدم كذلك للتلامذة تغذية راجعة مهمة حول مستوى إنجازهم مما يعزز وعيهم بمسار تعلمهم.

2- خطوات بناء الاختبارات محكية المرجع:

لكي يُصنّف اختبار ما ضمن فئة الاختبارات المرجعية إلى المحك لا بد من اتباع مجموعة من الخطوات المنهجية في بنائه وتقويمه، وقد أشار فؤاد أبو حطب وآخرون (2008، ص.495) إلى أهم هذه الخطوات على النحو الآتي:

- إعداد تخطيط للمحتوى:** يتم فيها تحديد المهارات والمعارف التي يسعى الاختبار إلى قياسها.
- تحديد الأداءات أو الأهداف السلوكية:** أي صياغة المرامي القابلة للقياس والتي يُفترض أن يكون الفرد قادراً على تحقيقها إذا امتلك الكفاءة المطلوبة في المهارات والمعارف المستهدفة.
- تحديد المجالات التفصيلية:** حيث يُعرّف المجال أو النطاق المرتبط بكل مرمى ثم تُكتب فقرات الاختبار وفقاً للتفاصيل المحددة في ذلك المجال مع اختيار عينات مناسبة، مثلاً: عنصرين لكل مرمى لبناء الاختبار.

هـ- التحقق من صدق العلاقة بين المهارات والمرامي: إذ ينبغي التأكد من أن المعارف والمهارات التي يقيسها الاختبار تمثل بالفعل متطلبات الأهداف الأدائية وتبدأ هذه المرحلة عادةً بحكم ذاتي (صدق ظاهري) ثم تمتد لتشمل آراء الخبراء أو بيانات تجريبية ناتجة عن تطبيق الاختبار على عينة تمتلك الكفاءة في مجال المهارات المستهدفة.

و- تحديد درجة القطع (Cutoff score): أو وضع درجة محك (Criterion score) التي يجب أن يحققها المتعلم ليُعد قد بلغ مستوى الكفاءة المطلوبة مما يبرهن على قدرته على أداء السلوك المحك بشكل مرضٍ.

3- أوجه الاختلاف بين الاختبارات التحصيلية مرجعية المحك ومرجعية المعيار:

يوجد عديد الاختلافات بين الاختبارات مرجعية المعيار والاختبارات مرجعية المحك والتي يمكن أن نلخصها في الجدول الآتي:

جدول رقم 23 يوضح المقارنة بين الاختبار مرجعي المعيار ومحكي المرجع.

وجه الاختلاف	الاختبار مرجعي المعيار	الاختبار مرجعي المحك
الهدف الرئيس للاختبار	مقارنة أداء التلميذ بأداء مجموعته المعيارية من خلال عدد الإجابات الصحيحة.	مقارنة أداء التلميذ بمحك محدد مسبقاً لمعرفة الأهداف التي تحققت والتي لم تتحقق.
استخدام الاختبار	إظهار الفروق الفردية بين التلامذة.	التأكد من تحقيق التلميذ للأهداف السلوكية المحددة.
خصائص أسئلة الاختبار	تنتشر الأسئلة على نطاق واسع من الأهداف.	تتركز الأسئلة حول عدد محدد من الأهداف.
تفسير الأداء	تفسير الدرجات بناءً على الأداء النسبي للزملاء (الدرجات المعيارية).	تفسير الدرجات بناءً على درجة قطع (Cutoff score) محددة مسبقاً.
بناء المفردات	يعتمد على تباين الدرجات ومعامل التمييز بين الأسئلة.	يعتمد على مدى تحقيق كل مفردة للهدف السلوكي المطلوب.

يوضح الجدول السابق الفروقات الأساسية بين الاختبارات التحصيلية مرجعية المعيار والاختبارات مرجعية المحك من عدة أبعاد بما في ذلك الهدف الرئيسي للاختبار، وطريقة استخدامه، وخصائص أسئلته، وكيفية تفسير أداء التلامذة، وكذلك طريقة بناء المفردات. يظهر من الجدول أن الاختبارات مرجعية المعيار تركز على مقارنة أداء التلميذ بأداء زملائه لتحديد الفروق الفردية، بينما تهتم الاختبارات مرجعية المحك بقياس مدى إتقان التلميذ لمجموعة محددة من الأهداف السلوكية مسبقاً. كما تختلف طريقة توزيع الأسئلة وتركيزها، حيث تنتشر أسئلة الاختبارات مرجعية المعيار على نطاق واسع من الأهداف، بينما تركز أسئلة الاختبارات مرجعية المحك على أهداف محددة. كذلك تختلف أساليب تفسير النتائج إذ تعتمد الاختبارات مرجعية المعيار على الأداء النسبي للزملاء، في حين تعتمد الاختبارات مرجعية المحك على درجة القطع أو الحد الأدنى للأداء المطلوب. وأخيراً يظهر الاختلاف في بناء المفردات، فبينما يعتمد الاختبار مرجعي المعيار على تباين الدرجات ومعامل التمييز يعتمد الاختبار مرجعي المحك على مدى تحقيق كل مفردة للهدف السلوكي مما يعكس تركيزه على قياس الكفاءة الحقيقية للمتعلم.

4- العدالة في التقييم التربوي:

تُعد العدالة (Fairness) مفهومًا محوريًا في ميدان القياس والتقويم التربوي وهي مفهوم تكتسب فيه التعريفات أهمية خاصة نظرًا لكونها كثيرًا ما تُفهم بصورة ضيقة ذات طابع تقني بحت. ومن ثم فإن النظر إلى العدالة ينبغي أن يتجاوز حدود البعد الإجرائي المرتبط ببناء الاختبار ليأخذ بعين الاعتبار السياقات الاجتماعية والثقافية التي تجري فيها عملية التقييم. فالعدالة لا تقتصر على ضمان تكافؤ الشروط التقنية للاختبار وإنما تمتد لتشمل كيفية إدراكها وتطبيقها عبر المجموعات والفئات المختلفة داخل المجتمع.

وفي هذا الإطار يتسع مفهوم التقييم التربوي ليشمل إلى جانب الاختبارات والامتحانات أحكام وتقويمات المعلمين لأداء الطلبة وهو ما يُعرف في المملكة المتحدة بمفهوم "التقييم". كما يتصل مفهوم العدالة بمسألة التحيز في القياس وعلاقته بالصدق ويتقاطع في الوقت ذاته مع المفهوم الأوسع للمساواة.

ويُتوقع أن تزداد الحاجة في القرن الحادي والعشرين إلى إعادة صياغة التقييمات التربوية على نحو يراعي الأبعاد الاجتماعية والثقافية للتعليم بعيدًا عن الاقتصار على الجانب الفني لبناء الأدوات. فالعدالة في التقييم ترتبط بما يسبق العملية التقييمية كفرص الوصول إلى التعليم والموارد المتاحة، وبما يعقبها كالتفسيرات المستخلصة من النتائج وآثارها العملية، فضلًا عن ارتباطها بمستوى تصميم التقييم ذاته (Gipps et al, 2009, p.105).

5- العدالة الاختبارية (Test Fairness):

للتأكد من عدالة الاختبار عبر المجموعات المختلفة كالذكور والإناث أو الطلبة الميسورين اجتماعيًا مقارنة بالمحرومين أو عبر الانتماءات العرقية، ينبغي النظر إلى ما هو أبعد من مجرد تساوي الدرجات أو النتائج. فالمعضلة الرئيسة تكمن في أن هذه المجموعات لا تتشابه في خصائصها وتجاربها مما يجعل مفهوم العدالة في التقييم غير قابل للاختزال في مخرجات متساوية. إن الفروق في الأداء قد ترجع إلى عوامل متعددة؛ منها تفاوت فرص الوصول إلى التعلم أو وجود تحيز في بنية الاختبار ذاته لصالح فئة معينة، وفي هذا السياق ميّز وود (1987) Wood بين بُعدين أساسيين للعدالة: الفرصة لاكتساب الموهبة وهو ما يرتبط بقضايا إتاحة فرص التعليم والتعلم؛ والفرصة لإبراز الموهبة بفعالية وهو ما يعكس العدالة في إجراءات التقييم ذاته.

وعليه فإن العدالة في التقييم لا يمكن أن تُفهم بمعزل عن فرص الوصول إلى المناهج التعليمية، ولا عن طبيعة الإمكانيات المتاحة أمام المتعلمين داخل بيئتهم المدرسية، فهاتان الركيزتان تمثلان

الأساس لبناء تصور أعمق وأكثر شمولاً لمفهوم العدالة في التقييم التربوي. وتشمل العدالة هنا ضمان تكافؤ الفرص في الوصول إلى التعليم المدرسي أي الحق في الدخول إلى النظام التعليمي؛ وإلى المنهاج الدراسي أي الحق في الاستفادة المتكافئة من محتواه وفرصه على حد سواء، بما يوفر "بيئة متكافئة" تُعد شرطاً مسبقاً لا غنى عنه لتحقيق مواقف تقويمية يمكن وصفها حقاً بالعدالة (Gipps et al, 2009, p.105).

6- العدالة والإنصاف (Fairness and Equity):

يُستخدم مفهومَا العدالة والإنصاف بشكل متقارب، حيث يعرف الإنصاف بأنه العدالة الأخلاقية وهو لا يعني بالضرورة المساواة في النتائج أو التجارب لأن ذلك غير واقعي بل يقتضي أن تكون ممارسات التقييم وتفسيرات نتائجه عادلة لجميع الفئات. فقد تتحقق نتائج مقاربة بين مجموعتين ومع ذلك يُعد الموقف غير عادل إذا كانت إحداها محرومة من فرص الوصول إلى المنهاج. وبالمقابل قد تكون هناك نتائج غير متساوية بين مجموعات مختلفة لكنها تُعتبر عادلة إذا كانت الفرص والموارد المتاحة متكافئة وكان الاختلاف راجعاً إلى الجهد والاستعداد للتعلم.

فالإنصاف بهذا المعنى لا يطابق المساواة؛ فالمساواة تُعنى بالقياس الكمي الفروق بين المجموعات، بينما ينظر الإنصاف إلى عدالة الظروف والترتيبات السابقة على التقييم والنتائج المترتبة عنه. وهو مفهوم ذو بعد اجتماعي وثقافي يتجاوز نظرية القياس وينسجم مع التوجهات السوسولوجية المستندة إلى أعمال فيجوتسكي (Vygotsky's) حيث يُنظر إلى التقييم كعملية اجتماعية وثقافية وسياسية لا يمكن فهمها بمعزل عن السياق الذي تجري فيه.

الجدول رقم 24 لمقارنة توضح الفرق بين المساواة والإنصاف والعدالة.

المفهوم	التعريف	التركيز	المثال التربوي
المساواة (Equality)	منح الجميع نفس الحقوق أو الموارد دون تمييز.	التشابه الكمي (إعطاء الشيء نفسه للجميع).	تزويد جميع الطلاب بعدد متساوٍ من الكتب المدرسية أو نفس مدة الامتحان.
الإنصاف (Equity)	تكليف الموارد أو الفرص بما يتلاءم مع احتياجات الأفراد المختلفة.	مراعاة الفروق الفردية والظروف.	تقديم دعم إضافي للطلاب ذوي الاحتياجات الخاصة أو للطلاب من بيئات محرومة.
العدالة (Fairness)	مفهوم أشمل يوازن بين المساواة والإنصاف، ويركز على عدالة الوصول والإجراءات والنتائج.	التوازن بين الحقوق والاحتياجات لتحقيق بيئة عادلة للجميع.	وضع اختبار يُبنى ويُطبق ويُفسر بطريقة لا تميز ضد أي فئة، مع مراعاة الظروف السابقة والآثار اللاحقة.

7- معاني العدالة في الاختبارات:

تعددت التعريفات المرتبطة بمفهوم العدالة في الاختبارات غير أن اتفاقاً عالمياً حول تعريف محدد لم يتحقق حتى الآن (الرابطة الأمريكية للبحوث التربوية، الجمعية الأمريكية لعلم النفس، والمجلس الوطني للقياس في التعليم، 2014، ص. 49). ورغم ذلك يمكن الإشارة إلى بعض الجوانب التي تحظى بقدر من الاتفاق النسبي إلى جانب مجالات عديدة ما تزال محل خلاف ونقاش. وفي هذا السياق سيتم عرض هذه التوجهات المختلفة مع إبراز الرؤية الأكثر فائدة وملاءمة بالنسبة لمطوري الاختبارات.

أ- الصدق (Validity):

يُعد الصدق المؤشر الأهم والأشمل للحكم على جودة أي اختبار. وقد عرّفه ميسك بأنه: "حكم تقييمي متكامل يوضح مدى دعم الأدلة التجريبية والأسس النظرية لكفاية وملاءمة الاستنتاجات والإجراءات المستندة إلى درجات الاختبار" (Messick, 1989, p.13). في المقابل ركّز "كين" في مناقشته للصدق على مدى معقولية الادعاءات المستخلصة حول المختبرين استناداً إلى درجاتهم، وعلى درجة ما تحظى به هذه الادعاءات من دعم منطقي وتجريبي (Kane, 2013).

ب- البناء (Construct):

يشير البناء إلى مجموعة المعارف أو المهارات أو السمات الأخرى ذات الصلة (KSAs) التي يُفترض أن يقيسها الاختبار. كالرشاقة، المهارات اليدوية الضرورية لطلاب طب الأسنان المبتدئين، القلق، المعرفة المقررة في دروس الكيمياء بالمرحلة الثانوية، الذكاء، الانطواء، التفكير المنطقي، القدرة الكمية، وفهم القراءة.

وتُسمى المعارف أو المهارات أو السمات التي تدخل ضمن نطاق البناء المراد قياسه بـ العناصر المرتبطة بالبناء (construct-relevant)، في حين تُعد أي عناصر لا ترتبط بالبناء من العناصر غير المرتبطة بالبناء (construct-irrelevant). فعلى سبيل المثال تُعتبر القدرة اللفظية عنصراً غير مرتبط بالبناء عند قياس التحصيل في مادة الرياضيات لكنها تُعد عنصراً مرتبطاً بالبناء عند قياس الفهم القرائي (Zieky, 2016, p.10).

ج- تباين الدرجات (Score Variance):

يشير تباين الدرجات إلى مقدار الاختلاف بين درجات مجموعة من الأفراد، فإذا حصل جميع المختبرين على نفس الدرجة تماماً فإن التباين يكون صفراً. أما كلما ابتعدت الدرجات عن المتوسط، زاد التباين. وتعود مصادر التباين إلى عوامل متعددة قد تؤثر على الفروق بين الدرجات، مثل:

- سمات لدى المختبرين: كالفهم الجيد أو الضعيف للقراءة.
 - خصائص الاختبار: مثل استخدام أسئلة مقالية مقابل أسئلة اختيار من متعدد.
 - العوامل البيئية: كأداء الاختبار في مكان هادئ أو في بيئة مليئة بالضوضاء.
- وقد يكون مصدر التباين مرتبطاً بالبناء المراد قياسه (construct-relevant) أو غير مرتبط به (construct-irrelevant) (Zieky, 2016, p.10).

د- التحيز (Bias):

يُستخدم مفهوم التحيز في بناء الاختبارات للدلالة على غياب العدالة بين الأفراد من خلفيات مختلفة رغم امتلاكهم نفس القدرة في المجال المقاس. ويُعد الاختبار متحيزاً إذا كان شخصان يتمتعان بالمستوى نفسه من القدرة، لكن احتمالية نجاحهما تختلف بسبب انتمائهما إلى مجموعات ثقافية متباينة (Shephard et al. 1987, p.317-375).

يظهر التحيز غالباً لأن الاختبارات تُصمم من منظور ثقافي محدد، بحيث تعكس خبرات تلك المجموعة وحدها، وينتج عن ذلك صعوبة أو إعاقة في أداء المتقدمين من ثقافات أخرى كما يحدث في بعض اختبارات الذكاء، وقد يكون التحيز راجعاً إلى:

- محتوى الاختبار: الذي يتضمن موضوعات مألوفة لمجموعة معينة دون غيرها.
 - غموض التعليمات: بما يؤدي إلى استجابات مختلفة باختلاف الخلفيات الثقافية.
 - أنظمة التقييم: التي قد تُهمّش استجابات صحيحة أو مناسبة لكنها أكثر شيوعاً لدى مجموعة معينة.
- ورغم أن الفروق بين المجموعات في درجات الاختبار تُعتبر مؤشراً على احتمال وجود تحيز، فإن ذلك لا يعني بالضرورة تحيزاً بالمعنى الفني للكلمة، إذ قد تكون هذه الفروق ناتجة عن عوامل بيئية أو عن تفاوت في فرص الوصول إلى المناهج الدراسية (Gipps et al., 2009, p.105-118).
- وبالتالي التحيز في الاختبارات لا يقتصر على وجود فروق بين المجموعات بل يتمثل في غياب تكافؤ الفرص في النجاح بين أفراد يملكون نفس القدرة نتيجة محتوى الاختبار أو طريقة تطبيقه أو تقييمه.

هـ الحيادية:

تعد الحيادية مبدأً أساسياً لتحقيق العدالة في الاختبار إذ تقتضي معاملة جميع المختبرين باحترام وتكافؤ الفرص للحصول على درجات صحيحة بغض النظر عن خلفياتهم أو انتماءاتهم. وقد يستلزم ذلك أحياناً إدخال تعديلات مناسبة تضمن عدم تأثر النتائج بخصائص لا علاقة لها بالبناء المقاس؛

فمثلاً يجب توفير مواد صوتية أو مكتوبة بطريقة برايل للمكفوفين عندما تكون القدرة على الرؤية غير مرتبطة بما يقيسه الاختبار (Zieky, M. J, 2016, p.10).

و- الفروق في الدرجات بين المجموعات:

لا تُعد الفروق في درجات الاختبار بين المجموعات دليلاً كافياً على انعدام العدالة، إذ يرى المتخصصون في القياس النفسي أن صعوبة الاختبار النسبية لمجموعة دون أخرى لا تعني بالضرورة تحيزه، بل إن الفروق تحتاج إلى تفسير أعمق، حيث يؤكد "كول وزيكى" أن "الاختلاف في الدرجات وحده لا يشكل دليلاً على التحيز" (Cole & Zieky, 2001).

4- العدالة القائمة على الصدق:

ترتبط العدالة في الاختبارات ارتباطاً وثيقاً بمدى صدق الاستنتاجات المستخلصة من الدرجات عبر مختلف المجموعات، كما تؤكد معايير ETS (2014). ويرى "شبيرد" أن التحيز هو "غياب الصدق" مما يجعل العلاقة بين الصدق والعدالة علاقة مباشرة؛ فتعزيز أحدهما يعزز الآخر وانخفاض أحدهما يعني بالضرورة انخفاض الآخر (Shepard, 1987).

5- الاختيارات من بين المحتوى الصادق:

رغم أن التباين الصادق يُعد عادلاً بطبيعته، إلا أن اختلاف المحتوى – حتى وإن كان يتمتع بالصدق ذاته – قد ينتج عنه فروق بين المجموعات (Willingham & Cole, 1997). وبما أن أي اختبار يمثل عينة من المجال المقاس فإن عملية تحديد محتواه غالباً ما تستند إلى قرارات تقديرية نظراً لصعوبة حصر جميع عناصر المجال أو اختيارها عشوائياً. وفي هذا السياق ينبغي مراعاة الفروق المحتملة بين المجموعات عند اختيار المحتوى ما دام ذلك لا يؤثر على الصدق. فعلى سبيل المثال أظهرت الأدلة أن النساء يميلن إلى أداء أقل من الرجال في مهام التصور المكاني (Voyer et al., 1995). لذا إذا توافرت مفردات تحقق أهداف الاختبار بكفاءة دون الاعتماد الكبير على هذه المهارة فإن تقليل التركيز على المهام المكانية يُعد خياراً أكثر إنصافاً في تصميم الاختبار.

6- التصميم المرتكز على الأدلة (Evidence-Centered Design - ECD):

يُعد التصميم المرتكز على الأدلة إطاراً منهجياً معتمداً في تطوير الاختبارات الحديثة يقوم على مبدأ التفكير المبني على الأدلة إذ يساعد هذا الإطار في تحديد ما ينبغي قياسه وكيفية قياسه بطريقة مناسبة مع توفير آليات واضحة لربط الادعاءات حول قدرات أو معارف المفحوصين بالأدلة التي يكشف عنها أدائهم في المهام المختلفة للاختبار.

ومن أبرز مرتكزاته الأساسية تعزيز الصدق والعدالة في الاختبارات من خلال تقليص التباين غير المرتبط بالبناء المستهدف، ويتطلب ذلك صياغة تعريف دقيق للبناء المراد قياسه وتحديد الفئة المستهدفة بدقة. ومن خلال هذا الضبط يتيح (ECD) تقليل مصادر الخطأ أو التحيز التي قد تؤثر على تفسير النتائج، كما يضمن هذا الإطار أن تكون عملية جمع الأدلة وتفسيرها متنسقة مع المعرفة الأساسية وأهداف التقييم، وهو ما أبرزه (Mislevy, Steinberg, & Almond, 1999). وبهذا يسهم (ECD) في تعزيز صدق النتائج، ويؤدي بالتالي إلى تحقيق العدالة في التقييم التربوي والنفسي.

7- التصميم الشامل (UD – Universal Design):

يمثل أخذ احتياجات الأفراد ذوي الإعاقات بعين الاعتبار أحد المحاور الرئيسية في عملية تصميم الاختبارات، لما لهذه الاحتياجات من أثر مباشر على القرارات المرتبطة بالبناء والقياس. ويسهم مفهوم التصميم الشامل (UD) في توجيه عملية التطوير نحو تعزيز العدالة والشمولية، بحيث تصبح الاختبارات أكثر ملاءمة لجميع المشاركين وأكثر قدرة على تحقيق أغراضها التقييمية (Thompson, Johnstone, & Thurlow, 2002).

ويتمثل الهدف الأساس لمصممي الاختبارات في قياس الكفاءات والمعارف المرتبطة بالبناء المستهدف مع تقليل أي تأثيرات جانبية ناتجة عن إعاقات لا ترتبط مباشرة بذلك البناء. وهنا يبرز دور مبادئ التصميم الشامل التي توفر وضوحاً منهجياً بشأن:

أ- الكفاءات والمعارف الضرورية لأداء الاختبار بشكل عام.

ب- الكفاءات والمعارف التي يسعى الاختبار إلى قياسها بصورة محددة.

فالمطلوبات التي لا ترتبط بالبناء المستهدف تمثل مصادر محتملة لتباين غير صادق وغير عادل. فعلى سبيل المثال إذا كان أداء الاختبار يتطلب قراءة نص قصير لا علاقة له بالبناء المستهدف فإن هذا النص يمثل عاملاً غير صادق للتباين. وإذا أثر هذا العامل سلباً وبشكل غير متكافئ على ذوي الإعاقات البصرية فإنه يصبح عاملاً مخلاً بالعدالة. وللتغلب على مثل هذه الحالات ينبغي تبني حلول عملية كتوظيف خطوط سهلة القراءة وتقليل الحواجز البصرية أو التقنية التي قد تحد من أداء بعض المشاركين. ويهدف التصميم الشامل للاختبارات إلى إزالة هذه المصادر غير ذات الصلة بالتباين عبر مجموعة من الإجراءات، من أبرزها:

- صياغة تعليمات الأداء بوضوح وإيجاز.
- تصميم صفحات الاختبار أو الشاشات الإلكترونية بحيث تكون سهلة القراءة والاستخدام.
- توفير أوصاف لفظية أو تمثيلات لمسية للخرائط والرسوم البيانية والمواد البصرية الأخرى.

- إتاحة تكييفات بصرية مثل تكبير حجم الخط أو زيادة التباين.

- تبسيط المفردات اللغوية بما يتناسب مع الغرض دون تعقيد غير ضروري.

- إشراك الأفراد ذوي الإعاقات في العينات القبلية أثناء مرحلة بناء الاختبار.

كما يُستحسن تجنب إدراج عناصر غير ضرورية التي قد تمثل عائقاً أمام بعض الفئات. فعلى سبيل المثال قد يستلزم سؤال رياضي استخدام رسم بياني إلا أنه لا حاجة لجعل هذا الرسم معقداً أو متعدد الألوان إذ قد يشكل ذلك عائقاً إضافياً للمصابين بعمى الألوان ويؤثر على عدالة الأداء (Zieky, M. J, 2016, p.14).

8- اللجنة الاستشارية للعدالة:

يُعَدُّ إشراك لجنة استشارية تُعنى بمسائل العدالة خطوة مهمة في مراحل تصميم الاختبارات وتطويرها اللاحقة. ورغم أنَّ مهامها قد لا تُعرَّف دائماً بهذا المسمى إلا أنَّ جوهر عمل اللجنة يتمثل في رصد وتحديد مصادر التباين غير المرتبط بالبناء المقاس والذي قد ينتج عن انتماء المفحوصين إلى مجموعات معينة. ولهذا يُطلب من مطوري الاختبارات تلبية متطلبات متعددة في آن واحد. ويساعد وجود لجنة استشارية للعدالة على ضمان مراعاة قضايا العدالة بشكل منهجي عند اتخاذ القرارات الخاصة بمحتوى الاختبار وأنماط الأسئلة المطروحة وأوقات إجراء الاختبار وأساليب التطبيق والاستجابة، وغيرها من الجوانب ذات الصلة (Zieky, 2016, p.15).

9- إرشادات كتابة الأسئلة والعدالة:

من المهم أن يحدد مطورو الاختبارات تصورهم الواضح لمفهوم العدالة في محتوى الأسئلة، وأن يتم تدريب كتّاب الأسئلة على صياغة بنود تتماشى مع هذا التصور قبل الشروع في عملية الكتابة، فإعداد الأسئلة يتطلب وقتاً وجهداً وتكلفة عالية ومن غير المنطقي صياغة أسئلة ثم استبعادها لاحقاً بسبب عدم توافقها مع معايير العدالة.

وضعت العديد من مؤسسات تطوير الاختبارات ودور النشر والوكالات الحكومية في الولايات المتحدة أدلة خاصة بإنتاج مواد تعليمية وتقويمية عادلة بما في ذلك الاختبارات وغالباً ما تحمل هذه الأدلة عناوين تتضمن مزيجاً من مصطلحات كالتحيز (Bias) والعدالة (Fairness) والحساسية (Sensitivity) مقترنةً بكلمات مثل "إرشادات" أو "مراجعة". ومن بين أهم هذه الإرشادات ما أصدرته الجمعية الأمريكية لعلم النفس (APA) (2010، ص. 71) من إرشادات عامة لتقليل التحيز. كما قدمت Data Recognition Corporation (2003) دليلاً يتناول قضايا العدالة والتحيز والحساسية في الاختبارات، إلى جانب إرشادات تدريبية ذات صلة. وما قدمته كذلك إرشادات خدمة

الاختبارات التعليمية ((Educational Testing Service (ETS -2015)). وبالمثل نشر (Smarter
(Balanced Assessment Consortium (2012)) إرشادات خاصة بمراجعة التحيز والحساسية.

10- أوجه التشابه بين الإرشادات:

إلى جانب التشابه في العناوين تميل الوثائق المختلفة إلى التوافق بدرجة كبيرة في الموضوعات التي تتناولها والاتجاه العام للملاحظات التي تقدمها. ورغم اختلاف مستويات التفصيل بين هذه الوثائق إلا أن الاختلافات الحقيقية نادرة وغالبًا ما تقتصر على تفاصيل بسيطة مثل ما إذا كان ينبغي كتابة عبارة "الأمريكيون الأفارقة" باستخدام الشرطة أم بدونها. في الواقع تتقاطع هذه الوثائق ليس فقط في المفاهيم التي تعرضها بل أيضًا في أسلوب الصياغة المستخدم للتعبير عنها (Ravitch, 2003). فالقواعد المتعلقة بالعدالة التي سيتم عرضها أدناه هي مقتطفات موجزة من أهم مضامين إرشادات خدمة الاختبارات التعليمية ((Educational Testing Service (ETS -2015) والتي سيشار إليها لاحقًا بـ(GFTC)). وقد استندت هذه القواعد الخاصة بالعدالة في صياغة الأسئلة إلى هذه الوثيقة التي تمثل بدورها العديد من النسخ الأخرى لإرشادات العدالة المتوفرة، وقد استخدمت الوثيقة على نطاق واسع في العديد من الاختبارات عالية الأهمية والتي صُممت لفئات متنوعة من المتقدمين منذ ثمانينيات القرن الماضي، حيث تم تعديلها وتوسيعها بشكل دوري لتواكب الآراء المتغيرة حول العدالة (Zieky, M. J, 2016, p.16).

تجدر الإشارة إلى أن قواعد العدالة الواردة في (GFTC) وضعت أساسًا للاستخدام داخل الولايات المتحدة. ومع أن القواعد العامة مثل: "التعامل مع المتقدمين للاختبار بعناية وتجنب أي تصرف قد يسيء إليهم" يمكن أن تنطبق في جميع السياقات إلا أن بعض القواعد التفصيلية قد تتطلب تعديلات لتناسب الدول الأخرى. فعلى سبيل المثال ما يُعتبر غير ملائم أو جارحًا عند الحديث عن "الأمريكيين الأفارقة" في السياق الأمريكي قد لا يكون ذا دلالة أو قد لا يُنظر إليه على أنه مسيء في ثقافات أخرى كما هو الحال بين باكستان واليابان. وللاطلاع على استراتيجيات التكيف لاستخدام هذه الإرشادات في سياقات دولية، يمكن الرجوع إلى (Educational Testing Service, 2009a).

11- الهدف من استخدام الإرشادات:

تتمثل الغاية الأساسية من تطبيق قواعد العدالة في الاختبارات في تجنب تضمين أسئلة أو محفزات تؤدي إلى تباين غير عادل في الدرجات ذلك التباين الذي لا يعكس حقيقة البناء المقاس. وتُوجّه هذه القواعد إلى مُعدّي الأسئلة ومنشئي الاختبارات عند صياغة المواد كما يعتمد عليها المراجعون للتأكد من انسجام المحتوى مع معايير العدالة.

وإلى جانب ذلك تسهم هذه الإرشادات في تحديد مفهوم العدالة في محتوى الاختبار مما يقلل من الاعتماد على القرارات الفردية أو المتباينة غالبًا بين المراجعين. ويُعد الهدف النهائي من هذه الإرشادات هو قياس البناء المستهدف بشكل مباشر دون التأثير بحواجز غير ذات صلة بالبناء، ومن أبرزها:

أ- **الحواجز المعرفية:** مثل قياس معارف ليست جزءًا من البناء المقاس.

ب- **الحواجز العاطفية:** مثل إثارة مشاعر قوية غير لازمة.

ج- **الحواجز الجسدية:** مثل فرض متطلبات حركية أو حسية لا ترتبط بالبناء المقاس داخل بيئة الاختبار.

إن أي محتوى يعد ضروريًا لقياس البناء المستهدف يعتبر محتوى عادلاً حتى وإن بدا غير ملائم وفقاً للإرشادات إذا كان مفتاحاً لتحقيق الصدق. فعلى سبيل المثال قد يُرفض تناول موضوع تلوث الطعام بالبكتيريا المعوية في اختبار لفهم القراءة لأنه مزعج ومسيء، لكنه في المقابل يكون مقبولاً وضرورياً في اختبار خاص بمراقبي المطاعم. وهنا يظهر أن الالتزام بالمواد الضرورية لصدق القياس لا يُعد شكلاً من أشكال الرقابة كما قد يُفهم خطأً بل ضماناً للعدالة في التقييم.

أما على مستوى القواعد التفصيلية التي تستهدف إزالة مصادر التباين غير المرتبط بالبناء فقد تختلف وجهات النظر حول بعض الأمثلة المحددة؛ كالسؤال حول ضرورة تجنب موضوع الرقص الاجتماعي، غير أن الهدف الرئيس يظل ثابتاً وواضحاً وهو قياس البناء المستهدف بعدالة لجميع الفئات (Zieky, 2016, p.17).

12- المصادر المعرفية للتباين غير المرتبط بالبناء:

تظهر المصادر المعرفية للتباين غير المرتبط بالبناء كعائق أمام العدالة في الاختبارات عندما يتحقق شرطان رئيسيان:

أ- أن تكون هناك معرفة أو مهارة لا يقصدها القياس، لكنها ضرورية للإجابة الصحيحة.

ب- أن تختلف المجموعات في مدى امتلاكها لتلك المعرفة أو المهارة.

في كثير من الأحيان تتضمن الأسئلة سياقات أو محتويات لا ترتبط مباشرة بالبناء المستهدف. فعلى سبيل المثال يقتضي اختبار الفهم القرائي عرض مقاطع نصية ذات موضوعات محددة، بينما الهدف هو قياس القدرة العامة على الفهم القرائي لا الفهم المتعلق بالمحتوى المحدد للنص. فإذا أدى هذا المحتوى إلى فروق بين المجموعات يصبح مصدرًا معرفيًا لتباين غير عادل في الدرجات.

وبالمثل في مسائل الرياضيات التطبيقية يشيع استخدام سياقات حياتية لتوضيح المسألة، مثل: تقسيم بيتزا بين مجموعة أطفال لقياس معرفة الكسور. هنا قد يصبح السياق نفسه رغم أنه غير ضروري للبناء مصدرًا للتباين غير العادل إذا تطلب معرفة تختلف مستوياتها بين المجموعات. ولتفادي ذلك لا بد من تجنب اللغة الصعبة غير الضرورية أو المطالبة بمعرفة متخصصة لا تمت بصلة للبناء المقاس. ومع ذلك تختلف الأمور في بعض الاختبارات كاختبارات الترخيص المهني، حيث قد تُقبل سياقات عملية تعكس خبرات متوقع أن يواجهها المتقدمون في بدايات عملهم. المشكلة الأساسية تكمن في التمييز بين:

أ- معرفة غير مرتبطة بالبناء لكن يمكن افتراض امتلاك جميع المتقدمين لها، فتُعد مقبولة.
ب- معرفة غير مرتبطة بالبناء وتؤثر بشكل غير متكافئ على بعض المجموعات، فتُعد غير عادلة.
أمثلة على ذلك:

- هل يتسبب نص عن الحياة في المزرعة في إضعاف فرص المتقدمين من المدن؟
- هل يُعد سؤال رياضي حول جدول مواعيد المترو غير عادل للمتقدمين من المناطق الريفية؟
- هل يعتبر تضمين نص عن سباقات الجري إشكالاً بالنسبة للمتقدمين ذوي إعاقات تمنعهم من المشاركة فيها؟

عملياً من النادر العثور على محتوى يمتلك جميع المتقدمين معرفة متساوية به. في الاختبارات الأكاديمية المدرسية (K-12) يمكن أن يمثل المنهاج الدراسي النموذجي مرجعاً مناسباً إذ إن المدارس الأمريكية مقارنةً بالبيئات المنزلية أكثر تقارباً في الخبرات التي تقدمها. فعلى سبيل المثال من المرجح أن يكون معظم طلاب الصف الثالث قد سمعوا بمفهوم القوارب حتى وإن لم يركبوا واحداً من قبل. أما على مستوى السياقات الفريدة لبعض الدول مثل الولايات المتحدة فقد يُشكّل إدراجها تحدياً إضافياً. لذا يُعد من الأنسب تجنب استخدام محتويات محلية خاصة ما لم تكن جزءاً جوهرياً من البناء المقاس. فعلى سبيل المثال لا ينبغي إدراج مسائل رياضية تتطلب معرفة بالعملات الأمريكية إلا إذا كانت هذه المعرفة جزءاً صريحاً من البناء المقاس (Zieky, 2016, p.18).

13- المبادئ العامة للعدالة الاختبارية:

لا توجد مجموعة من القواعد قادرة على تغطية جميع الحالات المحتملة ومن الطبيعي أن تبرز بعض الاختلافات بشأن تطبيق بعض القواعد. كما أن هذه القواعد قد تصبح غير مناسبة مع مرور الوقت. فعلى سبيل المثال كان استخدام كلمة "الرجل" للإشارة إلى جميع البشر أمراً شائعاً ومقبولاً

في فترة ماضية، لكنه لم يعد كذلك اليوم. لذلك عند ظهور تساؤلات حول عدالة أحد البنود من المفيد الرجوع إلى بعض المبادئ العامة للعدالة التي تظل عالمية وثابتة، وهي:

- تضمين جميع العناصر الضرورية لتحقيق قياس صحيح.
- إظهار الاحترام الكامل لجميع المتقدمين للاختبار.
- توفير فرص متساوية لجميع المجموعات للحصول على نتائج قابلة للتفسير بشكل سليم.
- تجنب إدخال مواد لا ترتبط بالبناء المقاس والتي قد تجعل الأفراد يشعرون بأن الاختبار غير عادل أو غير مناسب (Zieky, M. J., 2016, p.24).

14- تدريب كتاب البنود والمراجعين:

إن وجود القواعد الخاصة بالعدالة في صياغة البنود وحده لا يكفي بل يجب أن يتعلم كتاب البنود والمراجعون كيفية تطبيق هذه الإرشادات بشكل صحيح. ويتطلب التدريب الفعال المرور بعدة مراحل، حيث يُعد من المفيد تدريب الكتاب والمراجعين معًا لكون كل كاتب بند قد يكون مراجعًا محتملاً بينما ليس كل المراجعين بالضرورة كتاب بنود.

تبدأ المرحلة الأولى من التدريب بمناقشة معاني القواعد المرتبطة بالعدالة مع تقديم أمثلة واضحة على الانتهاكات المحتملة. أما المرحلة الثانية فتركز على مناقشة المواد الحدودية التي قد يراها بعضهم مقبولة بينما يعتبرها آخرون مخالفة للإرشادات. هذه المناقشات قد تؤدي إلى توافق أو تكشف عن قواعد غامضة أو مثيرة للجدل تحتاج إلى توضيح أو مراجعة. وبعد أن يطبق الكتاب والمراجعون هذه القواعد لعدة أشهر، تأتي المرحلة الثالثة التي تنسم بالاستمرارية، حيث يجتمعون بشكل دوري لمناقشة تجاربهم وتبادل الحلول للمشكلات الشائعة وتصحيح أي سوء فهم. وتجدر الإشارة إلى أن وجهات النظر حول العدالة قد تتغير بمرور الزمن؛ فعلى سبيل المثال كان يُنظر إلى مصطلح "نيغرو" في الماضي على أنه مقبول للإشارة إلى الأمريكيين الأفارقة، لكنه اليوم يُعتبر غير لائق. لذا يصبح من الضروري مراجعة القواعد وتحديثها بشكل مستمر كلما دعت الحاجة (Zieky, M. J., 2016, p.24).

8- الأداء التفاضلي للمفردة والعدالة الاختبارية:

يُعرّف الأداء التفاضلي للمفردة (Differential Item Functioning - DIF) بأنه الحالة التي يظهر فيها اختلاف في أداء مجموعات مختلفة من المتقدمين للاختبار على مفردة معينة، رغم أنهم قد تمت مطابقتهم وفقًا لمعيار محدد غالبًا ما يكون درجة الاختبار الكلية. فعلى سبيل المثال تتم مطابقة الرجال والنساء الذين حصلوا على نفس الدرجة (X) أو الدرجة السابقة لها (X-1) على افتراض أن لديهم

مستوى متقاربًا من المعارف والقدرات التي يقيسها الاختبار. يتم تجميع الفروق في الأداء عبر مستويات الدرجات المختلفة، وكلما زادت هذه الفروق ارتفعت القيمة المطلقة لـ(DIF) وغالبًا ما تُعطى قيم سلبية عندما تجد المجموعة المستهدفة (Focal Group) كالنساء أو الأمريكيون الآسيويون أو ذوو البشرة السوداء أو الأمريكيون الأصليون فالفردة أكثر صعوبة مقارنةً بـ المجموعة المرجعية (Referens Group) كالذكور أو ذوو البشرة البيضاء، ذات القدرة المماثلة في الاختبار.

يساعد فحص أنماط الأداء التفاضلي للمفردات (DIF) على تعزيز العدالة في الاختبارات. فعلى سبيل المثال أظهرت بعض الدراسات أن مفردات الاختيار من متعدد التي تُعرض خياراتها في صف أفقي تكون أصعب نسبيًا لبعض المجموعات المستهدفة مقارنةً بالمفردات التي تُعرض خياراتها في شكل عمودي (O'Neill & McPeck, 1993). وبما أن هذا الفارق لا يرتبط بالبناء المقاس فإن التنسيق الأفقي يُعد مصدرًا غير عادل للتباين. ومع ذلك فإن وجود أداء تفاضلي للمفردة (DIF) لا يُعد دليلًا قاطعًا على عدم العدالة؛ فقد تكون الفروق ناتجة عن البناء ذاته. فمثلًا المفردات التي تختبر الكلمات ذات الجذور اللاتينية قد تكون أسهل نسبيًا للمتقدمين من أصول إسبانية مقارنةً بغيرهم (Schmitt, Holland, & Dorans, 1993). وفي هذه الحالة لا يُعتبر وجود هذا النوع من المفردات غير عادل، لأنه يرتبط مباشرة بالبناء المراد قياسه.

كما أن ظهور (DIF) قد يكون مرتبطًا بسياق اختبار معين دون آخر فمفردة في الجبر قد تُظهر أداءً تفاضليًا في اختبار رياضيات يحتوي على عدد محدود من مفردات الجبر بينما قد يختفي ذلك الأثر في اختبار يضم عددًا كبيرًا منها، نظرًا لأن وفرة المفردات تعزز مطابقة المعرفة بين المجموعات.

ولتطبيق الأداء التفاضلي للمفردة (DIF) بصورة منهجية وُضعت مجموعة من الإجراءات الأساسية، من أبرزها:

أ- تصنيف المفردات إلى ثلاث فئات وفقًا للقيمة المطلقة لـ(DIF): (A) صغيرة أو غير مهمة، (B) معتدلة، (C) كبيرة.

ب- بناء الاختبار يعتمد بشكل أساسي على مفردات المجموعة (A).

ج- عند الحاجة إلى تلبية متطلبات ومواصفات معينة يمكن استخدام بعض مفردات المجموعة (B).

د- لا يُسمح باستخدام مفردات المجموعة (C) إلا إذا كان من الضروري تلبية المواصفات، وبعد أن تؤكد لجنة مستقلة عدالة هذه المفردات.

هـ- في حال تجميع اختبار تشغيلي (Operational test) دون بيانات (DIF)، يجب حساب (DIF) بعد التطبيق وقبل إصدار الدرجات.

و- تخضع أي مفردة من المجموعة (C) لمراجعة لجنة مستقلة، ولا تُعتمد إلا إذا تأكدت اللجنة من عدالتها (Zieky, 1993, 2011).

الفصل التاسع

ملفات أعمال الطالب بورتفوليو (Portfolio)

الأهداف (Objectives):

يهدف هذا الفصل إلى التعريف بالتقييم التوثيقي الشامل والكفايات من خلال التركيز على ملف الأعمال (Portfolio) باعتباره أحد الأساليب الحديثة في تقييم أداء المتعلمين. ويسعى إلى توضيح مفهوم ملف الإنجاز واستخداماته التربوية، مع استعراض مكوناته المختلفة وأنواعه المتعددة، فضلاً عن تحديد أهدافه والمكونات الأساسية التي ينبغي أن يتضمنها. كما يتناول الفصل خصائص ملف الإنجاز التي تميزه عن غيره من أساليب التقييم، والخطوات المنهجية لتصميمه وتوظيفه في الميدان التربوي، إضافة إلى طرق تصحيحه وأساليب تقويمه. ولا يغفل الفصل التطرق إلى ملف الإنجاز الإلكتروني (E-Portfolio) بوصفه امتداداً رقمياً للملف التقليدي، مع مناقشة بعض المشكلات التي قد تواجه عملية تقويم التحصيل الدراسي باستخدام الصحائف الوثائقية. ومن خلال ذلك يهدف الفصل إلى إكساب الطالب معرفة شاملة بالجوانب النظرية والعملية لملف الإنجاز كأداة تقييم بديلة وفاعلة.

المحتوى (Contents):

- 1- التقييم التوثيقي الشامل والكفايات Portfolio Assessment
- 2- مفهوم واستخدامات ملف الأعمال (Portfolio)
- 3- مكونات الصحائف الوثائقية أو ملفات الإنجاز (Portfolios)
- 4- أنواع ملفات أعمال الطالب (ملفات الإنجاز)
- 5- أهداف ملف الإنجاز
- 6- المكونات الأساسية لملف الإنجاز
- 7- خصائص ملف الإنجاز
- 8- خطوات تصميم واستخدام ملف الإنجاز
- 9- تصحيح ملف الإنجاز
- 10- تقييم ملف الإنجاز
- 11- ملف الإنجاز الإلكتروني (E-Portfolio)
- 12- بعض مشكلات تقييم التحصيل باستخدام الصحائف الوثائقية (ملفات الأعمال)

تمهيد:

شهدت الممارسات التربوية في العقود الأخيرة تحولاً ملحوظاً من الاعتماد شبه الكلي على الاختبارات التقليدية إلى تبني أساليب بديلة في التقييم والتي تهدف إلى قياس تعلم الطالب بصورة أكثر شمولاً وواقعية. ويأتي التقييم التوثيقي أو ملف الإنجاز (Portfolio) في مقدمة هذه الأساليب إذ يُعد أداة غنية ومرنة تسمح بتجميع أعمال الطالب وأنشطته المختلفة على نحو يعكس تقدمه وتطوره عبر الزمن. ولا يقتصر ملف الإنجاز على كونه وسيلة لجمع الأدلة والشواهد، بل يمثل أيضاً أداة للتأمل الذاتي وتعزيز المسؤولية الفردية لدى المتعلم. وفي هذا الإطار يتناول هذا الفصل المفاهيم الأساسية لملف الإنجاز واستخداماته، ومكوناته، وأنواعه، وأهدافه، إضافة إلى خصائصه وخطوات تصميمه وتوظيفه التربوي. كما يسلط الضوء على آليات تصحيح وتقييم ملفات الإنجاز مع التطرق إلى الملف الإلكتروني (E-Portfolio) بوصفه نقلة نوعية في مجال التقييم، فضلاً عن مناقشة أبرز المشكلات التي قد تعترض استخدام هذه الأداة في قياس التحصيل. ومن خلال ذلك يسعى الفصل إلى إبراز أهمية ملف الإنجاز كأحدى الممارسات التقييمية الحديثة التي تواكب متطلبات التعليم المعاصر.

1- التقييم التوثيقي الشامل والكفايات (Portfolio Assessment):

يُعد هذا الاتجاه في تقييم تحصيل المتعلمين وكفاياتهم من أبرز الاتجاهات الحديثة التي ظهرت في الميدان التربوي في العقود الأخيرة، وذلك استجابةً لمحدودية الأساليب التقليدية في القياس والتقييم. إذ يرى أنصار هذا التوجه أن الاختبارات السائدة كالاختبارات الموضوعية القائمة على مفردات الاختيار من متعدد أو الأسئلة المغلقة غالباً ما تُقدّم صورة مجتزأة عن قدرات الطالب، وتركز بشكل أساسي على استدعاء الحقائق أو التعامل مع مواقف مصطنعة لا تعكس بالضرورة واقع الأداء الفعلي.

ومن هنا جاء التقييم التوثيقي الشامل كمدخل بديل يندرج ضمن ما يُعرف بـ التقييم البديل (Alternative Assessment) أو التقييم الواقعي (Authentic Assessment)، حيث يسعى إلى تقديم صورة أكثر شمولية وواقعية عن أداء المتعلم. فهو يركز على تجميع شواهد وأدلة متنوعة توثّق مسيرة الطالب التعليمية ونموه المعرفي والمهاري والوجداني، ويمنح المتعلم دوراً فاعلاً في عملية التقييم عبر إشراكه في انتقاء الأعمال والأنشطة التي تُبرز تقدمه وتطوره عبر الزمن (علام، 2000، ص746).

وبذلك يُعد هذا النوع من التقييم أداة فعّالة لا تقتصر على قياس التحصيل الأكاديمي فحسب، بل تتجاوز به إلى الكشف عن الكفايات الحقيقية للطلاب وتعزيز مهارات التفكير الناقد والقدرة على حل المشكلات والابتكار، مما يجعله أكثر ملاءمة لمتطلبات التعلم في القرن الحادي والعشرين.

يهدف هذا النوع من التقييم إلى قياس القدرات العقلية العليا لدى المتعلمين مع التركيز على تتبع عمليات التعلم الأساسية وتطورها سواء في إطار المدرسة أو خارجها. ويعتمد في ذلك على صحيفة وثائقية (Portfolio) تجمع أعمال الطالب وأنشطته، بحيث تُقدّم صورة متكاملة عن تطوره ونموه عبر الزمن. ويرى بعض التربويين أن هذه الصحيفة ينبغي أن تتضمن جميع أنشطة الطالب وأعماله المنجزة خلال العام الدراسي لتكون بمثابة سجل شامل لمسيرته التعليمية. بينما يرى آخرون أن الهدف يكمن في اختيار عينة ممثلة من الأعمال تعكس مستوى إنجاز الطالب بصورة واقعية بحيث يمكن عرضها على ولي الأمر للتعرف على نوعية ما حققه ابنه من تطور. وهناك من يذهب إلى أن هذه الصحيفة يمكن أن تُبنى لغرض محدد كأن تشتمل على مجموعة من كتابات الطالب أو نماذج من محاولاته في حل المشكلات أو مختارات من أفضل أعماله بهدف عرضها على أقرانه أو أسرته أو حتى أفراد المجتمع المحلي.

ويبرز هذا الاستخدام الطابع العملي للتقويم التوثيقي؛ إذ يُشبه إلى حد ما يقوم به الفنانون والمصورون حين يعرضون عينات منتقاة من أعمالهم لتجسيد خبراتهم وإمكاناتهم. غير أن عينات أعمال الطلاب تختلف من حيث الغاية والمحتوى؛ فهي أوسع تنوعاً وأكثر ارتباطاً بالمجال الدراسي أو بالمهارات المستهدفة. ولهذا فقد شاع استخدام الـ Portfolio على نحو خاص في ميادين تعليم المهارات اللغوية – مثل القراءة والكتابة – والمهارات الفنية الإبداعية، كما استُخدم كذلك في الرياضيات والعلوم الإنسانية بوصفه أداة فعّالة لتقويم الأداء والتقدم المستمر (Kumar, et al, 2016, p.38).

أما التقييم التوثيقي الشامل (Portfolio Assessment) فيهدف بصورة أساسية إلى التحقق من جودة ونوعية التحصيل لدى المتعلم، وذلك من خلال تقصي مدى اكتسابه لمعارف ومهارات وقدرات واتجاهات عقلية محددة يتم التعبير عنها بصياغات تصف مستويات مختلفة من التعلم والاكتساب. ويُنفَّذ هذا التقييم عبر فحص عينات منتقاة من أعمال الطالب يقوم باختيارها بنفسه تحت إشراف وتوجيه معلمه ثم تُعرض هذه الأعمال على مقوّم خارجي (Audience) يختلف عن المعلم المباشر بما يضمن موضوعية الحكم وحياديته. ويسترشد هذا المقوّم في عمله بمحكّات متفق عليها تكون وثيقة الصلة بواقع المجتمع واحتياجاته.

تركز عملية التقييم في الغالب على الصحائف الوثائقية للطالب (Portfolio) من حيث شكلها ومحتواها إذ تمثل هذه الصحيفة مرآة لمسيرة الطالب التعليمية. غير أنَّ أهميتها لا تقتصر على كونها أداة لتقويم التحصيل فقط بل تتجاوز ذلك إلى كونها وسيلة تسهم في إعادة تشكيل الفكر التربوي حول عمليتي التعليم والتعلم. فهي تدفع المربين إلى تبني منظور أعمق للتدريس قائم على مراعاة الفروق الفردية وتتبع تطور المتعلمين وإبراز إنجازاتهم الفعلية بصورة أكثر واقعية وارتباطاً بالحياة اليومية (علام، 2000، ص746).

2- مفهوم واستخدامات ملف الأعمال (Protfolio):

يُعدّ ملف الأعمال واحدًا من أبرز أساليب التقويم المعتمد على الأداء وأكثرها شيوعًا في الممارسات التربوية الحديثة. وقد قُدِّمت له عدة تعريفات؛ حيث عرّفته مابري (Mabry, 1999) بأنه عبارة عن محفظة أو ملف يجمع فيه الطالب مجموعة منتقاة من أعماله بهدف تقديم صورة عامة عن مستوى تحصيله، بينما وصفه آرتر وسباند (Arter & Spande, 1999) بأنه انتقاء هادف لأعمال الطالب يُجسّد جهوده وتقدمه ومستوى تحصيله في مجال دراسي محدد يتم اختياره بمشاركته الفاعلة وفق قواعد ومعايير محددة تحكم عملية الانتقاء وتقويم جودة الأداء.

ويُظهر هذا التعريف أن ملف الأعمال لا يُعنى بمجرد تجميع أعمال الطالب بشكل عشوائي، بل يقوم على عملية انتقاء منظمة تُبرز مدى تطور الطالب وتقدمه وتُعزز دوره الإيجابي في بناء تجربته التعليمية. كما أن وجود معايير تقويم واضحة يضيفي على هذا الملف قيمة موضوعية تجعله أداة معتمدة في الحكم على جودة التحصيل والأداء.

أما من حيث الاستخدامات فقد استُخدم ملف الأعمال في البداية مع بعض المواد الدراسية ثم توسع ليشمل مختلف المجالات الأكاديمية وأثبتت فعاليته خاصة عند ربطه بالاتجاهات الحديثة في التدريس القائمة على النظرية البنائية التي تشجع المتعلم على المشاركة الفاعلة في بناء معرفته. ويكتسب ملف الأعمال أهمية إضافية من ميل الطالب الطبيعي إلى الاحتفاظ بإنجازاته والعودة إليها لاحقًا للتأمل فيها الأمر الذي يساعده على إدراك تطوره الذاتي والتخطيط لتحسين أدائه المستقبلي (Sweet, 1993) (السعداوي، 2018، ص173).

كما يُطلق على ملف إنجاز المتعلم أسماء متعددة، مثل: الحقيبة التعليمية أو البورتفوليو، ويُنظر إليه بوصفه مؤشرًا حقيقيًا على مستوى المتعلم واتجاهاته وميوله، شريطة أن يُوظف بطريقة منهجية صحيحة من خلال معلم يمتلك مهارات عالية تتسم بالمرونة والانفتاح على الإبداع وقبول الاختلاف.

ويُعد هذا الملف أداة تربوية فعّالة تُسهم في تحقيق مجموعة من الأهداف التربوية المهمة، من أبرزها:

أ- تنمية مهارات التفكير الناقد والإبداعي، إضافةً إلى تعزيز مهارات التنظيم والاستقلالية في التعلم لدى الطالب.

ب- دمج المتعلمين في عملية تقويم تقدمهم الذاتي، حيث تُمكن عينات الكتابة أو الأعمال المجمعة داخل الملف من توثيق التغيرات التي تحدث على امتداد فصل دراسي كامل. وهذا التوثيق يصبح أداة مساعدة للمتعلمين لملاحظة تطورهم وتقييم أدائهم بأنفسهم واتخاذ خطوات لتحسينه.

ج- تدريب المتعلمين على ممارسة التأمل الذاتي فيما ينجزونه من أعمال وذلك من خلال تسجيل تعليقاتهم الشخصية حول ما يتضمنه الملف من أعمال والإجابة عن تساؤلات عاكسة، مثل:

- ماذا تعلمت من هذا العمل؟

- ما شعوري تجاه هذا العمل؟

- كيف يمكنني تطوير هذا العمل؟

- ما جوانب القوة أو التميز في أدائي؟

إن هذا النوع من الممارسات لا يعزز فقط وعي الطالب بذاته وبقدراته، بل يُسهم كذلك في بناء شخصية متعلمة ناقدة، متجددة، وقادرة على التطوير المستمر (الفرطوسي والميداني، 2021، ص330).

كما استُخدم ملف الأعمال (Portfolio) على نطاق واسع في التقويم النهائي لا سيما لأغراض الترفيع من صف دراسي إلى آخر، حيث أصبح مكوناً أساسياً في برامج التقويم التي اعتمدتها العديد من المؤسسات التربوية في بعض الولايات الأمريكية كفيرمونت (Vermont) وكنتاكي (Kentucky) ونيوجيرسي (New Jersey). وقد شكّل هذا التوجه نقلة نوعية في أنماط التقويم إذ تجاوز دوره كأداة للتقويم التكويني ليصبح جزءاً محورياً من عملية اتخاذ القرارات التربوية المصيرية.

ولم يقتصر استخدام ملف الأعمال على الولايات المتحدة فحسب، بل جرى توظيفه بشكل تكاملي مع استراتيجيات أخرى للتقويم الأدائي في معظم الدول وخاصة المتقدمة منها ككندا وبريطانيا وأستراليا. فعلى سبيل المثال أشار مكتب التعليم لشمال أمريكا (North America Division Office of Education [NADOE], 2000) إلى استخدام ملف الأعمال في عدد من المجالات التعليمية المختلفة، مما يعكس اتساع نطاقه ومرونته في التوظيف العملي (السعداوي، 2018، ص173).

تُعد المعارف تراكمات منهجية وهادفة وذات معنى من نتائج الأفراد، جُمعت وصُنفت عبر الزمن لتشكل أساساً يُستفاد منه في مجالات متنوعة. وقد استُخدمت فكرة توثيق المعارف منذ عصور بعيدة من قبل الفنانين والمصورين والمبدعين في مجالات متعددة مثل تصميم الأزياء والفنون التشكيلية، حيث اعتمدوا على ملفات إنجاز شخصية لعرض أفضل أعمالهم والتعريف بمستوى قدراتهم وإمكاناتهم، ومع تطور الممارسات التربوية أصبح ملف الإنجاز (Portfolio) أداة تعليمية وتقييمية تساعد على تتبع تطور الطالب وتقدمه عبر الزمن وإبراز مخرجات تعلمه بصورة منظمة وهادفة.

في السياق التربوي قد تتشابه العناصر المضمّنة في ملف الإنجاز مع منتجات المشاريع الطلابية، إلا أنها تختلف من حيث كونها ناتجة عن عملية اختيار واعٍ ومقصود على فترة زمنية أطول ما يجعلها أكثر دلالة على النمو الأكاديمي والمعرفي. كما يمكن أن يتضمن ملف الإنجاز ملاحظات أو تغذية راجعة من المعلمين، أولياء الأمور، الأقران، أو حتى الإداريين، ليعكس صورة شاملة عن تجربة الطالب التعليمية. ورغم عدم وجود صيغة موحدة أو إرشادات نهائية لشكل ملف الإنجاز إلا أن التربويين غالباً ما يميزون بين نوعين رئيسيين:

1- ملف العرض (Showcase or Model Portfolio):

يركز على تقديم نماذج مختارة بعناية من أعمال الطالب تمثل أفضل إنجازاته وأكثرها جودة. ويستلزم هذا النوع من الملف امتلاك الطالب مهارة التقويم الذاتي والقدرة على التمييز بين الأعمال المتنوعة لتحديد ما يعكس تفوقه وتميزه.

2- الملف الوصفي (Descriptive Portfolio):

يهدف إلى تقديم صورة شاملة عن خبرات الطالب، حيث يتضمن أعمالاً تمثيلية دون الاختصار على المنتجات النهائية عالية الجودة. قد يحتوي أيضاً على المسودات، سجلات الإنجاز، خطط العمل، والمراحل المختلفة للتقدم، مما يجعله أداة مفيدة لفهم عملية التعلم وليس فقط مخرجاتها النهائية. غالباً ما يُحدد المعلم نوع ملف الإنجاز وأهدافه ومكوناته العامة، مع ضرورة وضع حدود لعدد الأعمال المضمّنة لتفادي تحوله إلى مجموعة عشوائية ومرهقة يصعب تقييمها، وللحفاظ على قيمته التربوية. ومن هنا تأتي أهمية أن يمارس الطالب الحكم الناقد في اختيار محتوياته بما يتوافق مع متطلبات المهمة ويبرز مستوى إنجازه (Morrow, et al, 2000, p.167).

3- استخدامات ملف الإنجاز (Portfolio):

تتسم المشاريع والأنشطة التي يمكن أن يتضمنها ملف الإنجاز بتنوع واسع يكاد يكون غير محدود، حيث تختلف تبعاً لطبيعة التخصص ومجالات الممارسة. ويشير كيرك (1997) إلى قائمة ثرية من المهام والمشاريع التطبيقية في ميدان النشاط البدني يمكن دمجها ضمن ملف الإنجاز وهو ما يبرز مرونته ويتيح آفاقاً مبتكرة وواسعة لتوظيفه كأداة تعليمية فعالة (Morrow et al., 2000, p.167).

ومع اتساع نطاق استخدام ملف الأعمال في العملية التعليمية أصبح يؤدي أدواراً متعددة تتصل بجوانب التدريس والتعلم والتواصل، ويمكن تلخيص هذه الأدوار فيما يلي:

أولاً: في مجال التدريس:

- إتاحة فرصة لتوضيح مدى تقدم الطلاب عبر فترة زمنية محددة، بما يساعد على تتبع نموهم الأكاديمي.
- تزويد المعلم برؤية واضحة وموضوعية حول مستوى تدريسه، بما يمكنه من تحسين استراتيجياته التعليمية.
- مساعدة المعلمين على تقويم مدى اقتربهم من تحقيق أهداف المقرر الدراسي.
- إيجاد حلقة تواصل فعالة بين البيت والمدرسة من خلال مشاركة أولياء الأمور بملفات إنجاز أبنائهم.
- توفير وسيلة منظمة لرصد شواهد التحصيل بهدف الترفيع من صف إلى آخر.

ثانياً: في مجال التعلم وتعزيز التحصيل الدراسي:

- تيسير عملية التعلم على الطلاب وتحفيزهم نحو المثابرة والاستقلالية.
- توعية الطلاب بأدوارهم في العملية التعليمية وتحميلهم مسؤولية تعلمهم الذاتي.
- تنمية مهارة التفكير النقدي لديهم عبر التأمل في أعمالهم ومراجعتها.
- تعزيز التقدير الذاتي من خلال عرض أفضل إنجازاتهم بصورة تعكس قدراتهم.
- مساعدة الطلاب على تقييم وتثمين أعمالهم باستخدام معايير موضوعية.
- تقديم أدلة ملموسة توثق مستوى إنجازاتهم وتطورهم.
- تنمية مهارات التحليل وحل المشكلات والتفكير المنظم بما يتوافق مع متطلبات التعلم في القرن الحادي والعشرين.

ثالثاً: في مجال التواصل:

- تقوية قنوات التواصل بين المتعلمين والمعلمين وأولياء الأمور من خلال الاطلاع المباشر على إنجازات الطالب.

- توفير شواهد واقعية تعكس مستوى التحصيل الدراسي للطلاب وجودة أعمالهم.

- تقديم صورة واضحة حول كيفية تحقيق الطلاب لأهداف المنهج الدراسي، مما يعزز الشفافية في العملية التعليمية (السعداوي، 2018، ص174).

3- مكونات الصحائف الوثائقية أو ملفات الإنجاز (Portfolios):

تتألف الصحائف الوثائقية أو ملفات الإنجاز من ثلاث مكونات رئيسة تُسهم في تقديم صورة شاملة عن تعلم الطالب وتطوره (Gredier, 1996; Storms et al., 1995):

أ- سيرة أعمال الفرد (Biographies of Works):

وتُشير إلى عملية توثيق مراحل تطور مشروع رئيسي يقوم به الطالب، بحيث يُعبر عن أفكاره وخبراته المتنامية في المجال الدراسي الذي اختاره. ويُظهر هذا المكون حجم الجهد المبذول من الطالب ويعكس مدى إصراره على تطوير عمل متكامل يمثل نتاجه الفكري والإبداعي.

ب- أعمال متنوعة للطالب (Range of Works):

ويشمل هذا المكون مجموعة من الأنشطة والأعمال التي أنجزها الطالب والتي تُظهر عمق إنتاجه وتنوع مهاراته. فعلى سبيل المثال يمكن أن يحتوي ملف إنجاز طالب في مجال المهارات اللغوية على مقالات كتبها أو نصوص شعرية نظمها أو محاورات صفية مسجلة أو غيرها من الأنشطة التي تبرز تدرج مهاراته اللغوية وتنوعها.

ج- انعكاسات الطالب (Student Reflections):

وفي هذا المكون يُطلب من الطالب على امتداد الفترة الدراسية أن يُعبّر عن رؤيته الخاصة لأعماله فيحدد جوانب القوة والتميز، ويُحلّل مواطن التحسن كما يُوضح أهدافه المستقبلية. وقد يُكَلِّف بدور "الناقد" لأعماله أو بدور "المؤرخ الشخصي" لمسيرته التعليمية. وتُعتبر هذه الانعكاسات عنصراً جوهرياً لأنها تُعزز وعي الطالب بذاته وتثري عملية تعلمه وتدفعه نحو النمو الذاتي والتفكير التأملّي (علام، 2000، ص746).

4- أنواع ملفات أعمال الطالب (ملفات الإنجاز):

تُعَدّ ملفات الإنجاز من الأساليب الحديثة في تقييم تعلم الطلبة وتتميز بمرونتها وتنوعها بحيث تتلاءم مع طبيعة الأهداف التعليمية واحتياجات المتعلم والمعلم على حد سواء. وتجدر الإشارة إلى

أن اختيار نوع الملف ليس قراراً نهائياً أو جامداً، إذ يمكن للمعلم أن يجرب أكثر من نوع قبل أن يستقر على الشكل الأكثر فاعلية ومناسبة لعملية التقويم، ويمكن تصنيف ملفات الإنجاز في أربع فئات رئيسية:

4-1- الملفات التجميعية (Collective Portfolios):

وهي ملفات تتضمن أعمال المتعلم التراكمية التي تبرز تطوره عبر الزمن، على سبيل المثال: عينات من الكتابة في كل مرحلة من مراحل التعبير الكتابي، تسجيلات صوتية لقراءة الشعر أو الخطابة أو نصوص نثرية بلغات مختلفة، إضافة إلى حلول مسائل رياضية مرتبطة بحل المشكلات. ويُستخدم هذا النوع عادة لتوثيق المسار التعليمي للطلاب بشكل متدرج.

4-2- ملفات العرض (Showcase Portfolios):

تركز على إبراز أفضل إنجازات المتعلم التي تعكس مستوى متقدماً من أدائه وغالباً ما تُختار بعناية لتعكس مهاراته المتميزة. قد تحتوي هذه الملفات أيضاً على بعض عناصر الملفات التجميعية، كما يمكن أن تشمل أعمالاً شخصية كالصور أو الجوائز أو شهادات التقدير لتمنح صورة شمولية عن شخصية المتعلم وتقدمه.

4-3- الملفات التقويمية (Assessment Portfolios):

تضم وثائق وبيانات تُستخدم لأغراض تقويمية مباشرة كعينات أعمال مرتبطة بمهام محددة، تقارير لقاءات بين المعلم والمتعلم أو أدلة على كفاءات مهنية متخصصة. وقد تحتوي كذلك على ملاحظات مكتوبة أو شهادات الالتحاق بدورات أو ورش عمل أو صور ووثائق داعمة. ويهدف هذا النوع إلى توفير معلومات موضوعية تفيد في إصدار الأحكام واتخاذ القرارات بشأن مستوى الطالب.

4-4- الملفات الرقمية (Digital Portfolios):

وهي ملفات تعتمد على البرمجيات والتقنيات الرقمية لتنظيم وتوثيق إنجازات المتعلم، حيث يُنشأ أرشيف إلكتروني يتضمن نصوصاً مكتوبة وصوراً ومواد سمعية أو مرئية. وتمتاز هذه الملفات بالمرونة وسهولة التحديث والاطلاع عليها، كما تتيح فرصاً لتقويم مهارات التفكير الإبداعي والتقني لدى الطالب، وغالباً ما تصاحب المتعلم في مسيرته التعليمية على المدى الطويل (الفرطوسي والميداني، 2021، ص331).

5- أهداف ملف الإنجاز:

يمثل ملف الإنجاز أداة تقويمية وتربوية متعددة الأبعاد إذ لا تقتصر أهميته على الطالب وحده بل تمتد لتشمل المعلم وولي الأمر والمشرّف التربوي، وتتجلى أهدافه في الآتي:

5-1- بالنسبة للطالب:

- تنمية مهارات تنظيم الأعمال وترتيبها وفق تسلسل منطقي يعكس مسار التعلم.
- تعزيز قدراته على التعلم الذاتي والتقييم الذاتي بما يسمح له باكتشاف نقاط القوة والضعف.
- تطوير مهارات التفكير العليا من خلال التحليل والنقد والمقارنة والتفسير.

5-2- بالنسبة للمعلم:

- توفير وسيلة موضوعية لمتابعة مدى تقدم المتعلم خلال فترة زمنية معينة.
- المساعدة في تنظيم عملية التقييم وربطها بالأنشطة الصفية واللاصفية.
- دراسة أوجه العلاقة بين أدوات التقييم المختلفة، مما يثري عملية إصدار الحكم على أداء المتعلم.

5-3- بالنسبة لولي الأمر:

- الحصول على دليل ملموس يوضح إنجازات المتعلم في المواد الدراسية المختلفة.
- تفعيل دوره في عملية التواصل التربوي مع المدرسة بما يعزز الشراكة بين البيت والمؤسسة التعليمية.

- التعرف على احتياجات المتعلم والإسهام في تقديم الدعم المناسب له بما يحقق تقدمه الدراسي.

5-4- بالنسبة للمشرفين التربويين:

- توفير مؤشر عملي على مستوى إنجازات المتعلم في ضوء الأهداف التربوية.
- المساعدة في اكتشاف القدرات الفردية والاستعدادات الخاصة لدى الطلبة.
- تزويد المدرسة بأدلة نوعية حول مستوى التحصيل الدراسي للمتعلمين.
- الكشف عن طبيعة الأنشطة التعليمية التي يقدمها المعلم للطلبة وجودتها (الفرطوسي والميداني، 2021، ص. 332).

6- المكونات الأساسية لملف الإنجاز:

- يتضمن ملف الإنجاز مجموعة من العناصر الرئيسة التي تهدف إلى إبراز تطور المتعلم وتوثيق خبراته التعليمية والتربوية، ومن أبرز هذه المكونات:
- أ- الغلاف: يعمل على حفظ محتويات الملف وحمايتها ويعطي انطباعاً أولياً عن جدية العمل وتنظيمه.
 - ب- الرسالة التوضيحية: وهي مقدمة قصيرة يوضح فيها الهدف من الملف وما يعكسه من تقدم المتعلم، إضافة إلى عرض موجز لشواهد تعلمه ونقاط تطوره عبر الفترة الزمنية المحددة.
 - ج- فهرس المحتويات: يسهل الاطلاع على مكونات الملف وتنظيمها مما يعزز من وضوح بنيته الداخلية.

د- الأعمال المضمّنة: وتشمل الواجبات والمهام التي كُلف بها المتعلم إلى جانب العناصر التي اختارها بنفسه، مما يحقق التوازن بين التوجيه الذاتي والتوجيه التعليمي.

هـ- التواريخ الزمنية: تدوين تاريخ إضافة كل عمل إلى الملف بما يسمح بتتبع المسار الزمني للتعلم ورصد مدى التقدم.

و- نسخ الأعمال: إدراج الأعمال في صورتها الأولية (المسودة) وصورتها النهائية لإبراز عملية التطوير والتحسين التي مر بها المتعلم.

ز- ورقة التغذية الراجعة: وهي عنصر محوري حيث يدوّن المتعلم فيها أفكاره ومشاعره تجاه الأعمال التي أنجزها، استنادًا إلى مجموعة من الأسئلة الإرشادية، مثل:

- ماذا تعلمت من هذا العمل؟

- ما الجوانب الإيجابية في أدائي؟

- لماذا اخترت هذه الفقرة تحديدًا وفقًا للمعايير المتفق عليها؟

- ما الجوانب التي أُرغب في تحسينها؟

- ما درجة الرضا عن أدائي؟

- ما أبرز الصعوبات والتحديات التي واجهتني أثناء الإنجاز؟

وتكمن أهمية هذه المكونات في كونها تتيح للمتعلم والمعلم معًا وسيلة منظمة لتوثيق عملية التعلم وتُظهر التطور المرحلي بموضوعية وشفافية (الفرطوسي والميداني، 2021، ص333).

يمكن أن يأخذ ملف الإنجاز أشكالًا متعددة، بدءًا من مجلد ورقي تقليدي إلى منصات إلكترونية حديثة (E-Portfolio) تعتمد على الوسائط المتعددة، وتشمل محتوياته عادةً عناصر كسجل تدريبي، دفتر يوميات، تقارير مكتوبة، صور فوتوغرافية، رسومات، مقالات، خرائط، جداول ورسوم بيانية، نسخ من الشهادات، منتجات رقمية، نتائج اختبارات، خطط تعليمية أو تدريبية، بالإضافة إلى ملفات سمعية وبصرية كالتسجيلات الصوتية ومقاطع الفيديو (Morrow, et al, 2000, p.167).

وقد أشار أنتوينات (Antoinette, 1996) إلى قائمة واسعة من المحتويات التي يمكن أن يتضمنها ملف أعمال الطالب كالتسجيلات القصصية وشرائط التسجيل السمعية والكتابات التأملية والملخصات الخاصة بالزيارات أو الرحلات الميدانية، إلى جانب المراجعات وتقييمات الزملاء واللوحات والأعمال الفنية وأشرطة الفيديو والرسوم التخطيطية والصور والرسوم البيانية، فضلًا عن موضوعات الإنشاء أو التعبير والتقارير الجماعية والمذكرات واللقاءات والحوارات ومنتجات أعمال الحاسوب. ويُعزى نجاح ملف الأعمال إلى مجموعة من العوامل الجوهرية، من أبرزها:

أ- تنمية مهارات المتعلمين وتعديل اتجاهاتهم من أساليب التعلم التقليدية القائمة على الحفظ والتلقين إلى أساليب أكثر حداثة تركز على النمو الشامل للطلاب وتعزيز ثقته في تحليلاته الذاتية لأدائه وتقييمه لذاته مع إتاحة الفرصة له لتوثيق إنجازاته وتحمل مسؤولية تعلمه.

ب- تعزيز قنوات التواصل بين المعلمين والآباء والطلبة من خلال عقد لقاءات دورية تُعنى بمناقشة الملفات ومتابعة الإنجازات المحققة.

ج- تصميم إجراءات دقيقة للتقويم الوثائقي الشامل تشمل آليات جمع البيانات والمعلومات وتحليل محتوى الملفات بشكل منهجي.

د- تحديد محكات واضحة للتقويم والحكم على نوعيته والإفادة من نتائج هذه العملية في تحسين الممارسات التعليمية.

وبذلك يصبح ملف الأعمال أداة تعليمية وتقويمية فعّالة تُسهم في دعم تعلم الطالب وتعزيز نموه المتكامل (علي مهيدات، 2008، ص163).

7- خصائص ملف الإنجاز:

يمتاز ملف الإنجاز بمجموعة من السمات التي تجعله أداة فعّالة ليس فقط للتقويم، بل أيضاً لتخطيط الأنشطة التعليمية وتنفيذها بما يتناسب مع خصوصية تعلم الطالب، ومن أبرز هذه الخصائص ما يلي:

أ- سهولة تكيفه مع تفرد التعلم:

يُعد كل طالب ملفاً خاصاً به يضم أعماله وإنجازاته ويُقيّم بشكل مستقل مما يجعل الملف انعكاساً صادقاً لتجربة تعليمية شخصية تعكس تفرد كل متعلم. فمثلاً في مهارة الكتابة قد يبدأ الطالب من كتابة كلمة أو فقرة ويتطور تدريجياً إلى إنتاج مقال يُنشر في صحيفة أو حتى إصدار كتاب مع مراعاة الفروق الفردية وتوظيف أقصى طاقات كل طالب.

ب- تركيزه على تقويم النتائج:

لا يقتصر الملف على عرض أعمال الطلبة النهائية فحسب بل يتضمن كذلك مراحل الإنجاز المختلفة كالمسودات والتعديلات ويظهر بوضوح تطور العمل عبر ملاحظات المعلم أو المناقشات الصفية، وهذا يكشف عن أصالة العمل ويؤكد أنه جهد حقيقي للطلاب نفسه.

ج- إبراز جوانب القوة:

يسلط الملف الضوء على نقاط القوة لدى المتعلم ويشجعه على اختيار أفضل أعماله مما يعزز الدافعية ويجعل من الملف تجربة إيجابية. أما الأخطاء ونقاط الضعف فيُنظر إليها على أنها فرص

للتحسين وليست مؤشرات عجز، الأمر الذي يقلل من القلق المرتبط بالتقويم ويُشعر الطالب بإمكانية التطور المستمر.

د-إشراك الطالب في عملية التقويم:

يتعاون الطالب والمعلم في جميع مراحل إعداد الملف، بدءًا من تحديد النتائج المطلوبة وصولًا إلى اختيار الأعمال النهائية، كما يُشارك الطالب في تقويم أدائه بالاستناد إلى معايير واضحة. وهذا يعزز الشعور بالمسؤولية الذاتية ويُكوّن لديه تصورًا مسبقًا عن طبيعة الأداء الجيد.

هـ- تسهيل انتقال وتبادل الخبرات:

يوفر الملف أدلة ملموسة على تعلم الطالب يمكن مشاركتها مع الزملاء والمعلمين وأولياء الأمور مما يعزز فرص الحوار والتغذية الراجعة. كما يمكن أن تُعرض الأعمال المتميزة كنماذج أمام الطلبة بما يسهم في نقل الخبرات وتبادلها بين الطلاب أنفسهم.

و- الحاجة إلى وقت أطول:

يُعد ملف الإنجاز من أدوات التقويم التي تستهلك وقتًا أطول مقارنة بالأساليب التقليدية إذ قد تمتد فترة استخدامه لأكثر من عام. كما أن مراجعته تتطلب وقتًا كافيًا من المعلم لمناقشته مع الطالب، حيث أشارت الدراسات إلى أن المراجعة الأولية تستغرق نحو 30 دقيقة إضافة إلى 30 دقيقة أخرى للمناقشة إلا أن هذه المدة تتناقص تدريجيًا مع اكتساب المعلم خبرة أكبر في إدارة الملف. وتوصي الدراسات بمراجعته مرة كل ستة أسابيع عند استخدامه إلى جانب أدوات تقويم أخرى.

ز- ثبات ملف الإنجاز:

تشير الدراسات إلى أن معامل ثبات ملف الإنجاز كأداة تقويم يتراوح بين (0.31 – 0.43)، وهي قيمة منخفضة نسبيًا، إلا أن هذا المستوى من الثبات يُعتبر مقبولًا إذا ما قورن بالاختبارات الصفية التي يعدها المعلمون والتي لا يتجاوز معامل ثباتها في المتوسط (0.45) مما يبرر اعتماد الملف كأداة تقويم داعمة (على مهيّدات، 2008، ص165).

8- خطوات تصميم واستخدام ملف الإنجاز:

أ- **تحديد الهدف والنوع:** يبدأ العمل بتحديد الهدف من ملف الأعمال كمف التحصيل الدراسي، ملف التقدم، مع مراعاة أن يكون التركيز في البداية على مادة واحدة أو وحدة دراسية محددة.

ب- **اختيار الوعاء المناسب:** يحدد الوعاء الذي يتناسب مع الهدف (حقيقية، ملف ورقي، ملف إلكتروني...)، كما يُفضل تخصيص مكان محدد في الصف لحفظ الملفات وتيسير استخدامها بشكل روتيني.

ج- تقديم الفكرة للطلاب: يتم شرح مفهوم ملف الأعمال وأهميته للطلاب مع إعطاء أمثلة من واقع العلماء والمشاهير والسماح للطلاب بإضفاء لمسة شخصية على ملفاتهم.

د- تصميم المحكات: يتفق مع الطلاب على المحكات والمعايير التي سستخدم في اختيار الأعمال، وتعرض هذه المحكات في الفصل ليتم تطويرها باستمرار بما ينسجم مع أهداف التعلم.

هـ جمع الأعمال: يُسمح للطلاب بوضع أعمالهم في الملفات وفقاً للمحكات المتفق عليها.

و- التدريب على الاستخدام: يُدرّب الطلاب على كيفية استخدام المحكات في انتقاء أعمالهم، وكذلك على استخدام النماذج الخاصة بالتأمل الذاتي أو غيرها من النماذج المرتبطة بالملف.

ز- تصميم محكات التقويم: تُصمم محكات تقويم ملف الأعمال بمشاركة الطلاب وتستخدم كأداة للتواصل المستمر معهم.

ح- جدولة المؤتمرات التعليمية: يُعد جدول زمني لعقد لقاءات فردية أو جماعية مع الطلاب لمراجعة الملفات وتحديد أهداف جديدة للتعلم.

ط مشاركة أولياء الأمور: يُنظم احتفال أو معرض لعرض إنجازات الطلاب ومناقشتها بحضور أولياء الأمور والزملاء مع اختيار أفضل عمل لكل طالب.

ي- التطوير المستمر: يتم تطوير ملفات الأعمال باستمرار بالاستفادة من إبداعات الطلاب وأفكارهم المبتكرة (السعدوي، 2018، ص179).

9- تصحيح ملف الإنجاز:

يُعدّ ملف إنجاز المتعلم أداة مهمة إذا استخدم بشكل منهجي ومنظم، حيث يتيح تقويماً أكثر شمولية وعمقاً لأداء الطالب مقارنة بالاختبارات التقليدية. غير أنّ فعاليته تتوقف على مدى دقة وموضوعية إجراءات التصحيح. إذ قد تتباين الدرجات بشكل ملحوظ بين أفراد الصف الواحد إذا لم يُستند في التقويم إلى محكات واضحة ومعايير محددة، ولتحقيق العدالة والموضوعية وضمان اتساق عملية التقويم، ينبغي مراعاة ما يلي:

أ- تحديد الهيكل الأساسي للملف: على المعلم إعداد جدول يوضح الأقسام الرئيسية التي يجب أن يتضمنها ملف الإنجاز كالتقارير، الأنشطة الإبداعية، التأملات الذاتية، مع تحديد مستوى الأداء والخصائص المطلوبة في كل قسم.

ب- تخصيص درجات مناسبة لكل قسم: بحيث تعكس هذه الدرجات الوزن النسبي لأهمية القسم في العملية التعليمية.

ج- توزيع الدرجات وفق مستويات الأداء: يجب أن تُمنح أعلى الدرجات للأداء المتميز، مع توفير مجال من الدرجات لكل مستوى حتى يُتاح التدرج في التقدير بدلاً من الاختصار على نقطة تقييمية واحدة مما يضمن مرونة ودقة أكبر.

د- حساب الدرجة النهائية بشكل موضوعي: تُجمع الدرجات المحصلة من مختلف الأقسام ثم تُحوّل إلى نسبة مئوية من الدرجة الكلية بما يوفر صورة شاملة عن مستوى الطالب وأدائه.

إنّ تبني مثل هذا الإطار المنهجي في تقويم ملفات الإنجاز لا يسهم فقط في رفع مستوى الموضوعية والدقة، بل يعزز أيضاً وعي المتعلم بمستوى أدائه ويحفزه على تحمل مسؤولية تعلمه وتطوير مهاراته بصورة مستمرة (الفرطوسي والميداني، 2021، ص.338).

10- تقييم ملف الإنجاز:

يُعدّ ملف الإنجاز وسيلة تقويمية حديثة تتطلب استخدام أدوات ومعايير واضحة تضمن دقة وموضوعية النتائج، ولتحقيق فعالية هذا النوع من التقويم ينبغي السير وفق خطوات منهجية منظمة تتمثل فيما يلي:

أ- تحديد الأهداف الرئيسية من الملف: إذ لا يمكن الحكم على أعمال المتعلم دون وضوح الغاية، سواء أكانت تشخيصاً لمستواه أم متابعة لتطوره أم حكماً على كفاءته النهائية.

ب- تحديد المحكات والمعايير: بحيث يضع المعلم مؤشرات أداء دقيقة ومحددة يمكن الاحتكام إليها عند الحكم على جودة الأعمال.

ج- تجميع العينات المناسبة من إنتاج المتعلم: على أن تكون هذه العينات ممثلة لأهم جوانب تعلمه ومعبرة عن نموه في فترة زمنية محددة.

د- إجراء عملية التقويم وفق المحكات الموضوعية مسبقاً: مما يضمن الشفافية ويقلل من الذاتية في الحكم.

هـ- صياغة تقرير مفصل (تغذية راجعة): يقدم للمتعلم وولي أمره بحيث يتضمن نقاط القوة وجوانب التحسين وخطط المتابعة المستقبلية الأمر الذي يجعل الملف أداة تعليمية بنائية وليس مجرد وسيلة للحكم أو التصنيف.

وبهذا، يصبح ملف الإنجاز أداة تقويمية ذات بعد تشاركي يجمع بين البعد الكمي المتمثل في الدرجات والبعد الكيفي المتمثل في التحليل النوعي للأداء (الفرطوسي والميداني، 2021، ص.335).

11- ملف الإنجاز الإلكتروني (E-Portfolio):

ظهر ملف الإنجاز الإلكتروني استجابةً للقصور الذي تعاني منه الطرائق التقليدية في التقييم، وخاصة تلك التي تعتمد على الانطباعات الذاتية للمعلم والتي غالباً ما يشعر المتعلمون بأنها غير عادلة ولا تعكس جهودهم الحقيقية.

يُعد ملف الإنجاز الإلكتروني وعاءً رقمياً متكاملًا يتيح تخزين مختلف المحتويات البصرية والسمعية بما في ذلك النصوص، مقاطع الفيديو، الصور، والتسجيلات الصوتية، كما يُصمم لدعم عمليات تربوية متعددة وأغراض متنوعة في مجال التقييم.

ويتضمن هذا النوع من الملفات مجموعة من النتائج التعليمية كالعروض التقديمية، الشروحات، الموارد، والإنجازات التي تعبّر عن الفرد أو الجماعة أو المؤسسة التعليمية. ويمكن أن تُحفظ هذه الأعمال بأشكال مختلفة (نصوص، رسومات، وسائط متعددة) سواء على مواقع إلكترونية عبر شبكة الإنترنت أو على وسائط رقمية مثل الأقراص المضغوطة و (DVD).

كما يشكل ملف الإنجاز الإلكتروني سجلاً منظماً لأفضل أعمال الطالب المتميزة من دروس ومشاريع وأنشطة، بحيث تختلف مكوناته من طالب لآخر تبعاً لفلسفته التربوية في تنظيمه. ويعتمد الملف في عرض محتوياته على تقنيات الوسائط المتعددة مع إمكانية التنقل بين عناصره عبر روابط إلكترونية (Links). وإضافة إلى ذلك يمكن نشره ومشاركته عبر الإنترنت أو توزيعه باستخدام وسائط إلكترونية مختلفة (الفرطوسي والميداني، 2021، ص339).

12- بعض مشكلات تقييم التحصيل باستخدام الصحائف الوثائقية (ملفات الأعمال):

يُعدُّ تحديد ما يجب تقييمه – أي الأهداف التي يسعى نظام التقييم الوثائقي الشامل Portfolio Assessment إلى تحقيقها – من أبرز التحديات التي تواجه القائمين على تطبيق هذا النظام. فقد أظهرت خبرات بعض الدول المتقدمة في هذا المجال وجود تباين واختلاف بين المسؤولين حول ما ينبغي أن يُقيّم في ملفات الأعمال، الأمر الذي دفع العديد من البرامج إلى الاعتماد على الأهداف الوطنية العامة، ومرشد المناهج الدراسية، وتوجيهات الخبراء، لاستخلاص عناصر أساسية محدودة من هذه المصادر المتنوعة لتكون مرجعاً للتقييم، بدلاً من الاقتصار على قوائم مهارية جامدة (علام، 2000، ص749).

كما أن تعدد تعريفات الصحائف الوثائقية أدى إلى تنوع أساليب تقييمها وفقاً للظروف المحلية وإمكانات القائمين عليها. ففي بعض النظم التربوية يتم منح درجات لكل جزء من الملف ثم حساب المتوسط للتوصل إلى تقدير عام، بينما تركز نظم أخرى على إصدار أحكام كلية حول مجالات أداء

محددة في ضوء معايير منصوص عليها في دليل المعلم. وعلى الرغم من تباين المداخل إلا أن جميعها تؤكد ضرورة الاستناد إلى محكات أداء واضحة ومتفق عليها مجتمعياً تظهر مستويات الإنجاز المختلفة وقد تُصاغ أحياناً في صورة سلالمة تقدير تتكون من خمسة أو ستة مستويات.

ويتطلب هذا الدور من المعلم أن يشارك بفعالية في تصميم مكونات الملف وإرشاد الطلاب أثناء إعدادهم لأعمالهم، وتنظيم الجدول الزمني لعملية بناء الملفات داخل الصف، والمساهمة في بناء معايير دقيقة لتقويمها نظراً لكونه الأقدر على التمييز بين الأداء المناسب وغير المناسب والتعرف على أنسب الظروف لتطبيق هذا النظام ومراجعة نتائجه. وهذا يستدعي بطبيعة الحال تأهيلاً مهنيّاً متقدماً للمعلم ليتمكن من أداء هذه الأدوار بكفاءة (علام، 2000، ص750).

أما فيما يتعلق بصيغ التقييم المتنوعة للصحائف فإن بعضها يركز على الجوانب التوثيقية أو التحصيلية دون الاهتمام الكافي بانعكاسات الطالب أو تنمية مهاراته في التأمل الذاتي. وغالباً ما تُستخدم الصيغة التقويمية لتقديم بيانات حول نوعية تحصيل الطالب والمساهمة في تطوير التعليم داخل الصف. إلا أن هناك تعارضاً محتملاً بين هدف قياس التحصيل وتطوير العملية التعليمية إذ إن التركيز على تحقيق أحدهما قد يضعف فاعلية الآخر الأمر الذي ينعكس على طبيعة وبنية ملفات الأعمال.

ومن أبرز المشكلات أيضاً قضية الصدق والثبات في تقدير ملفات الأعمال، فالسؤال الجوهرى المتعلق بالصدق هو: إلى أي مدى تعكس إنجازات الطالب المدونة في الملف البنية الفرضية الحقيقية التي يُراد قياسها؟ وبسبب اتساع نطاق محتوى المواد الدراسية وتعدد تفسيراتها يصبح من الصعب الاعتماد على هذه التعريفات وحدها للتحقق من صدق التقدير. وقد حاول بعض الباحثين معالجة هذه الإشكالية من خلال دراسة الارتباط بين درجات الملفات ونتائج الاختبارات التقليدية في المجال نفسه، غير أن النتائج أظهرت ارتباطات ضعيفة.

أما بالنسبة للثبات فقد استخدمت دراسات عديدة درجة اتفاق المحكمين كأساس لقياس ثبات التقديرات إلا أن النتائج كانت منخفضة أيضاً. ويرجع ذلك في كثير من الأحيان إلى عدم وضوح مصطلحات سلالمة التقدير (Scoring Rubrics) أو تعقيدها أو إلى قلة تدريب القائمين بعملية التقدير، إضافة إلى تباين المهام بين الصفوف الدراسية المختلفة أو اختلاف الظروف المصاحبة لعملية التطبيق (علام، 2000، ص750).

ومن المهم التأكيد على أن تفعيل متطلبات نظام التقييم التوثيقي الشامل في البيئة الصفية Classroom-Based Portfolio Assessment يستلزم مواكبة التطورات التقنية في المدارس من خلال

تعزيز استخدام الحاسوب وتوظيف برمجيات حديثة، إضافة إلى ابتكار صيغ وأساليب تنظيمية جديدة لهذا النظام بحيث تكون عملية سهلة الاستخدام ولا تستهلك مساحة كبيرة.

ويُعدّ هذا النظام من الاتجاهات الحديثة والفعّالة في تقويم المتعلم إذ يتيح توثيق نموه التحصيلي بصورة تدريجية وشمولية عبر مجموعة واسعة من الأدوات والأساليب كقوائم المراجعة، الملاحظات الصفية، السجلات القصصية، مقاييس التقدير، كتابات الطلاب وبحوثهم، المحاكاة بالحاسوب، العروض الشفوية والسمعية والبصرية، إضافة إلى الصور، الرسوم، القصص، التسجيلات، المقابلات الفردية والجماعية، والمؤتمرات التعليمية. إن هذا التنوع والثراء في أدوات التوثيق والتقويم يتيح فرصة لتقييم إنجازات الطالب تقييماً شاملاً Holistic Evaluation يعكس واقعية الموقف التعليمي بشكل أعمق وأكثر مصداقية مقارنة بما تقدمه الاختبارات التقليدية التي ما تزال تمثل الأداة السائدة في الممارسات التقييمية المعاصرة (علام، 2000، ص751).

الفصل العاشر

القياس التربوي في العصر الرقمي

الأهداف (Objectives):

يهدف هذا الفصل إلى تعريف الطالب بأهم التطورات المعاصرة في مجال القياس والتقويم النفسي والتربوي، وذلك من خلال تسليط الضوء على ثلاثة محاور رئيسية. أولاً، التعرف على الاختبارات الإلكترونية (E-tests) من حيث مفاهيمها ومميزاتها وتحدياتها، وما أحدثته من نقلة نوعية في طرائق تطبيق الاختبارات. ثانياً، فهم آلية عمل الاختبارات التكيفية المحوسبة (CAT) ودورها في تحسين دقة القياس وكفاءته من خلال تكييف الفقرات بحسب مستوى المفحوص. وثالثاً، استكشاف ملامح القياس النفسي في عصر الآلة الذكية وما يحمله من فرص وتحديات مستقبلية في ضوء الذكاء الاصطناعي وتكنولوجيا البيانات. ومن خلال هذه الأهداف، يسعى الفصل إلى تزويد القارئ بفهم متكامل حول انعكاسات التكنولوجيا الحديثة على القياس النفسي وأساليبه.

المحتوى (Contents):

- 1- الاختبارات الإلكترونية (E-tests)
- 2- الاختبارات التكيفية المحوسبة (CAT)
- 3- القياس التربوي والنفسي في عصر الآلة الذكية

تمهيد:

أدى التطور التكنولوجي المتسارع إلى إحداث نقلة نوعية في أساليب القياس النفسي والتربوي، حيث لم تعد الاختبارات التقليدية هي الوسيلة الوحيدة لتقويم المتعلمين وتشخيص قدراتهم. فقد ظهرت الاختبارات الإلكترونية (E-tests) بوصفها بديلاً عصرياً يوفر السرعة والدقة وسهولة التطبيق، تلتها الاختبارات التكيفية المحوسبة (CAT) التي تمثل شكلاً أكثر تطوراً يعتمد على توجيه أسئلة متناسبة مع مستوى أداء المفحوص بشكل لحظي. ومع دخول عصر الذكاء الاصطناعي أصبح القياس النفسي والتربوي أمام آفاق جديدة تسعى إلى استثمار إمكانات الآلة الذكية في تقديم اختبارات أكثر موضوعية وفاعلية، بما يواكب متطلبات التعليم الحديث ويعزز جودة القرارات التربوية والنفسية. ومن هذا المنطلق، يهدف هذا الفصل إلى استعراض هذه المستجدات وبيان أهميتها في تطوير الممارسة التقييمية.

1- الاختبارات الإلكترونية (E-tests):

أصبح التعليم الإلكتروني في الوقت الراهن أحد أبرز مظاهر التطور في العملية التعليمية، إذ جاء استجابة للتغيرات السريعة التي يشهدها العالم في ميدان التكنولوجيا الرقمية، فضلاً عن التحديات التي فرضتها الأزمات الصحية العالمية كجائحة كورونا، والتي أدت إلى توقف التعليم الجاهي في معظم دول العالم، مما استدعى التحول نحو أنماط جديدة من التعليم تعتمد على الوسائط المتعددة والأجهزة التكنولوجية الحديثة، لتغدو ضرورة لا غنى عنها في شتى مجالات الحياة.

ومن بين الأدوات الرئيسة في هذا النمط من التعليم تبرز الاختبارات الإلكترونية، وهي تلك الاختبارات التي تُبنى وتُجرى عبر بيانات رقمية من خلال الاتصال بالإنترنت باستخدام أجهزة متنوعة كالحواسيب، والأجهزة اللوحية، والهواتف الذكية. وقد أثبتت هذه الاختبارات كفاءتها وفعاليتها في دعم العملية التعليمية، لما تتميز به من خصائص لا توفرها الاختبارات التقليدية، إذ تتصف بالمرونة والسرعة في الإعداد والدقة في التصحيح إلى جانب مساهمتها في تقليل الجهد والوقت وإضفاء عنصر التشويق والمتعة على تجربة المتعلم (التميمي والشيخ، 2021، ص.169).

1-2- مفهوم الاختبارات الإلكترونية:

يمكن تعريف الاختبار الإلكتروني بأنه أداة حديثة لتقويم المتعلم رقمياً تتيح للمعلم تصميم اختبارات بصورة مبسطة وقابلة للتطبيق المباشر، حيث تُصحح آلياً وفورياً بما يضمن مستوى عالٍ من المصداقية والشفافية ويعزز من موضوعية النتائج ودقتها.

وهي إحدى تقنيات التكنولوجيا التي يمكن توظيفها للتغلب على بعض الصعوبات التي قد تعيق تنفيذ الاختبارات التقليدية (الورقية)، فضلاً عن إمكانية استثمارها في رفع مستوى التحصيل العلمي للمتعلمين وتعزيز مهاراتهم في التعلم الذاتي. فالاختبارات الإلكترونية هي اختبارات تُبنى وتُنشر فقراتها عبر شبكة الإنترنت، بحيث يتاح للمتقدمين الإجابة عنها في الوقت نفسه باستخدام الوسائط الرقمية المتنوعة (Billo et al., 2016).

وفي هذا السياق يلعب التقويم الإلكتروني دوراً محورياً في التحول الجوهري الذي شهدته العملية التعليمية، إذ أسهم في إعادة صياغة المنهج بحيث أصبح يركز على المتعلم بوصفه المحور الرئيس للتعلم. ويُنظر إلى هذا النوع من التقويم باعتباره عملية منظمة لجمع وتحليل البيانات المستقاة من مصادر متعددة للوصول إلى فهم أعمق لمعارف المتعلم بما يتيح تقييم أدائه بدقة وموضوعية وبصورة بناءة (Dermo, 2009, p.305).

ومن هذا المنطلق يمكن القول إن الاختبارات الإلكترونية تمثل أحد أهم تطبيقات التقويم الإلكتروني، حيث تُجرى باستخدام أجهزة مختلفة كالهواتف الذكية، والأجهزة اللوحية، والحواسيب، بهدف قياس أداء المتعلم وتقييمه بصورة رقمية. ويتولى المعلم مسؤولية إعداد هذه الاختبارات بدءاً من تحديد المادة العلمية المستهدفة مروراً باختيار نمط الأسئلة وصياغتها وآلية الإجابة والتصحيح الإلكتروني وصولاً إلى تحديد آليات توزيع الدرجات وضبط الوقت المخصص للاختبار وإغلاقه بشكل تلقائي.

ولتعزيز نزاهة الاختبارات الإلكترونية والحد من فرص الغش، ظهرت مجموعة من الآليات والتقنيات التي يمكن للمعلم أو المؤسسة التعليمية توظيفها لتحقيق أعلى درجات المصداقية. ومن أبرز هذه الآليات تصميم اختبارات تركز على الفهم والتحليل أكثر من الحفظ والاسترجاع، والاعتماد على المشاريع والاختبارات الشفوية كوسائل بديلة أو مكملّة، بالإضافة إلى تنويع النماذج وتوليد أسئلة جديدة لم يسبق استخدامها، مع فرض قيود تمنع التصوير أو التواصل بين الطلبة أثناء الاختبار. كما يمكن الاستعانة بكاميرات مراقبة مزوّدة بتقنيات التعرف على الوجه (Facial Recognition) لمتابعة سلوك المتقدمين، إلى جانب منع استخدام سماعات الأذن، وتطبيق متصفحات متخصصة، مثل: (LockDown Browser) التي تحدّ من النسخ واللصق أو فتح أي برامج موازية أثناء أداء الاختبار.

وفي إطار التطوير المستمر يجري العمل حالياً على إدماج تقنيات الذكاء الاصطناعي في منظومة الاختبارات الإلكترونية، حيث طُوّرت تطبيقات قادرة على التحقق من هوية المتقدمين عبر

مقارنة أنماط ضغوطات المفاتيح الخاصة بهم بما تم تسجيله مسبقاً لكل طالب وهو ما يُسهم في رفع مستوى الأمان والموثوقية لهذه الاختبارات (التميمي والشيخ، 2021، ص.169).

1-3- خصائص الاختبارات الإلكترونية:

تتسم الاختبارات الإلكترونية بمجموعة من الخصائص التي تجعلها أداة فعالة ومفيدة في العملية التعليمية، وقد أشار إلى بعضها عدد من الباحثين (زيتون، 2005، ص259؛ صبحي، 2005، ص. 222؛ Basu et al., 2007)، ويمكن تلخيصها فيما يلي:

- أ- تسهم في التخفيف من بعض المشكلات النفسية التي يواجهها المتعلمون عادة في الاختبارات التقليدية كقلق الامتحان والخوف وصعوبة فهم صياغة الأسئلة.
- ب- تمتاز بسهولة الإعداد وذلك من خلال الاستعانة ببنوك الأسئلة ونشرها إلكترونياً.
- ج- تساعد على توفير الوقت والجهد مقارنة بالاختبارات الورقية.
- د- تتيح إمكانية تعديل الأسئلة بسهولة سواء بالحذف أو الإضافة أو معالجة الأخطاء الطباعية.
- هـ- توفر إمكانية إعداد صور متكافئة متعددة للاختبار نفسه بسرعة ويسر.
- و- تقلل من التكاليف المادية المرتبطة بالطباعة والتوزيع.
- ز- تسمح بتنوع أنماط الأسئلة بما يتناسب مع الأهداف التعليمية.
- ح- تسهم في زيادة دافعية المتعلمين بفضل اعتمادها على وسائط متعددة في عرض الأسئلة.
- ط- توفر فرصاً للتفاعل المباشر بين المتعلم والاختبار.
- ي- تتيح إمكانية قياس مهارات يصعب رصدها عبر الاختبارات التقليدية.
- ك- تتسم بالمرونة من حيث الزمان والمكان إذ يمكن للمتعلمين أدائها في أوقات وأماكن مختلفة.
- ل- تقدم ميزة التصحيح الإلكتروني الفوري بمجرد انتهاء المتعلم من الإجابة.
- م- تحتفظ ببيانات دقيقة عن سلوك المتعلم أثناء الاختبار كوقت الدخول والمدة المستغرقة للإجابة.
- ن- تتيح إرسال النتائج بسهولة إلى المتعلمين عبر البريد الإلكتروني.
- س- تزود المعلم بخيارات متعددة في تصميم وبناء الاختبار بما يحقق أهدافه التعليمية.
- ع- تقدم تغذية راجعة فورية وتعزيزاً مباشراً، مما يساعد المتعلم على ترسيخ معارفه واستخدامها بشكل أفضل.

1-4- أهداف الاختبارات الإلكترونية:

تسعى الاختبارات الإلكترونية إلى تحقيق مجموعة من الأهداف التربوية التي تعزز من فاعلية العملية التعليمية وجودتها، ويمكن تلخيص أبرزها فيما يلي:

أ- مراعاة الفروق الفردية بين المتعلمين من خلال تنويع أنماط الأسئلة وطرق عرضها بما يتناسب مع مستوياتهم المختلفة.

ب- الكشف عن نقاط القوة والضعف لدى المتعلمين بصورة دقيقة بما يساعد على توجيه الجهود التعليمية نحو احتياجاتهم الفعلية.

ج- توفير نتائج سريعة وموثوقة وهو ما يمنح المعلم والمتعلم فرصة الاستفادة من الوقت في تحسين الأداء.

د- الإسهام في تحسين وتطوير العملية التعليمية عبر الاستفادة من البيانات التي توفرها الاختبارات الإلكترونية وتحليلها.

هـ- تزويد المتعلمين بتغذية راجعة فورية تمكّنهم من معرفة مستوى تقدمهم وتصحيح مسار تعلمهم بشكل مستمر (التميمي والشيخ، 2021، ص.172).

1-5- عناصر بناء الاختبار الإلكتروني:

أشار السيد علي (2016) إلى أن هناك مجموعة من العناصر الأساسية التي ينبغي مراعاتها عند تصميم وبناء الاختبارات الإلكترونية، ويمكن تلخيصها فيما يلي:

أ- صياغة الأسئلة من حيث النوع، العدد، والزمن المخصص للإجابة على كل سؤال.

ب- اختيار الوسائط المتعددة الملائمة لدعم محتوى الاختبار.

ج- تقديم تغذية راجعة فورية وهادفة للمتعلمين.

د- وضع تعليمات واضحة ومباشرة للاختبار الإلكتروني.

هـ- توفير أدوات فعّالة للتفاعل مع الاختبار.

و- تحديد أنواع الإجابات المتوقعة من المتعلمين.

1-6- أشكال الاختبارات الموضوعية:

تُعد الاختبارات الموضوعية من أكثر أشكال الاختبارات شيوعاً في الميدان التربوي والتعليمي، وذلك لما تتميز به من وضوح في الصياغة وسهولة في التصحيح وإمكانية تحقيق قدر عالٍ من الموضوعية والعدالة بين المتعلمين. وتتعدد صور هذه الاختبارات لتناسب مع طبيعة الأهداف التعليمية ومستوياتها المعرفية المختلفة، ومن أبرز هذه الأشكال:

أ- **اختبارات التكملة (Completion Tests):** يُطلب من المتعلم إكمال جملة أو عبارة ناقصة بكلمة أو أكثر، مما يساعد على قياس مدى استيعاب المفاهيم الأساسية والقدرة على التذكر.

ب- اختبارات ملء الفراغ (Fill-in-the-Blank Tests): تشبه اختبارات التكملة لكنها غالباً ما تُصاغ في شكل عبارات تتطلب إدراج مصطلح محدد أو رقم أو رمز لإتمام المعنى بشكل صحيح.

ج- اختبارات الإجابات القصيرة (Short Answer Tests): يُقدّم للطالب سؤال مباشر يتطلب إجابة موجزة وغالباً ما تستخدم لقياس الفهم الأساسي والتفسير البسيط.

د- اختبارات المقارنة (Matching Tests): تعتمد على مطابقة عناصر من قائمتين كربط المفاهيم بتعريفها أو الأسماء بإنجازاتها، مما يساعد في قياس قدرة الطالب على الربط بين المعلومات.

هـ- اختبارات الاختيار من متعدد (Multiple Choice Tests): من أكثر الأنماط استخداماً حيث يُعرض سؤال واحد مع عدة بدائل على أن يختار المتعلم الإجابة الصحيحة، ويتميز هذا النمط بقدرته على قياس مستويات معرفية متعددة بدءاً من التذكر وحتى التحليل.

و- اختبارات الإجابات المتعددة (Multiple Response Tests): تشبه أسئلة الاختيار من متعدد إلا أن الطالب مطالب بتحديد أكثر من إجابة صحيحة، وهو ما يعكس تعقيداً أكبر في صياغة السؤال وفي قياس الفهم العميق.

ز- أسئلة دراسة الحالة (Case Study Questions): يقدم موقف أو حالة واقعية ويُطلب من الطالب تحليلها أو تفسيرها أو اقتراح حلول لها، مما يتيح قياس مستويات عليا من التفكير كالتطبيق والتحليل.

ح- اختبارات الصحة والخطأ (True/False Tests): يقدم للطالب عبارات ليحدد مدى صحتها أو خطئها. ورغم سهولتها وسرعة الإجابة عليها إلا أنها قد لا تعكس بدقة مستوى الفهم العميق ما لم تُصنع بعناية (التميمي والشيخ، 2021، ص.173).

1-7- أنواع الاختبارات الإلكترونية:

يمكن تصنيف الاختبارات الإلكترونية إلى نوعين أساسيين:

أ- الاختبارات الإلكترونية غير المتكيفة (Non-Adaptive Tests):

وهي صورة مطابقة تقريباً للاختبارات الورقية التقليدية ولكن في بيئة محوسبة. حيث تُعرض الأسئلة لجميع المتعلمين بالعدد والترتيب نفسه وبالشكل ذاته الذي يظهر على الورق دون أي تغيير أو تعديل في مستوى الصعوبة. هذا النوع من الاختبارات يسهل تصميمه وإدارته ويُستخدم عادة في الامتحانات التي تهدف إلى قياس التحصيل العام أو المقارنة بين متعلمين في بيئات مختلفة.

ب- الاختبارات الإلكترونية المتكيفة (Computerized Adaptive Tests - CAT):

وهي اختبارات أكثر تطوراً وذكاءً إذ تتكيف مع مستوى المتعلم بشكل لحظي. حيث يحدد مستوى صعوبة السؤال التالي بناءً على إجابة المتعلم للسؤال السابق: فإذا أجاب المتعلم بشكل صحيح،

ينتقل الاختبار إلى سؤال أكثر صعوبة أما إذا أخطأ في الإجابة فينتقل إلى سؤال أبسط. هذه الآلية تسمح بتقدير مستوى المتعلم بدقة وكفاءة دون الحاجة لعرض جميع الأسئلة مما يقلل من زمن الاختبار ويزيد من موثوقية نتائجه. ويُنظر إلى هذا النوع على أنه يمثل مستقبل التقييمات التعليمية بفضل قدرته على مراعاة الفروق الفردية وتقديم صورة أدق عن مستوى الأداء (الخزي، 2016؛ الدامغ، 2006).

جدول 25 يوضح الفروق الجوهرية بين الاختبارات الإلكترونية المتكيفة وغير المتكيفة

البعد المقارن	الاختبارات الإلكترونية غير المتكيفة	الاختبارات الإلكترونية المتكيفة (CAT)
آلية تقديم البنود	تُعرض جميع البنود على جميع المتعلمين بالترتيب نفسه	تُعرض البنود تبعاً لإجابات المتعلم السابقة (تصاعدياً أو تنازلياً في الصعوبة)
مراعاة الفروق الفردية	محدودة، حيث يواجه جميع المتعلمين أسئلة متماثلة	عالية، إذ يتكيف الاختبار مع مستوى كل متعلم بشكل فردي
الدقة في تقدير مستوى المتعلم	أقل دقة، لاحتمال عدم ملائمة بعض البنود لمستوى المتعلم	عالية جداً، حيث يُقدَّر مستوى الأداء بسرعة وبأقل عدد من البنود
زمن الاختبار	عادة أطول، لأن جميع البنود تُقدَّم	أقصر نسبياً، إذ يكفي النظام بعدد قليل من البنود لتحقيق التقدير
العدالة	متساوٍ شكلياً (جميع المتعلمين يجيبون عن الأسئلة نفسها)	عدالة وظيفية، حيث يحصل كل متعلم على بنود متناسبة مع قدراته
سهولة الإعداد	سهولة الإعداد والتحويل من الاختبارات الورقية	أكثر تعقيداً، وتتطلب خبرة تقنية وإحصائية متقدمة، مثل نظرية الاستجابة للمفردة
التغذية الراجعة	عامة، ولا تعكس دائماً مستوى الأداء بدقة	دقيقة وشخصية، حيث تُبنى على مستوى المتعلم الفعلي

8-1- مميزات الاختبارات الإلكترونية:

أصبحت الاختبارات الإلكترونية من أبرز أساليب التقويم الحديثة، نظراً لما توفره من خصائص تجعلها أكثر كفاءة ومرونة مقارنة بالاختبارات التقليدية الورقية، ومن أهم مميزاتها ما يلي:

أ- دعم الوسائط المتعددة: تتيح إمكانية إدماج الصور والرموز والمقاطع الصوتية والمرئية في بنود الاختبار مما يزيد من وضوح الأسئلة ويعزز التفاعل.

- ب- تنوع أنماط الأسئلة: يمكن أن تتضمن أنماطاً متعددة كأسئلة الاختيار من متعدد، الصواب والخطأ، التوصليل، ملء الفراغ، أو حتى الأسئلة العملية التفاعلية.
- ج- تحديد زمن الأداء: تسمح بوضع وقت محدد لكل اختبار أو حتى لكل سؤال مما يساعد على ضبط عملية التقويم وتحقيق العدالة بين الممتحنين.
- د- سهولة التنقل: توفر بيئة إلكترونية مرنة تمكّن المتعلم من الانتقال بين الأسئلة للأمام أو العودة للمراجعة دون تعقيد.
- هـ- التصحيح الفوري: يتم تصحيح الإجابات إلكترونياً بشكل تلقائي مع إظهار النتيجة النهائية مباشرة بعد انتهاء الاختبار.
- و- التغذية الراجعة الفورية: تمنح المتعلم فرصة الاطلاع على إجاباته الصحيحة والخاطئة بعد الانتهاء بما يعزز عملية التعلم الذاتي.
- ز- إرسال النتائج إلكترونياً: يمكن إرسال نتائج المتعلمين عبر أنظمة التعليم الجامعي كنظام (ODUS) إلى المعلم أو الجهات المعنية بسرعة وكفاءة.
- ح- إمكانية الطباعة: تتيح للمتعلمين أو المعلمين طباعة النتائج في ملفات قابلة للمعالجة كبرنامج Excel، مما يسهل حفظها وتحليلها.
- ط- المرونة المكانية: لا تتطلب الوجود في مكان محدد، حيث يمكن إجراء الاختبار من أي مكان يتوفر فيه اتصال بالإنترنت.
- ي- الاستقلالية: لا تحتاج إلى مراقبة مباشرة من المعلم، إذ يقوم المتعلم بقراءة الأسئلة واختيار الإجابة وإرسالها إلكترونياً مع الالتزام بالوقت المحدد.
- ك- الكفاءة في إدارة البيانات: تسهّل عملية تخزين النتائج ومقارنتها وتحليلها إحصائياً، مما يرفع من قيمة الاختبارات الإلكترونية كأداة للبحث العلمي والتقويم المستمر (التميمي والشيخ، 2021، ص.173).

جدول 26 مقارنة بين الاختبارات الورقية والإلكترونية

الجانب	الاختبارات الورقية	الاختبارات الإلكترونية
الوسائط	نصوص وصور محدودة مطبوعة	وسائط متعددة (صور، صوت، فيديو، رموز)
تنوع الأسئلة	غالباً أسئلة تقليدية (اختيار من متعدد، مقالية)	أنماط متنوعة (اختياري، سحب وإفلات، تفاعلية)
التصحيح	يدوي، يستغرق وقتاً	إلكتروني وفوري، مع إمكانية معرفة النتيجة مباشرة

المراجعة بعد الاختبار	محدودة أو غير متاحة	يمكن مراجعة الأخطاء ومعرفة الإجابة الصحيحة
إرسال النتائج	يحتاج جهداً يدوياً	إرسال آلي للنتائج عبر النظام (ODUS) أو غيره.
الطباعة والتقارير	يحتاج وقتاً وتجهيزاً	إمكانية طباعة النتائج مباشرة عبر Excel
المرونة المكانية	يجب الحضور للمكان المحدد	يمكن إجراء الاختبار من أي مكان
الحاجة للمراقب	يحتاج مراقبة مباشرة	لا يحتاج لمراقب دائم، النظام يضبط الوقت
الزمن	يتطلب وقتاً أطول للتصحيح والإعلان عن النتائج	أسرع بكثير في الأداء والتصحيح

1-9- إيجابيات استخدام الاختبارات الإلكترونية:

تتميز الاختبارات الإلكترونية بعدد من الإيجابيات التي تعود بالنفع على كلٍّ من المعلم والمتعلم، ويمكن إجمال أبرزها فيما يلي:

- أ- تقديم تغذية راجعة فعّالة: تتيح الاختبارات الإلكترونية للمعلم تزويد المتعلمين بتغذية راجعة فورية ودقيقة حول إجاباتهم، مما يساعدهم على تصحيح أخطائهم وتطوير استراتيجيات تعلمهم.
- ب- تشخيص مستوى المتعلمين: تمكّن المعلم من تحديد مستوى التحصيل الأكاديمي للطلاب بدقة، والكشف عن مواطن القوة والضعف لديهم بشكل فردي أو جماعي.
- ج- تحليل جودة الفقرات: تساعد المعلم على التعرف إلى الفقرات الغامضة أو غير المفهومة بالنسبة للمتعلمين، مما يساهم في تحسين جودة الاختبار مستقبلاً.
- د- تخفيف الجهد والوقت: تسهّل عملية اختبار أعداد كبيرة من المتعلمين في وقت قصير، مقارنةً بالاختبارات التقليدية التي تتطلب وقتاً وجهداً أكبر في الإعداد والتصحيح.
- هـ- جاذبية أكبر للمتعلمين: غالباً ما يجد الطلاب الاختبارات الإلكترونية ممتعة ومشوقة، حيث توفر لهم إحساساً بالتحكم والحرية، وتُشبه في طبيعتها الألعاب أو المحاكاة التفاعلية.
- و- السرعة والسهولة في الاستخدام: يسهل إعدادها وتنفيذها من قبل المعلم، كما يسهل على الطالب التعامل معها دون تعقيد.
- ز- تحسين الأداء الأكاديمي: أظهرت بعض الدراسات أن استخدام الاختبارات الإلكترونية يساهم في رفع مستوى التحصيل الدراسي لدى الطلاب، نظراً لما توفره من تفاعل فوري وبيئة محفزة.
- ح- زيادة الدافعية للتعلم: تساهم طبيعتها التفاعلية والمرنة في رفع دافعية الطلاب للمثابرة وتحسين أدائهم الأكاديمي.

ط. المرونة في الزمان والمكان: تمنح المتعلمين فرصة إجراء الاختبارات في أي وقت وأي مكان، مما يدعم التعلم الذاتي والتعلم عن بُعد.

ي- تصحيح إلكتروني فوري: تقلل من الجهد المبذول من قبل المعلمين في التصحيح اليدوي حيث تُعطي النتائج بشكل تلقائي وسريع مما يضمن الدقة والموضوعية (المندلأوي، 2025، ص156).

10-1- معايير الجودة في تصميم الاختبارات الإلكترونية:

تُعد جودة تصميم الاختبارات الإلكترونية شرطاً أساسياً لضمان فعاليتها ودقتها في قياس أهداف التعلم، إذ لا تقتصر أهميتها على الميدان المدرسي فقط بل تمتد لتشمل مختلف المراحل التعليمية والتخصصات الأكاديمية والتدريبية، ويمكن تصنيف هذه المعايير إلى نوعين رئيسيين:

أ- معايير تتعلق بخصائص الاختبار:

ينبغي أن يتضمن الاختبار الإلكتروني واجهة منظمة وواضحة، تبدأ عادة بصفحة مخصصة لإدخال البيانات الأساسية للممتحنين كالاسم، المرحلة الدراسية أو الأكاديمية، التخصص، المادة، البريد الإلكتروني، الدولة، وغيرها من البيانات التعريفية الضرورية. كما يُستحسن أن تكون واجهة الاستخدام سهلة التصفح ومتوافقة مع مختلف الأجهزة (حاسوب، جهاز لوحي، هاتف ذكي) بما يحقق تجربة مستخدم مريحة وفعّالة.

ب- معايير تتعلق بمحتوى الاختبار:

يجب أن يتميز محتوى الاختبار بالوضوح والدقة بحيث تصاغ الفقرات بلغة سهلة ومباشرة خالية من الغموض والالتباس. كما ينبغي أن تكون الفقرات مرتبطة بالأهداف التعليمية الموضوعية مسبقاً وأن تتسم بالإيجاز دون الإخلال بعمق المعنى مع مراعاة التدرج المنطقي في مستوى الصعوبة. إضافة إلى ذلك يُراعى التنوع في أنماط الأسئلة (اختيار من متعدد، صح أو خطأ، أسئلة مقالية قصيرة، أسئلة تفاعلية) بما يعزز قياس مختلف المستويات المعرفية لدى المتعلمين.

ويمكن إضافة معايير أخرى داعمة مثل:

- **الأمان والسرية:** أي حماية بيانات الممتحنين ومنع الغش؛
- **الشفافية:** أي إتاحة معرفة نظام التصحيح والزمن المخصص؛
- **العدالة:** أي مراعاة الفروق الفردية وتوفير فرص متكافئة لجميع المتعلمين (التميمي والشيخ، 2021، ص176).

11-1- مراحل تصميم وبناء الاختبارات الإلكترونية:

تمر عملية تصميم وبناء الاختبارات الإلكترونية بعدة مراحل متتابعة تهدف إلى ضمان جودتها وفعاليتها، وقد أشار خليل (2017، ص 5-26) إلى هذه المراحل على النحو الآتي:

أ- **مرحلة التحليل:** في هذه المرحلة يتم تحديد الهدف الرئيس من بناء الاختبار وتحليل خصائص الفئة المستهدفة، إضافة إلى تحليل المادة التعليمية التي سيبنى عليها محتوى الاختبار، كما يتم تحليل الواقع التكنولوجي للمؤسسة التعليمية وتحديد الأجهزة والبرامج اللازمة للتنفيذ.

ب- **مرحلة التصميم:** وتشمل صياغة أسئلة الاختبار وتحديد تعليماته وضبط الزمن المخصص واختيار أنماط الأسئلة والاستجابات المناسبة وتضمين الوسائط المتعددة مع تحديد نوع التغذية الراجعة وطريقة التصحيح.

ج- **مرحلة الإنتاج:** يتم في هذه المرحلة اختيار البرامج البرمجية لتأليف الاختبار وإجراء التجريب الأولي، ثم عرضه على المحكمين المتخصصين للتحكيم وتطويره في ضوء ملاحظاتهم مع توثيق النسخة النهائية من البرمجية.

د- **مرحلة النشر الإلكتروني:** يتم خلالها نشر الاختبار عبر الإنترنت أو من خلال أقراص وأسطوانات رقمية بما يتيح وصوله إلى المتعلمين في أماكن تواجدهم.

هـ- **مرحلة التطبيق:** تتضمن تجربة الاختبار فعلياً على عينة من المتعلمين وجمع بيانات الأداء ثم إعلان النتائج إلكترونياً.

و- **مرحلة التقويم:** وهي المرحلة النهائية التي يُقيم فيها مدى ملاءمة البيئة الإلكترونية للاختبار وصلاحيته نقله وتوصيله مع التأكد من سرية وأمن الاختبار.

12-1- التغذية الراجعة في الاختبارات الإلكترونية:

تعد التغذية الراجعة من أهم المزايا التي تتفرد بها الاختبارات الإلكترونية مقارنة بالاختبارات التقليدية، إذ تمنح المتعلم معلومات آنية ومباشرة حول أدائه مما يساهم في تعزيز عملية التعلم الذاتي وتحفيز الدافعية الداخلية، ويرى زيتون (2005) أن أشكال التغذية الراجعة التي تقدمها الاختبارات الإلكترونية يمكن تلخيصها فيما يلي:

أ- **عرض الدرجة النهائية:** يحصل الطالب مباشرة على نتيجته، مما يعزز الشفافية ويساعده على تقييم مستوى إنجازه بشكل فوري.

ب- **توضيح جوانب القوة والضعف:** لا يقتصر النظام على عرض الدرجة فقط، بل يحدد المجالات التي أتقنها الطالب وتلك التي تحتاج إلى مزيد من المراجعة.

ج- تصحيح الأخطاء: يتم إبراز الإجابات غير الصحيحة، الأمر الذي يساعد المتعلم على تجنب تكرارها مستقبلاً.

د- إعطاء إرشادات للتصحيح: يقدم النظام تغذية تفسيرية أو توجيهية توضح سبب الخطأ وتقدم أساليب للوصول إلى الإجابة الصحيحة، وهو ما يجعل التغذية الراجعة تعليمية بامتياز.

هـ- تحديد المدة الزمنية للاختبار: يعد عنصر الوقت أحد أشكال التغذية الراجعة غير المباشرة، إذ يساعد المتعلم على تنمية مهارة إدارة الزمن وتحسين سرعته في إنجاز المهام.

و- تعزيز الدافعية والتعلم الذاتي: بفضل هذه الأنواع من التغذية، يصبح الطالب أكثر تحفزاً للتعلم الذاتي والمراجعة المستمرة، حيث يشعر بملكية أكبر لعملية التعلم.

جدول 27 يوضح أنواع التغذية الراجعة في الاختبارات الإلكترونية وفوائدها التربوية

النوع	الوصف	الفائدة التربوية
عرض الدرجة النهائية	إظهار النتيجة مباشرة بعد إنهاء الاختبار	تعزيز الشفافية وتمكين الطالب من تقييم إنجازاته فوراً
تحديد جوانب القوة والضعف	إبراز ما أتقنه الطالب وما يحتاج إلى تحسين	مساعدة الطالب على التخطيط لتحسين تعلمه
تصحيح الأخطاء	إظهار الإجابات الخاطئة	تجنب تكرار الخطأ وتثبيت التعلم الصحيح
إرشادات للتصحيح	تقديم ملاحظات وتفسيرات حول الإجابة الصحيحة	تعميق الفهم وتنمية التفكير النقدي
تحديد المدة الزمنية	ضبط زمن إنجاز الاختبار	تنمية مهارة إدارة الوقت وتحسين سرعة الأداء
تعزيز الدافعية والتعلم الذاتي	زيادة تحفيز الطالب على متابعة التعلم	تنمية الاستقلالية وتحسين أساليب التعلم الذاتي

13-1- طرق تقديم الاختبارات الإلكترونية:

تتعدد طرق تقديم الاختبارات الإلكترونية وفقاً لآلية الوصول إليها، ويمكن التمييز بين طريقتين أساسيتين:

أ- الاختبارات عبر الإنترنت (Online e-assessment):

في هذه الطريقة يُرفع الاختبار على شبكة الإنترنت، ويقوم المتعلم بالدخول إلى الموقع أو المنصة المخصصة لأداء الاختبار، ويُعد الاتصال بالإنترنت شرطاً أساسياً لإجرائه حيث لا يمكن للطلاب الدخول للاختبار أو الحصول على نتائجه بدون (Romiszowski, 2004). وتمتاز هذه الطريقة

بالسرعة في نشر الاختبارات وإمكانية مراقبة الأداء اللحظي للطلاب وتطبيق الخصائص كالتوقيت الزمني التلقائي أو إغلاق الاختبار عند انتهاء الفترة المحددة.

ب- الاختبارات دون اتصال بالإنترنت (Offline e-assessment):

تعتمد هذه الطريقة على تنزيل برامج أو تطبيقات الاختبار مسبقًا على الأجهزة الذكية كالهواتف المحمولة، الحواسيب، أو الأجهزة اللوحية، بحيث يتمكن الطالب من أداء الاختبار دون الحاجة إلى الاتصال بالإنترنت (Naidu, 2006; Algahtani, 2011). وتمتاز هذه الطريقة بأنها مناسبة أكثر للاختبارات الطويلة أو المعقدة، حيث تضمن استمرار الطالب في الإجابة دون انقطاع بسبب مشاكل الاتصال وتوفر له بيئة أكثر استقرارًا مع إمكانية رفع النتائج لاحقًا عند توفر الإنترنت.

تشير الدراسات الحديثة إلى أن بعض المؤسسات التعليمية باتت تدمج بين الطريقتين (Hybrid e-assessment) بحيث يمكن للطالب بدء الاختبار عبر الإنترنت ثم استكماله أو مراجعته في وضع غير متصل، مما يعزز المرونة ويقلل من أثر الأعطال التقنية.

كما أن انتشار تقنيات الحوسبة السحابية (Cloud computing) أسهم في تطوير بيئات اختبارات إلكترونية أكثر أمانًا وموثوقية سواء عبر الإنترنت أو بدونه.

14-1- سلبيات الاختبارات الإلكترونية:

على الرغم من المزايا المتعددة للاختبارات الإلكترونية، إلا أن استخدامها لا يخلو من بعض السلبيات والتحديات التي قد تؤثر في مصداقيتها وفعاليتها، ومن أبرزها:

- **ضعف كفاءة المتعلمين في التعامل مع التكنولوجيا:** حيث يواجه بعض المتعلمين صعوبات في استخدام الحاسوب أو الإنترنت نتيجة قلة خبرتهم أو افتقارهم إلى المهارات الرقمية الأساسية، مما قد يعرقل أداءهم في الاختبار.

- **الحاجة إلى التدريب المسبق:** يتطلب تطبيق الاختبارات الإلكترونية تخصيص وقت وجهد لتدريب المتعلمين على كيفية الدخول إلى المنصة والتعامل مع أنماط الأسئلة الإلكترونية وآليات الإجابة وهو ما قد يشكل عبئًا إضافيًا على المؤسسات التعليمية.

- **مشكلات البنية التحتية:** تعاني بعض البيئات التعليمية من ضعف شبكات الإنترنت أو انقطاعها المتكرر بالإضافة إلى انقطاع التيار الكهربائي وهو ما قد يؤدي إلى تعطل الاختبارات وضياع الجهد المبذول من قبل المتعلم.

- **المشكلات التقنية الطارئة:** مثل توقف البرنامج المفاجئ أو تعطل الأجهزة أثناء أداء الاختبار، مما قد يحرم الطالب من استكماله أو يقلل من عدالة التقييم.

- **التأثير على صدق النتائج:** قد تتداخل مهارات غير أكاديمية في تحديد مستوى أداء الطالب كمهارات استخدام الأجهزة أو إتقان البرمجيات، مما يجعل الدرجة المحصلة غير معبرة بدقة عن المستوى الحقيقي للمعرفة أو المهارة المستهدفة.

- **زيادة فرص الغش والتحايل:** حيث تتيح بعض بيانات الاختبار الإلكتروني للمتعلمين إمكانية استخدام مصادر خارجية أو الاستعانة بآخرين، مما قد يؤثر سلبًا على نزاهة النتائج.

- **محدودية أنماط الأسئلة:** رغم التطور الحاصل إلا أن بعض الاختبارات الإلكترونية ما تزال تركز على أسئلة الاختيار من متعدد أو الأسئلة المغلقة، مما يحد من قدرتها على قياس المهارات العليا كال تفكير الناقد والإبداع (المندلاوي، 2025، ص156).

يمكن أن نختم بالقول إن الاختبارات الإلكترونية تمثل نقلة نوعية في أساليب التقويم الحديثة، إذ تجمع بين السرعة والمرونة وتواكب التطور التكنولوجي، غير أنها ما تزال تواجه بعض التحديات التقنية والتربوية التي تستدعي التخطيط الجيد والتدريب المسبق لتفاديها، وبالتالي فإن نجاحها يعتمد على تحقيق التوازن بين مزاياها وقيودها وتوظيفها بما يخدم العملية التعليمية بأعلى قدر من الكفاءة والعدالة.

2- الاختبارات التكيفية المحوسبة (Computerized Adaptive Testing - CAT):

يُعتبر الاختبار التكيفي المحوسب (CAT) من أبرز التطبيقات العملية لنظرية الاستجابة للمفردة (Item Response Theory - IRT) حيث يتيح تقدير مستوى قدرة الفرد بشكل ديناميكي أثناء عملية الاختبار. في هذا النوع من الاختبارات تُقاس قدرة الممتحن أولاً من خلال الإجابات على البنود الأولية ثم تُستخدم هذه التقديرات لاختيار البنود التالية التي تتناسب مع مستوى قدرته بشكل دقيق، مما يضمن تحدياً معرفياً مناسباً لكل فرد.

يمكن أن تحتوي الاختبارات التكيفية على عدد ثابت من البنود أو قد تنتهي بمجرد تحقق قاعدة إيقاف محددة بمجرد الوصول إلى تقدير موثوق للقدرة أو السمة المقاسة، مما يقلل من طول الاختبار ووقت الأداء بشكل ملحوظ، كما يسهم هذا الأسلوب في تقليل الإحباط والتوتر الذي قد يواجهه الممتحنون في الاختبارات الورقية التقليدية، خاصة عند مواجهة بنود أعلى أو أقل كثيراً من مستوى قدراتهم.

لقد اعتمدت مؤسسات كبرى كهيئة الاختبارات التعليمية (ETS) ووزارة الدفاع الأمريكية، هذا النوع من الاختبارات منذ سنوات نظراً لما يقدمه من كفاءة عالية، دقة في القياس، ومرونة في التطبيق عبر مجالات متنوعة. ومع ذلك يظل هناك تحديات تتعلق بأمن الاختبار، تكاليف التطوير، وصعوبة

السماح للممتحنين بمراجعة أو تعديل إجاباتهم، وهو ما يفرض وضع استراتيجيات إدارة دقيقة لضمان نزاهة وفعالية الاختبار (Urbina, 2004, p.101).

2-1- الاختبارات التكيفية المحوسبة في القياس والتقويم التربوي:

تُعد الاختبارات التكيفية المحوسبة (Computerized Adaptive Testing - CAT) أحد التطبيقات العملية لنظرية الاستجابة للمفردة (Item Response Theory - IRT) إذ تختلف عن الاختبارات التقليدية التي تقدم نفس مجموعة الأسئلة لكل الطلاب. في هذا النوع من الاختبارات يتم اختيار كل سؤال من بنك أسئلة واسع بناءً على تقدير لحظي لقدرة الطالب أثناء أداء الاختبار بحيث تتكيف صعوبة البنود مع مستوى المتعلم مما يزيد من دقة وموضوعية القياس.

ووفقاً لـ (Eignor, Way, Stoking, & Steffen, 1993) تعتمد هذه الاختبارات على اختيار متسلسل للبنود بحيث يكون كل سؤال ملائماً للتقدير الحالي لقدرة الطالب وفقاً لنموذج الاستجابة للمفردة الذي يوضح العلاقة بين خصائص السؤال ومستوى أداء الفرد. ولتحقيق دقة القياس يجب أن يكون هناك توافق بين صعوبة الاختبار ومستوى قدرة الطالب، إذ قد تؤدي البنود شديدة الصعوبة أو السهولة إلى نتائج غير دقيقة أو مضللة.

يمثل هذا النهج في القياس التربوي وسيلة للاستفادة القصوى من كل سؤال وتقليل زمن الاختبار مع الحفاظ على موثوقية وجودة التقييم (علام، 2013، ص.296). كما يتيح للمعلمين والمقيمين تقديم تجربة اختبارية أكثر تخصيصاً للمتعلمين مع تقليل الإحباط وزيادة الكفاءة في عملية التقييم.

2-3- الحاجة إلى الاختبارات المواءمة الفردية:

برزت فكرة الاختبارات المواءمة الفردية استجابةً للتحديات التي تعاني منها الاختبارات الجماعية التقليدية، فهذه الأخيرة تُبنى عادة لتشمل مفردات تغطي مدى واسعاً نسبياً من القدرات بغرض تحقيق عدالة في القياس لجميع المختبرين. ولتحقيق ذلك ينبغي أن يحتوي الاختبار على بنود متفاوتة في مستوى الصعوبة؛ بحيث يتضمن بنوداً سهلة لذوي الكفاءة المنخفضة، وبنوداً متوسطة للمستوى المتوسط، وبنوداً صعبة لذوي الكفاءة العالية.

في حال افتقار الاختبار إلى البنود الصعبة يصبح من الصعب التمييز بين الأفراد ذوي الكفاءة المرتفعة الذين ينجحون بسهولة في البنود المتوسطة والسهلة. وعلى العكس إذا لم يتضمن الاختبار بنوداً شديدة السهولة، فلن يتمكن من التمييز بين الأفراد الأقل كفاءة الذين قد يعجزون عن الإجابة على البنود متوسطة الصعوبة، ومن ثم فإن التوزيع المتوازن للبنود عبر مستويات الصعوبة يُعد شرطاً أساسياً لقياس فعال يراعي تنوع قدرات مجتمع المختبرين.

غير أن الواقع يشير إلى أن معظم الاختبارات الجماعية تميل إلى الاعتماد على مفردات متوسطة الصعوبة مع وجود عدد محدود من البنود شديدة السهولة أو شديدة الصعوبة الأمر الذي يُضعف قدرتها على التمييز بدقة بين جميع مستويات الأداء (ديفيليس، 2021، ص.295).

2-4- مميزات الاختبارات التكيفية المحوسبة:

أظهرت الأدبيات المتخصصة أن الاختبارات التكيفية المحوسبة تتفوق بوضوح على الاختبارات التقليدية من حيث الكفاءة والفاعلية إذ تجمع بين الدقة العلمية والمرونة التكنولوجية . ويمكن تلخيص أبرز هذه المميزات فيما يلي:

- **زيادة دقة القياس:** إذ تسمح خوارزميات التكيف بتقدير أكثر موضوعية لقدرات الفرد مقارنة بالاختبارات ذات الطول الثابت.

- **تخصيص الاستجابة لمستوى الفرد:** حيث تُعرض البنود بما يتناسب مع معدل أداء الممتحن لحظة بلحظة، مما يجعل تجربة الاختبار أكثر عدالة.

- **تحفيز الأداء الأمثل:** نتيجة توافق صعوبة البنود مع قدرات المختبر، مما يقلل من الإحباط أو الملل.

- **التغذية الراجعة الفورية:** إذ يحصل الفرد على نتائج أو مؤشرات مباشرة عن أدائه.

- **المراقبة المستمرة للخصائص السيكومترية للبنود:** مما يضمن تحديث جودة المفردات وملاءمتها باستمرار.

- **تنوع صيغ البنود الاختبارية:** حيث يمكن أن تشمل أسئلة تقليدية (اختيار من متعدد) أو تفاعلية تعتمد على الوسائط المتعددة.

- **التسجيل الفوري للدرجات وتخزينها:** وهو ما يسهل عمليات المعالجة الإحصائية وتحليل النتائج لاحقاً.

- **الحد من مشكلات سرية الامتحانات:** نتيجة اختلاف مسار الأسئلة بين الأفراد بحيث يصبح من الصعب تداول نسخة موحدة أو تسريبها (علام، 2013، ص.300).

2-5- تصميم نظام الاختبارات التكيفية المحوسبة:

يتطلب بناء نظام فعال للاختبارات التكيفية المحوسبة (CAT) توفر مجموعة من المكونات الأساسية التي تضمن دقة القياس ومرونته، ويمكن تلخيصها فيما يلي:

أ- **مجموعة واسعة من المفردات الاختبارية** تغطي مستويات متعددة من الصعوبة.

ب- **أحد نماذج الاستجابة للمفردة (IRT Models)** كأساس نظري لتقدير العلاقة بين خصائص البنود وأداء الأفراد.

ج- بنك أسئلة مُعايير مسبقًا باستخدام النموذج المختار، بحيث تُخزن مع كل مفردة معلماتها، مثل: (الصعوبة والتمييز).

د- نظام حاسوبي مخصص لاستدعاء المفردات بسرعة من بنك الأسئلة وعرضها على شاشة الممتحن.

هـ- إجراءات بدء الاختبار، مثل آلية اختيار السؤال الافتتاحي.

و- قواعد اختيار المفردة التالية استنادًا إلى أداء الفرد في المفردة السابقة.

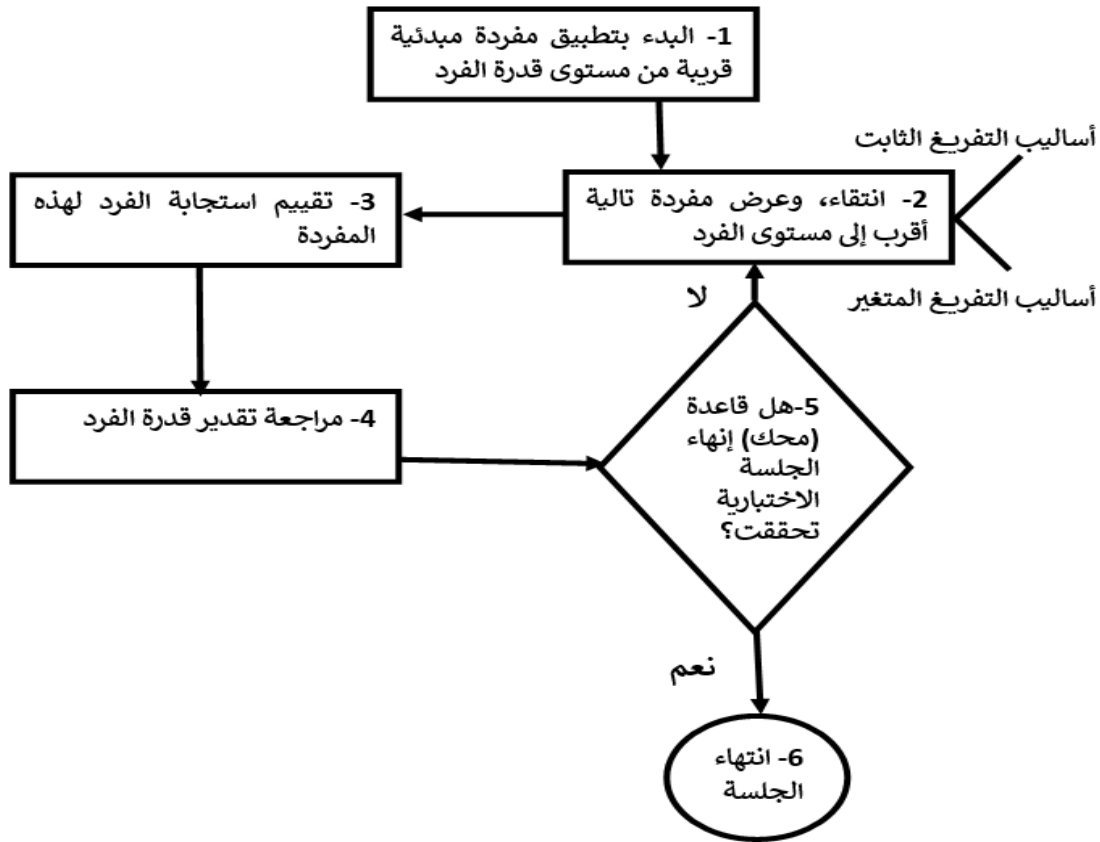
بمجرد أن يجيب الممتحن على السؤال الأول، يبدأ برنامج الاختبار التكيفي في العمل وفق خوارزميات متقدمة، من أبرزها:

1- طريقة المعلومات القصوى (Maximum Information Method): حيث يُنتقى السؤال الذي يوفر أكبر قدر من المعلومات عن قدرة الممتحن.

2- الطرائق المستندة إلى نظرية بايز (Bayesian Methods): التي تأخذ في الحسبان احتمالية مستويات مختلفة من القدرة بناءً على استجابات الفرد.

وبناءً على نمط الاستجابات السابقة (صحيحة أو خاطئة) يقوم الحاسوب بانتقاء السؤال الأنسب من بنك الأسئلة مسترشدًا بالمعلومات المخزنة مع كل مفردة. ولذا يصبح من الضروري إعادة تقدير مستوى قدرة الفرد مباشرة بعد كل استجابة لضمان دقة مستمرة في تحديد البنود اللاحقة. ومن المعتاد أن يتم توضيح هذه الخطوات في شكل مخططي يعرض تدفق العملية بدءًا من السؤال الأول وصولًا إلى قرار الإيقاف بعد الوصول إلى تقدير مستقر لقدرة المختبر.

شكل 29 يوضح الخطوات التسلسلية المنطقية لتقدير قدرة الفرد باستخدام نظام الاختبارات التكيفية المحوسبة (CAT).



ز- قاعدة تقدير القدرة، إذ يُعاد تقدير درجة قدرة الفرد بعد كل استجابة صحيحة أو خاطئة اعتماداً على المعلومات المخزنة للبنود ونموذج الاستجابة للمفردة.

ح- قاعدة إنهاء الجلسة الاختبارية، حيث يتوقف الاختبار عند تحقق أحد المعايير المحددة مسبقاً، كالوصول إلى مستوى ثابت من الدقة في تقدير القدرة أو استيفاء حد أقصى لعدد البنود أو انتهاء الزمن المخصص للجلسة (علام، 2013، ص.310).

وبهذا التصميم المتكامل يتحقق التكيف اللحظي المستمر بين مستوى الممتحن ومحتوى الاختبار، مما يُسهم في توفير تجربة تقييم دقيقة وفعّالة مع تقليل طول الاختبار والوقت اللازم لإنجازه.

2-6- التحديات التي تواجه استخدام الاختبارات التكيفية المحوسبة:

على الرغم من المزايا المتعددة للاختبارات التكيفية المحوسبة، إلا أن هناك العديد من المشكلات التي يعكف خبراء القياس على دراستها وبحثها في محاولة للتغلب عليها، ولعل أهمها:

أ- يعتمد تصميم وبناء الاختبارات التكيفية المحوسبة اعتماداً أساسياً على نظرية الاستجابة للمفردة الاختبارية؛ لذلك فإن جميع المشكلات المتضمنة في استخدام هذه النظرية تعد أيضاً مشكلات لهذه الاختبارات.

ب- على الرغم من تأكيد خبراء القياس أن هذه الاختبارات تميز بدرجة عالية ومتسقة بين الأفراد على متصل القدرة المقاسة بعدد من المفردات أقل مما تحتوي عليه الاختبارات الخطية التقليدية، إلا أنه لا يوجد اتفاق حول أفضل إستراتيجيات انتقاء مفردات الاختبارات التكيفية المحوسبة.

ج- تتطلب إجراءات تطبيق الاختبارات التكيفية المحوسبة إنشاء بنك أسئلة يشتمل على عدد كبير من المفردات يبلغ عدة مئات تختزن في الحاسوب لتيسير انتقاء المفردات المناسبة لقدرة الأفراد من ذوي المستويات المختلفة، والتوصل إلى تمييزات دقيقة بينهم في القدرة المقاسة. وبناء هذه المفردات وتقدير بارامترات استخدامها باستخدام أحد نماذج الاستجابة للمفردة يتطلب خبرة فنية عالية وجهداً كبيراً وتمويلًا كافياً. كما يتطلب توافر حواسيب وتقنيات متقدمة وبرمجيات حديثة لتطبيق الاختبارات وتصحيحها وتقديم نتائجها الفورية للأفراد أثناء الجلسات الاختبارية وتحليل هذه النتائج باستخدام الحاسوب. وهذه الخبرات والتقنيات ربما لا تكون متوافرة في كثير من الدول النامية.

د- تحتاج المشكلات المتعلقة بتأثير سياق عرض المفردات وترتيبها وتوازن محتواها إلى مزيد من الدراسات والبحوث التطويرية.

هـ- عدم توافر البنى التحتية التي يستند إليها في تنفيذ برامج الاختبارات التكيفية المحوسبة على نطاق واسع فهذه البنى تحتاج إلى كلفة باهظة (علام، 2013، ص.322).

و- صعوبة تحقيق العدالة وتكافؤ الفرص بين المختبرين، إذ قد تؤدي طريقة التكيف مع استجابات الأفراد إلى إثارة شكوك حول مدى إنصاف المقارنات بين من خضعوا لمفردات مختلفة.

ز- التحديات الأخلاقية المرتبطة بسرية بنك الأسئلة وحمايته من التسريب أو القرصنة، خصوصاً أن أي اختراق قد يؤدي إلى انهيار صلاحية الاختبار بالكامل.

ح- مقاومة بعض المؤسسات التربوية أو الأكاديمية لاعتماد هذه الاختبارات نتيجة ضعف الوعي بها أو الخوف من صعوبة تطبيقها عملياً مقارنة بالاختبارات الورقية التقليدية.

ط- الحاجة إلى تدريب المختصين (معلمين، مقيمين، مطوري اختبارات) على بناء البنوك الاختبارية وتفسير نتائج هذه الاختبارات، وهو ما قد يستغرق وقتاً وجهداً وميزانيات إضافية.

2-7- الخطأ المعياري للقياس (SEM) في الاختبارات التكيفية المحوسبة:

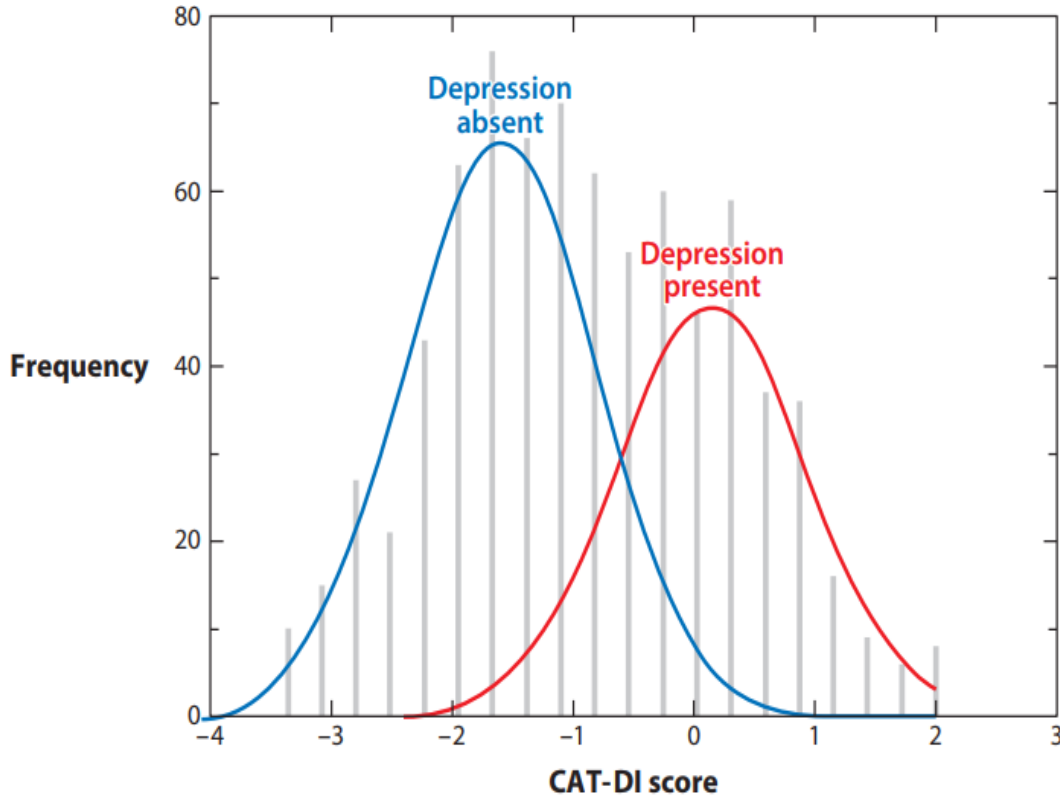
تُبنى معظم أنظمة الاختبارات التكيفية المحوسبة على أسس نظرية الاستجابة للمفردة (IRT)، إذ تُستخدم معلمات البنود (الصعوبة، التمييز، واحتمال التخمين) التي جرى تقديرها مسبقاً عند معايرة بنك الأسئلة لتكييف عملية اختيار البنود بما يتناسب مع استجابات الممتحنين. ويسمح هذا النهج بتحقيق قياس أكثر دقة للقدرة الكامنة (θ) مقارنة بالاختبارات التقليدية ذات الطول الثابت.

عند بدء الاختبار يتم تحديد نقطة أولية لتقدير القدرة، ومن أبسط الأساليب المتبعة افتراض قيمة متوسطة ($\theta = 0.0$) تمثل مستوى الأداء العام للمجموعة المرجعية. يبدأ الاختبار عادةً ببند متوسط الصعوبة ثم تُعدل صعوبة البنود اللاحقة وفق استجابات الممتحن فإذا كانت الإجابة صحيحة تُقدم مفردة أكثر صعوبة وإذا كانت خاطئة تُعرض مفردة أسهل. تستمر هذه العملية التكيفية إلى أن يصل التقدير إلى مستوى دقة مقبول، يتمثل غالباً في انخفاض قيمة الخطأ المعياري للقياس (SEM) إلى الحد المسموح به مسبقاً.

يُستخدم الخطأ المعياري للقياس كمعيار رئيسي لاتخاذ قرار إنهاء الاختبار، إلى جانب معايير أخرى كعدد البنود التي تم تقديمها أو نفاذ بنك الأسئلة المناسب للممتحن، وكلما انخفض الخطأ المعياري للقياس (SEM) ازدادت الثقة في أن الدرجة المقدرة (θ) قريبة من القدرة الحقيقية للفرد. وقد بينت دراسات سابقة (Rudner, 2000) أن الاختبارات التكيفية المحوسبة المبنية على نظرية الاختبار الكلاسيكية يمكن أن تكون فعالة في تصنيف الأفراد، إلا أن التطبيقات التي تهدف إلى تقدير دقيق للقدرة الفردية تعتمد بشكل أفضل على نظرية الاستجابة للمفردة لما توفره من خصائص سيكومترية عالية.

وفي إطار التطبيقات العملية، طور Gibbons وآخرون (2013) أول أداة فحص تشخيصية تكيفية محوسبة لاضطراب الاكتئاب بهدف تقليل العبء على المريض والأخصائي النفسي مع تعزيز دقة التشخيص وفق معايير DSM-IV الخاصة باضطراب الاكتئاب الشديد (MDD). وقد استُخدم في هذا التطبيق بنك أسئلة مكون من 88 بنداً تم اشتقاقها من 73 مقياساً للاكتئاب، حيث طبق على عينة متنوعة شملت 656 فرداً من المرضى الذين يعانون من درجات مختلفة من الاكتئاب وأفراداً من المجموعات الضابطة. وقد مكّن هذا الاختبار من تصنيف الأفراد بشكل تكيفي إلى فئتين أساسيتين: فئة وجود الاكتئاب (Depression present)، وفئة غياب الاكتئاب (Depression absent)، مع تحقيق مستويات عالية من الحساسية والخصوصية في التشخيص السريري، كما يمثلها الشكل الآتي:

شكل رقم (30) يمثل توزيع التكرارات المشاهدة لتقدير شدة الاكتئاب باستخدام الاختبار التكيفي المحوسب (CAT) لقائمة الاكتئاب (DI)، بحيث يمثل المكون السفلي عدم وجود الاكتئاب (المنحنى الأزرق)؛ ويمثل المكون العالي مستويات الشدة المرتبطة بالاكتئاب السريري (المنحنى الأحمر). (Gibbons, 2016, p96).



على الرغم من الانتشار الواسع للاختبارات التكيفية المحوسبة (CAT) المعتمدة على نظرية الاستجابة للمفردة الاختبارية (IRT) في مجال القياس التربوي، إلا أن استخدامها في مجال قياس الصحة العقلية ظل محدوداً نسبياً، ويُعزى ذلك إلى سببين رئيسيين:

أولاً: ندرة بنوك الأسئلة الخاصة بمقاييس الصحة العقلية، إذ إن بناء مثل هذه البنوك يتطلب جهوداً طويلة ومعقدة تشمل عمليات تطوير البنود والتحقق من صدقها وثباتها ومعايرتها وفق نماذج الاستجابة للمفردة الاختبارية (IRT) المختلفة.

ثانياً: أن اضطرابات الصحة العقلية كالاكتئاب والقلق تنسم بطبيعتها بالتعدد البُعدي، في حين أن الاختبارات التكيفية المحوسبة صُممت تقليدياً للعمل بكفاءة أكبر في المجالات أحادية البعد كما هو الحال في اختبارات التحصيل الدراسي في مادة الرياضيات أو القراءة مثلاً.

وقد أشار Gibbons (2016, p.87) إلى أن هذه الخصائص تجعل من عملية تطوير وتطبيق الاختبارات التكيفية المحوسبة (CAT) في الصحة العقلية أكثر تعقيداً، وتتطلب تعديلات منهجية وتقنية تتجاوز ما هو مألوف في المجالات التربوية.

2-8- دالة المعلومات (Information Function) في الاختبارات التكيفية المحوسبة:

تُعد دالة المعلومات أحد المفاهيم الجوهرية في نظرية الاستجابة للمفردة (IRT)، وتشكل أساساً مهماً في تصميم وتطوير الاختبارات التكيفية المحوسبة، فهي تحدد مقدار المعلومات السيكومترية التي توفرها المفردة الاختبارية عند مستويات مختلفة من القدرة الكامنة، وبالتالي تُعد مؤشراً مباشراً على دقة تقدير مستوى الفرد المختبر، وتنقسم دالة المعلومات إلى:

1- دالة معلومات المفردة (Item Information Function – IIF): تمثل المنحنى الخاص بمفردة معينة ضمن نموذج الاستجابة للمفردة، سواء في حالة المفردات ثنائية الاستجابة (Dichotomous) أو متعددة التدرج (Polytomous). ويمكن تحويل هذه المنحنيات إلى منحنيات معلومات المفردة (Item Information Curve – IIC) والتي توضح بدقة كمية المعلومات التي توفرها المفردة عند كل مستوى من مستويات القدرة.

2- دالة معلومات الاختبار (Test Information Function – TIF): وهي مجموع معلومات جميع المفردات المكونة للاختبار، وتُظهر مدى دقة الاختبار ككل عند تقدير القدرة الكامنة عبر مستويات مختلفة.

في الاختبارات التكيفية المحوسبة (CAT) تلعب دالة المعلومات دوراً محورياً في آليتي اختيار المفردات وتقدير القدرة، إذ يتم في كل خطوة اختيار المفردة التي تقدم أعلى مقدار من المعلومات عند مستوى القدرة الحالي للمختبر، مما يضمن زيادة دقة التقدير مع طرح أقل عدد ممكن من المفردات، وكلما ارتفعت قيمة دالة المعلومات للمفردة أو للاختبار، انخفض الخطأ المعياري للقياس (SEM) وزادت دقة النتائج.

وعلى خلاف النظرية الكلاسيكية للقياس التي تفترض أن الثبات ثابت عبر جميع مستويات القدرة، تُظهر دالة المعلومات في نظرية الاستجابة للمفردة (IRT) أن دقة القياس تتغير بحسب مستوى القدرة، وهذا ما يجعلها أداة حيوية عند تصميم بنوك الأسئلة التي تُستخدم في بناء الاختبارات التكيفية المحوسبة، حيث يتم انتقاء المفردات ذات المعلومات الأعلى عند مستويات مختلفة من القدرة لضمان كفاءة القياس (Thomas, 2019, p.06).

3- القياس التربوي النفسي في عصر الآلة الذكية:

3-1- تاريخ الحوسبة في القياس التربوي والنفسي:

منذ نشأته ارتبط القياس التربوي والنفسي ارتباطاً وثيقاً بالحوسبة نظراً لاعتماده الكبير على الأساليب الكمية والإحصائية، غير أنّ خصوصية هذا المجال الذي يتعامل مباشرة مع البنى النفسية للأفراد جعلت من مسألة تصحيح الاختبارات وتخزين نتائجها عبر الحاسوب أو الإنترنت قضية حساسة خصوصاً فيما يتعلق بحماية الخصوصية وسرية البيانات.

ومع التقدم السريع في تقنيات الذكاء الاصطناعي (AI) والتعلم الآلي (ML) أصبح بالإمكان استخراج معلومات شخصية دقيقة من بيانات تبدو سطحية للوهلة الأولى. وقد جذب هذا التطور اهتمام قطاعات عديدة، مثل: وكالات التوظيف، وشركات التأمين والتسويق، وحتى الأجهزة الأمنية، التي تسعى جميعها للتنبؤ بالسلوك المستقبلي بدرجة متزايدة من الدقة. وهو ما يفرض ضرورة التعامل بحذر بالغ مع هذه الأدوات لما تحمله من قدرة على كشف معلومات شخصية عميقة وحساسة. لقد أخذت تكنولوجيا المعلومات مكانة محورية متنامية في مجالات كالقبول الجامعي، والامتحانات المهنية، والاختبارات التحصيلية المعيارية، والأدوات السريرية. وبرغم أن الكثيرين لم يستوعبوا بعد مدى هذا التأثير فإن أتمتة عمليات التصحيح وتصميم الاختبارات وتحليل الصدق والثبات قد أسهمت في رفع دقة النتائج بدرجة ملحوظة وهي دقة ما تزال تتطور بوتيرة متسارعة. أما فيما يخص التعامل مع البيانات الضخمة فليس جديداً على القياس النفسي؛ إذ تكفي الإشارة إلى اختبار (Army Alpha) الذي خضع له أكثر من مليون مجند خلال الحرب العالمية الأولى، وهو حجم بيانات هائل بمقاييس ذلك العصر حتى إن نتائجه كانت تكفي لملء كيلومتر كامل من خزائن الملفات.

بدأ الذكاء الاصطناعي أولاً عبر أنظمة الخبراء المعتمدة على محركات استدلال وقواعد معرفة لتقديم الاستشارات في مجالات كالطب والقانون والاقتصاد. ومع مرور الزمن تطور ليعتمد على خوارزميات التعلم الآلي القائمة على البيانات الضخمة حتى بات المصطلحان (AI) و (ML) يُستخدمان أحياناً بشكل مترادف. وتجدر الإشارة إلى أن جذور هذه الخوارزميات تعود إلى البحوث في علم النفس العصبي حول الشبكات العصبية (Rust, et al, 2021, p.152).

وعليه قد يبدو الجمع بين القياس النفسي والتعلم الآلي أمراً غير مألوف إلا أن بينهما قواسم مشتركة تجمع بين علم النفس والإحصاء. فبينما طوّر السيكونوميون الأوائل تقنيات كالتحليل العامل في زمن لم يعرف الحاسوب يقف القياس النفسي اليوم عند مفترق طرق حاسم، حيث

تتقاطع تحليلات البيانات الضخمة مع سرعة المعالجة الفائقة ومع قدرات الذكاء الاصطناعي. وما كان يُنظر إليه يوماً كتطور تدريجي في تقنيات الاتصال الحاسوبي أصبح اليوم بمثابة ثورة عميقة أحدثت تحولاً جذرياً في هذا التخصص (المندلاوي، 2025، ص152) (Rust, et al, 2021, p.152).

3-2- بنوك الأسئلة المحوسبة:

يُعد تطوير وإدارة بنوك الأسئلة (Item Banks) من أبرز المجالات التي تأثرت بالحوسبة في القياس النفسي. فقد أسهم ظهور نظرية الاستجابة للمفردة (IRT) مقروناً بتوافر بيانات تفصيلية على مستوى السؤال في إتاحة أتمتة واسعة لعمليات إدارة هذه البنوك. غير أن النماذج الإحصائية اللازمة لتقدير معايير الأسئلة كانت معقدة وتتطلب عمليات تكرارية كثيرة. وفي سبعينيات القرن العشرين كان تحليل نماذج المعلمين والثلاثة معالم مكلفاً جداً من حيث وقت المعالجة الحاسوبية مما دفع معظم بنوك الأسئلة آنذاك للاعتماد على نموذج راش أحادي المعلم رغم قصوره المعروف. لكن مع دخول الثمانينيات أصبح تطبيق النماذج الأكثر تعقيداً ممكناً ومتاحاً وانتشر استخدامها تدريجياً. ومن أبرز التطبيقات التي استفادت من هذه النماذج الاختبارات التكيفية المحوسبة (CAT)، حيث تُستخدم استجابات المفحوص الأولية لتقدير مستوى قدرته بشكل مبدئي ثم يُوجَّه الاختبار لاحقاً لاختيار أسئلة تتناسب صعوبتها مع هذا المستوى. وقد أثبتت هذه الطريقة فعاليتها في تقليص عدد الأسئلة المطلوبة بما يزيد عن 50% الأمر الذي ينعكس إما على تقليل زمن الاختبار أو تحسين دقته. ومن بين التطبيقات العملية المهمة نجد مقاييس القدرات التفاضلية (Differential Ability Scales - DAS) التي طورتها شركة بيرسون Pearson Assessment في الولايات المتحدة عام 1990. ورغم الجدل في أواخر السبعينيات حول الاعتماد على نموذج راش في بنائها أثبتت هذه المقاييس متانتها وجودتها التطبيقية خصوصاً في الاستخدام السريري مع الأطفال. كما استمرت بعض مقاييسها الفرعية – مثل القدرات العددية والحسابية – في التوصية بها بفضل صلابتها (Rust, et al, 2009, p.202).

فيما أوضح كولين د. إليوت (Colin D. Elliott) مطوّر النسخة الأصلية لهذه المقاييس والمعروفة باسم مقاييس القدرات البريطانية أن معظم الانتقادات الموجهة إلى نموذج راش ارتبطت باستخدامه في بيانات أو اختبارات لم تُصمَّم أصلاً لملاءمته. أما عند توظيفه بوعي وبمعرفة حدوده، كما جرى في تطوير مقياس (DAS) فإنه يتيح الاستفادة من خصائصه الجوهرية كاستقلالية البند عن المفحوص واستقلالية المفحوص عن البند.

3-3- توليد الفقرات بالاعتماد على الحاسوب (Computerized Item Generation):

لم تعد الاختبارات النفسية المحوسبة مقصورة على الاعتماد على بنوك فقرات جاهزة مسبقاً، إذ أصبح بالإمكان توليد فقرات جديدة أو تعديل الفقرات القائمة أثناء سير الاختبار من خلال الحاسوب. وتكمن الفكرة الرئيسية في هذا النهج في الاستعانة بقوالب أو صيغ معيارية تُبنى عليها الفقرات.

عادةً ما تتكون الفقرة من إطار ثابت تُدمج فيه عناصر متغيرة. في صيغة "العلاقات اللفظية (Object-Relations Format)" على سبيل المثال، والتي تأخذ الشكل: «أ هو لـ ب كما أن س هو لـ ص» ويمكن للحاسوب توليد عدد هائل من الأمثلة، مثل: القفاز بالنسبة لليد كما أن الجورب بالنسبة لـ ____ (القدم). وبالاعتماد على هذا النمط يمكن إنشاء ملايين التركيبات الدلالية بصورة آلية. وينطبق الأمر ذاته على أشكال أخرى من الفقرات؛ إذ رغم محدودية الصيغ الأساسية فإن تنوع العناصر الممكن إدراجها يتيح إمكانيات شبيهة غير محدودة في توليد الفقرات.

في كثير من الحالات تُخزن مجموعات من العناصر المحتملة ليُعاد دمجها في قوالب ثابتة مما يسمح بتوليد عدد ضخم من الفقرات الجديدة. ويُستفاد من هذه الآلية خصوصاً في اختبارات الذاكرة، إضافةً إلى إمكانية تطبيقها في الاختبارات الإدراكية والعديدية بل وتكييفها مع معظم أنماط الفقرات الأخرى.

ومن أبرز مزايا هذا الأسلوب قدرته على توفير فقرات مبتكرة ومتنوعة عند الحاجة إلى تكرار الاختبار مع المفحوص نفسه من دون الإضرار بدقة القياس. ويُعد استخدام الجيش البريطاني لهذه الطريقة في بطارية (British Army Recruitment Battery (BARB خلال تسعينيات القرن الماضي مثالاً بارزاً على ذلك، كما أصبحت لاحقاً الأساس الذي انطلق منه مشروع (ICAR) والمسمى بالموارد الدولي المفتوح للقدرات العقلية منذ عام 2018 وما بعده (International Cognitive Ability Resource, 2018; Sun, Liu & Loe, 2019). (Rust, et al, 2021, p.154).

3-4- تطور الذكاء الاصطناعي في علم القياس النفسي:

ترتبط جذور الذكاء الاصطناعي بفترة ما قبل العصر الإلكتروني حين كانت الآلات الذكية الأولى ذات طبيعة ميكانيكية. فقد استُخدمت أدوات حسابية بسيطة كالمعداد والإسطرلاب منذ قرون طويلة أي قبل تصميم تشارلز باباج (Charles Babbage) عام 1820 وهو عالم رياضيات ومخترع بريطاني، لما يُعرف بمحرك الفروق (Difference Engine) والذي كان يهدف إلى تسهيل إعداد الجداول الرياضية المعقدة. ولاحقاً صمم باباج المحرك التحليلي (Analytical Engine) الذي يُعتبر

أول نموذج للحاسوب القابل للبرمجة. المحرك التحليلي الذي لم يُبنَ فعليًا ولكنه جمع بين وظائف الذاكرة والتعليمات باستخدام أجزاء ميكانيكية وبطاقات مثقبة، ليكون بمثابة نموذج أولي للحوسبة الحديثة.

وفي عام 1837 نشرت عالمة الرياضيات آدا لوفلايس (Ada Lovelace) ابنة الشاعر اللورد بايرون (Lord Byron) أول خوارزمية مخصصة للتنفيذ على هذا "الحاسوب الميكانيكي" مما جعلها تُعتبر أول مبرمجة في التاريخ، وأكدت دورها الريادي في ربط الحسابات الميكانيكية بالخوارزميات المنهجية.

خلال بدايات القرن العشرين ظهرت أجهزة تناظرية متعددة حاولت تجسيد أفكار باباج، إلا أن التحول الحقيقي نحو الحوسبة الفعلية أو الحوسبة الرقمية جاء بفضل رؤية آلان تورنغ (Alan Turing) في منتصف القرن العشرين وهو عالم رياضيات ومنطق وعالم حاسوب بريطاني (1912-1954)، حين أوضح الطبيعة الرقمية الكامنة في عملية الحساب وطرح مفاهيم أساسية كآلة تورنغ (Turing Machine) وهي نموذج نظري للحاسوب يوضح كيفية معالجة البيانات والخطوات الحسابية بشكل رقمي. واختبار تورنغ (Turing Test) وهو معيار لتقييم قدرة الآلة على إظهار سلوك ذكي يشبه الإنسان والذي ظل حتى اليوم رمزًا لكيفية تقييم الذكاء الاصطناعي.

أما الذكاء الاصطناعي الحديث فيمكن تقسيمه بوجه عام إلى مجالين رئيسيين: الأنظمة الخبيرة التي تعتمد على قواعد معرفة ومحركات استدلال لتقديم حلول دقيقة في مجالات محددة، والتعلم الآلي (Machine Learning) الذي يركز على تطوير خوارزميات قادرة على التعلم من البيانات الضخمة وتحسين الأداء تلقائيًا (Rust et al., 2009, p. 203).

ويكتسب الذكاء الاصطناعي أهمية خاصة في علم القياس النفسي، حيث أصبح يُستخدم في تحليل البيانات الضخمة وتصميم الاختبارات المحوسبة وتوليد الفقرات آليًا، مما يتيح دقة أكبر في القياس وتنوعًا واسعًا في أدوات التقييم.

المراجع:

- أبو حطب، ف. (1993). التقويم النفسي (بالتعاون مع آخرين). القاهرة: مكتبة الأنجلو المصرية.
- أبو ناهية، صلاح الدين. (1994). القياس التربوي. مصر: مكتبة الأنجلو المصرية.
- أمين محمد، علي سليمان. (2009). القياس والتقويم في العلوم الإنسانية: أسسه وأدواته وتطبيقاته. (مراجعة رجاء محمود أبو علام). القاهرة: دار الكتاب الحديث.
- بركات، زياد. (2018). القياس والتقويم النفسي والتربوي بين النظرية والتطبيق. الناشر: عمادة البحث العلمي بجامعة القدس المفتوحة. فلسطين.
- بيكر، ف. ب. (2001). أسس نظرية الاستجابة للمفردة (ت. عبد الله هاشم، سوسن أبو لهب شلبي، أحمد الطوري، عبد الرحمن بن سليمان أبو هاشم). الرياض: دار القياس والتقويم. العمل الأصلي نُشر بعنوان (The Basics of Item Response Theory).
- بيكر، فرانك (2010). أسس نظرية الاستجابة للمفردة (الطريري، أبو هاشم، سوسن، مترجم). السعودية: النشر العلمي والمطابع جامعة الملك سعود.
- التقي، أحمد محمد (2013). النظرية الحديثة في القياس. ط2. عمان: دار المسيرة للنشر والتوزيع.
- التميمي، ر. ر. ح.، والشيخ، ف. م. أ. (2021). اتجاهات حديثة في التكنولوجيا والتعليم الإلكتروني. بغداد، العراق: مؤسسة دار الصادق الثقافية.
- ثورندايك، روبرت & هيجن، إليزابيث. (1989). القياس والتقويم في علم النفس والتربية. ترجمة: عبد الله زيد الكيلاني، عبد الرحمن عدس. عمان: مركز الكتب الأردني.
- جودة، أحمد سعادت. (2001). صياغة الأهداف التربوية والتعليمية في جميع المواد الدراسية. عمان: دار الشروق.
- الخزي، فهد. (2016). دراسة أثر بعض المتغيرات على أداء طلبة الصف الحادي عشر في مدارس دولة الكويت في الاختبارات الإلكترونية، مجلة اتحاد الجامعات العربية للتربية وعلم النفس، (3). (174)
- خليل، حنان حسن علي (2017). التقويم الإلكتروني، المملكة الأردنية الهاشمية، عمان: دار الميسرة للطباعة والنشر.
- الدامغ، خالد عبد العزيز (2006). المستقبل الرقمي للقياس والتقييم التربوي، المجلة العربية للتربية، (3)، ص 101-130.

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. الرسالة دكتوراه غير منشورة. كلية التربية. قسم علم النفس. جامعة أم القرى.

ديفيليس، ر. ف. (2021). تطوير المقاييس: النظرية والتطبيق (ترجمة س. ب. س. القحطاني، وط. ع. عبد الرحمن؛ مراجعة ع. ب. س. العنقري). مكتبة الملك فهد الوطنية، مركز البحوث والدراسات. المملكة العربية السعودية.

زياد، رشيد. (2023). التقويم السيكمومتري للاختبارات والمقاييس النفسية والتربوية. الناشر: دار المتنبي للطباعة والنشر. المسيلة. الجزائر.

زيتون، ح. ح. (2005). رؤية جديدة في التعليم: التعلم الإلكتروني – المفهوم، القضايا، التطبيق، التقييم. الرياض: الدار الصولتية للتربية.

سامي، محمد ملحم. (2011). القياس والتقويم في التربية وعلم النفس. عمان. دار المسيرة للنشر والتوزيع.

السعداوي، عبد الله بن صالح. (2018). دليل المعلم للتقويم المعتمد على الأداء من النظرية إلى التطبيق، ط2، الناشر: مكتب التربية العربي لدول الخليج، الرياض. المملكة العربية السعودية.

سوسن، شاكراً مجيد (2005). أساسيات بناء الاختبارات والمقاييس النفسية والتربوية. ط3. دار النشر: مؤسسة علاء الدين للطباعة والتوزيع. دمشق. سوريا.

سوسن، شاكراً مجيد. (2014). أساسيات بناء الاختبارات والمقاييس النفسية والتربوية. عمان: دار رسلان للطباعة والنشر والتوزيع.

السيد علي، أسامة زكي (2016). الاختبارات اللغوية، المملكة العربية السعودية، الرياض: مركز الملك عبد الله بن عبد العزيز لخدمة اللغة العربية.

صبحي، سالي وديع. (2005). الاختبارات الإلكترونية عبر الشبكات، جمهورية مصر العربية، القاهرة: عالم الكتب.

علام، صلاح الدين محمود (2000). القياس والتقويم التربوي والنفسى أساسياته وتطبيقاته وتوجهاته المعاصرة. القاهرة: دار الفكر العربي.

علام، صلاح الدين محمود (2013). نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي. القاهرة: دار الفكر العربي.

علي مهيدات، عبد الحكيم؛ المحاسنة، إبراهيم محمد. (2008). التقويم الواقعي، الناشر: دار جرير للنشر والتوزيع. عمان. الأردن.

عودة، أحمد سليمان. (1985). القياس والتقويم في العملية التدريسية. دار الأمل للطباعة والنشر والتوزيع.

الفرطوسي، علي سموم؛ الميداني، شذى فؤاد. (2021). التقويم التربوي المستمر وتطبيقاته العملية، الناشر: دار الفكر العربي، القاهرة، مصر.

فؤاد، أبو حطب وسيد، عثمان وآمال، صادق. (2008). التقويم النفسي. ط4. القاهرة: مكتبة الأنجلو المصرية.

القُدومي، عبد الناصر. (2008). الاختبارات التحصيلية وطرق إعدادها. كلية التربية الرياضية، مركز القياس والتقويم.

كاظم، أمينة محمد. (1988أ). دراسة نظرية نقدية حول القياس الموضوعي للسلوك نموذج راش، مؤسسة الكويت للتقدم العلمي.

كاظم، أمينة محمد. (1988ب). استخدام نموذج راش في بناء اختبار تحصيلي في علم النفس وتحقيق التفسير الموضوعي للنتائج، مطبوعات جامعة الكويت.

الكلوت، أحمد اسماعيل (1991). الخصائص السيكمترية لمقياس (مايرز – بريجز) لأنماط الشخصية وقدرة المقياس على التنبؤ بالتحصيل الدراسي. أطروحة دكتوراه. كلية الدراسات العليا. الجامعة الاردنية.

محاسنة، ابراهيم محمد (2013). القياس النفسي في ظل النظرية التقليدية والنظرية الحديثة. عمان. الاردن: دار جرير للنشر والتوزيع.

مخائيل، امطانيوس. (1996). القياس والتقويم في التربية الحديثة، منشورات جامعة دمشق، دمشق، سوريا.

مخائيل، امطانيوس. (2020). القياس النفسي (الجزء الأول)، منشورات جامعة دمشق، دمشق، سوريا.

مراد، صلاح أحمد وسليمان، أمين علي. (2005). الاختبارات والمقاييس في العلوم النفسية والتربوية وخطوات إعدادها وخصائصها. الجزائر: دار الكتاب الحديث.

معمرية، بشير. (2007). القياس النفسي وتصميم أدواته. ط2. سلسلة دراسات، منشورات الحبر.

المندلأوي، ض. ع.، والمندلأوي، ع. ع. (2021). القياس والتقويم في التربية وعلم النفس. بغداد: مكتب الأمير للطباعة والاستنساخ.

المندلأوي، علاء عبد الخالق. (2025). القياس والتقويم التربوي الأسس والنظرية والتطبيقات العملية. مؤسسة دار الصادق الثقافية. بغداد. العراق.

ميمي السيد، أحمد اسماعيل (2007). الخصائص السيكمترية لاختبار القدرة العقلية باستخدام نموذج راش لدى طلبة المرحلة الثانوية العامة. رسالة ماجستير غير منشورة. كلية التربية. جامعة الزقازيق.

ميمي السيد، أحمد إسماعيل (2014). التوجهات الحديثة في القياس النفسي والتربوي. القاهرة: دار الكتاب الحديث.

النبهان، موسى. (2013). أساسيات القياس في العلوم السلوكية. ط2. عمان. الناشر: دار الشروق للنشر والتوزيع.

النعمي، عز الدين (2011). أثر الزيادة في عدد المفردات المرتبطة على الخصائص السيكمترية للمفردة والاختبار. مجلة اتحاد الجامعات العربية للتربية وعلم النفس. 9 (3). 158-178.

وليد، مسعود (2014). محاضرات في القياس نظرية الاستجابة للمفردة مع تطبيقات عملية. كلية التربية قسم علم النفس. جامعة أم القرى.

اليعقوبي، حيدر. (2013). التقويم والقياس في العلوم التربوية والنفسية- رؤيا تطبيقية. الناشر: مركز المرتضى للتنمية الاجتماعية. بغداد. العراق.

Algahtani, A.F. (2011). Evaluating the Effectiveness of the E-learning Experience in Some Universities in Saudi Arabia from Male Students' Perceptions. Durham theses, Durham University.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.

American Psychological Association. (2010). Publication manual of the American Psychological Association. Washington, DC: American Psychological Association.

Anastasi, A., & Urbina, S. (1997). Psychological Testing (7th ed.). Prentice Hall

Ayala, R.j . (2009). The Theory and Practice of item Response Theory. A division of Guilford Publications, Inc 72 Spring, Street, New York, Ny 10012. www.guilford.com

Baker, F.B. (2001). The Basics of Item Response Theory. ERIC clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD.

- Basu, A.; Cheng, I.; Prasad, M and Rao, G.(2007). Multimedia adaptive computer-based testing: an overview, Special Session, July Beijing, 1850-1853.
- Billo, Redwan Oluwaseun; Olugbebi, Muiyiwa; Babatunde, Abdulraulph; Bello, Bashir Omoloran; & Billo, Shakirat Iyabo (2016). Design and implementation of web-based examination system for the university. *Journal of Computer Science and Control Systems*9 (2): (p5-9.
- 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.
- Christine, D.(2010). Item Response Theory Understanding Statistics Measurement. Oxford
- Coaley, K. (2010). An introduction to psychological assessment and psychometrics. SAGE Publications.
- Cohen, R. J., Swerlik, M. E., & Sturman, E. D. (2013). *Psychological Testing and Assessment: An Introduction to Tests and Measurement* (8th ed.). McGraw-Hill.
- Cole, N. S., & Zieky, M. J. (2001). The new faces of fairness. *Journal of Educational Measurement*, 38, 369–382.
- Dermo.j (2009) .E-Assessment and the student learning experience, A survey of student perception of assessment, *British jornal of Education Technology*, 40(2).
- DeVellis, R. F. (2016). *Scale Development: Theory and Applications* (4th ed.). SAGE Publications.
- DiIorio, C. K. (2005). *Measurement in health behavior: methods for research and evaluation*. John Wiley & Sons.
- Educational Testing Service. (2009a). ETS international principles for fairness review of assessments. Princeton, NJ: Educational Testing Service.
- 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 .
- Embretson, S.E. and Reise, S.P. (2000) .*Item Response Theory for Psychologists*. Lawrence Erlbaum Associates, Inc., Mahwah.
- Francis ,Tuerlinckx ; Paul, DeBoeck.(2001). Non – modeled item interactions lead to distorted discrimination parameters : A case Study . *Methods of Psychological Research Online* 2001,Vol 6,No 2 ,p159-174. <http://www.mpr-online.de>
- Garson, G. D. (2013). *Validity and reliability*. Statistical Associates Publishing. <https://www.statisticalassociates.com>
- Gibbons, R. D., Weiss, D. J., Frank, E., & Kupfer, D. (2016). Computerized adaptive diagnosis and testing of mental health disorders. *Annual review of clinical psychology*, 12, 83-104.

- Gipps, C., & Stobart, G. (2009). Fairness in assessment. In *Educational assessment in the 21st century: Connecting theory and practice* (pp. 105-118). Dordrecht: Springer Netherlands.
- Hambleton, R.K. (1989). Principles and applications of item response Theory in R.L.Linn (Ed). *Educational measurement* (3rd edition; pp147-200). New york.
- Harrow, A.J. (1972). *A taxonomy of the psychomotor domain*. New York: David McKay Co.
- Kane, M. T. (2013). Validating the interpretation and uses of test scores. *Journal of Educational Measurement*, 50, 1–73.
- 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.
- Kumar, H., & Rout, S., Dalalh, M., Ahmad, J., Khan, A., Chandan, J. S., Kothari, C. R., & Koul, L. (2016). *Measurement and evaluation in education* [English edition]. Tripura University, Directorate of Distance Education. Vikas Publishing House Pvt. Ltd.
- Louis L. Thurstone. (1925). A Method of Scaling Psychological and Educational Tests. *Journal of Educational Psychology*. 16: 433- 451.
- Magno, C., & Ouano, J. (2010). *Designing written assessment for student learning*. Philippines: Phoenix Pub.
- Masters G.N. (1988). Item discrimination: when more is worse. *Journal of Educational Measurement* 25:1, 15-29.
- Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: Macmillan.
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (1999). *Evidence-centered assessment design*. Princeton, NJ: Educational Testing Service.
- Mislevy, R. J., & Verhelst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika*, 55(2), 195-215.
- Morrow, J. R., Jr., Jackson, A. W., Disch, J. G., & Mood, D. P. (2000). *Measurement and evaluation in human performance* (2nd ed.). Human Kinetics.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (3rd ed.). McGraw-Hill.
- O'Neill, K. A., & McPeck, W. M. (1993). Item and test characteristics that are associated with differential item functioning. In P. Holland & H. Wainer (Eds.), *Differential Item Functioning* (pp. 255–276). Hillsdale, NJ: Erlbaum.
- 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.

- Ravitch, D. (2003). *The language police: How pressure groups restrict what students learn*. New York: Knopf.
- Reise, S. P., & Haviland, M. G. (2005). Item response theory and the measurement of clinical change. *Journal of personality assessment*, 84(3), 228-238.
Retrieved from www.ets.org/s/about/pdf/fairness_review_international.pdf
- Rex B, Kline. (2011). *Principles and practice of structural equation modeling*. 3rd ed. A Division of Guilford Publications, Inc. 72 Spring Street, New York, NY 10012.
- Romiszkowski, A. (2004). How's the E-learning Baby? Factors Leading to Success or Failure of an Educational Technology Innovation. *Educational Technology*, 44(1):5-27.
- Rust, J., & Golombok, S. (2009). *Modern psychometrics: The science of psychological assessment*. Routledge.
- Rust, J., Kosinski, M., & Stillwell, D. (2021). *Modern psychometrics: The science of psychological assessment* (4th ed.). Routledge.
- Schmitt, A. P., Holland, P. W., & Dorans, N. J. (1993). Evaluating hypotheses about differential item functioning. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp.281–315). Hillsdale, NJ: Erlbaum Associates.
- Shepard, L. A. (1987). The case for bias in tests of achievement and scholastic aptitude. In S. Modgil & C. Modgil (Eds.), *Arthur Jensen: Consensus and controversy* (pp. 210–226). London: Falmer Press.
- Sivakumar, A., & Thirumoorthy, G. (2019). *Measurement and evaluation in education*. New Delhi, India: A.P.H. Publishing Corporation.
- Tes and Rabu. (2011, June 8). Reliabilitas instrument tes (Bag. 1). *Serpihan Media dan Ruang untuk Berbagi Pengetahuan*. <https://gerbangbuku.blogspot.com/2011/06/reliabilitas-instrumen-tes-bag-1.html>
- Thomas, M. L. (2019). Advances in applications of item response theory to clinical assessment. *Psychological assessment*, 31(12), 1442.
- Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal Design Applied to Large Scale Assessments*. Synthesis Report.
- Tripura University, Directorate of Distance Education. (2016). *Measurement and evaluation in education* (M.A. [Education], 2nd semester, EDCN 804E) [English edition]. Vikas Publishing House Pvt. Ltd. University Press, Inc.
- Urbina, S. (2004). *Essentials of behavioral science series. Essentials of psychological testing*.

- Voyer, D., Voyer, S., & Bryden, P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, 117, 250–270.
- Willingham, W. W., & Cole, N. S. (1997). *Gender and fair assessment*. Mahwah, NJ: Erlbaum.
- 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
- Wright, B. D., & Stone, M. H. (1979). *Best test design*.
- Wu, M., & Adams, R. (2007). *Applying the Rasch model to psycho-social measurement: A practical approach* (p. 87). Melbourne: Educational Measurement Solutions.
- Zieky, M. J. (1993). Practical questions in the use of DIF statistics in test development. In P. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 347–348). Hillsdale, NJ: Lawrence.
- Zieky, M. J. (2016). Fairness in test design and development. In *Fairness in educational assessment and measurement* (pp. 9-31). Routledge.

المؤلف/ الدكتور زياد رشيد



أستاذ محاضر قسم "أ" بجامعة الشهيد الشيخ العربي التبسي – تبسة، الجزائر، من مواليد 10 نوفمبر 1974 بمدينة تڨرت – ورقلة. زاول دراسته الابتدائية والمتوسطة ببلدية المرارة، ثم درس المرحلة الثانوية ببلدية جامعة – المغير. حصل على شهادة الليسانس في علم النفس المدرسي من جامعة الوادي، ثم واصل مساره الأكاديمي بجامعة وهران 2، حيث نال شهادة الماجستير والدكتوراه علوم في تخصص القياس والتقويم النفسي والتربوي. إلى جانب مساره التدريسي والبحثي، أشغل بعدة مهام أكاديمية وعلمية؛ حيث يعمل كـ مسؤول تخصص ماستر إرشاد وتوجيه، وعضو في اللجنة العلمية لقسم علم النفس، بالإضافة إلى كونه رئيس مركز المساعدة النفسية الجامعي. كما أنه عضو في مخبر الدراسات النفسية والاجتماعية – فرقة علم النفس العصبي والكرونيولوجيا بجامعة بسكرة، وسبق له أن كان باحثًا وعضوًا في مخبر تطبيقات علوم النفس وعلوم التربية من أجل التنمية في الجزائر (L.A.S.P.E.D.A.L) فرقة الاختبارات (تصميم، تكييف، تعريب) بجامعة وهران 2. يمثل اهتمامه الأكاديمي والبحثي تقاطعًا بين القياس والتقويم والإحصاء النفسي والتربوي، مع تركيز خاص على تطوير وتكييف وتقنين الأدوات النفسية والتربوية بما يخدم احتياجات المتعلمين والباحثين في السياقات الجزائرية والعربية.

لقد أسهم المؤلف في إثراء مجال القياس والتقويم والإحصاء النفسي والتربوي من خلال نشر عدة بحوث ومقالات علمية، والمشاركة في نوات وأيام دراسية وملتقيات وطنية ودولية. كما أشرف على تنظيم ورشات تكوينية متخصصة في القياس والإحصاء، تناولت موضوعات متقدمة مثل: النمذجة بالمعادلات البنائية (SEM) باستخدام برنامجي *Amos* و *SmartPLS*، ونظرية الاستجابة للمفردة الاختبارية (IRT) باستعمال برنامجي *Winsteps* و *Bilog-mg3*، إضافة إلى التدريب على تطبيقات الحزمة الإحصائية للعلوم الاجتماعية (SPSS). ويضاف إلى ذلك خبرته العملية بالمعهد الوطني للبحث في التربية (INRE) التابع لوزارة التربية الوطنية، والتي مكنته من الجمع بين العمق الأكاديمي والتطبيق الميداني في تطوير الممارسات البحثية والتربوية.

عن هذا الكتاب

يأتي هذا الكتاب ليلبي حاجات طلبة السنة الثانية علوم التربية، حيث يهدف إلى تزويدهم بالمعارف النظرية والمهارات التطبيقية المرتبطة بمادة القياس التربوي وبناء الاختبارات المدرسية. وقد تم إعداد محتواه بما يتماشى مع متطلبات عرض التكوين الجامعي (Canevas) في هذا المستوى، مع التركيز على تبسيط المفاهيم وتوضيح الإجراءات خطوة بخطوة.

يتناول الكتاب جملة من الدروس والمحاور الأساسية، من أبرزها:

- الخصائص السيكمترية لأدوات القياس النفسي والتربوي وكيفية التحقق منها.
- أسس بناء وتصميم الاختبارات المدرسية وفق معايير علمية دقيقة.
- إعداد جدول المواصفات وربطه بالأهداف التربوية.
- التعريف بالأهداف التربوية وتصنيفاتها وعلاقتها بالعملية التعليمية.
- استعراض أهم تطورات القياس التربوي الحديث وانعكاساتها على الممارسة التربوية.
- التعرف على نظرية القياس الكلاسيكية (CTT) ومحدداتها.
- مدخل إلى نظرية الاستجابة للمفردة (IRT) وتطبيقاتها الحديثة.

وبهذا يوفر الكتاب إطارًا علميًا وتطبيقيًا يساعد الطلبة على فهم أعمق لأسس القياس التربوي، ويكسبهم كفاءات عملية في بناء الاختبارات وتحليلها بما يتوافق مع متطلبات البحث الأكاديمي والممارسة التربوية في المدرسة.