1

الفصل الأول

مبادئ القياس النفسى

الفصل الأول

مبادئ القياس النفسى

إن القياس العقلى، أو القياس النفسى، هو مجال ملئ بالمصطلحات والمفاهيم الفريدة من نوعها؛ وتتضمن بعض الأمثلة مصطلحات مثل الدرجات التائية، الدرجات المعيارية، معاملات الثبات والصدق التلازمى. وغالبًا ما تكون أسماء الاختبارات ذاتها أسماء معقدة وتحتاج إلى مهارة وتدعى مزاعم مؤثرة بأنها تقيس السمات. وتسود نتائج الاختبار الكثير من الأرقام، وتلخص هذه الأرقام أحيانًا بطريقة مثل «كانت درجة هذا الاختبار الفرعى ٨، وكانت درجة ذلك الاختبار الفرعى ١١، وكانت نسبة ذكاء المقياس الكلية ٨٣، وكانت الرتبة المئوية ١٣».

ويبدو أن الاخصائيين النفسيين يدعون أنهم باستخدام درجات الاختبار هذه يستطيعون التوصل إلى معلومات غير متاحة لمعظم الناس.

كيف تعمل هذه الاختبارات؟

يحاول هذا الفصل أن يساعدك على فهم الكيفية التى تعمل بها الاختبار، عن طريق الكشف عن بعض المنطق وراء القياس النفسى ومبادئه. إن تفرد القياس العقلى أو الاختبار النفسى يقوم على أساس عدة مفاهيم ومبادئ رئيسية. من بين هذه المفاهيم والمبادئ، التطبيق المنتظم والتصحيح الموضوعى، التقدير الكمى للنتائج، استخدام الإسناد المعيارى والدرجات المشتقة، ومفهومان يُسميان الثبات والصدق. وعلى الرغم من أن بعض هذه المفاهيم قد تكون معقدة، إلا أنه من أجل أهداف هذا الكتاب بُذلت مجهودات لتبسيط هذه المفاهيم ولاختصار المناقشة. وعن طريق تنمية فهم لهذا المفاهيم، سيتحسن فهم القراء إلى حد بعيد للكيفية التى يمكن بها استخدام الاختبارات النفسية.

الانتظام والموضوعية

معظم الاختبارات النفسية لها تطبيق على نسق واحد وتصحيح موضوعى؛ ومعظم الإجراءات غير الرسمية ليست كذلك. (تشير بعض المراجع مثل كرونباخ، ١٩٨٤، إلى هذا الثبات والموضوعية كتقنين).

وتعتبر اختبارات نسبة الذكاء المقننة أمثلة للاختبارات النفسية، حيث يقتضى هذا الاختبار تقديم فقرة محكمة وتماسك دقيق لمعايير التصحيح الموضوعى. أحد هذه الاختبارات التى تقيس نسبة الذكاء وكسلر لذكاء الأطفال (وكسلر، ١٩٩١)، وقد تم تنقيح هذا المقياس إلى الدرجة التى يحدد فيها دليل المقياس مقدار الإيضاح المسموح بها، عندما يقدم المفحوصين إجابات ملتبسة على الأسئلة اللفظية وكيفية تقدير حتى الأخطاء البسيطة التى يقع فيها الأطفال فى مهام تركيب المكعبات. وتتطلب الكفاءة فى تطبيق مقاييس وكسلر وتصحيحها مقرر دراسى بعد التخرج، يستغرق فصلاً دراسياً

يتباين هذا الانتظام وهذه الموضوعية بشدة مع الإجراءات غير الرسمية التى يستخدمها غير الأخصائيين النفسيين دائمًا لقياس النمو. فعلى سبيل المثال، قد يستخدم أخصائى الأطفال ما يرى أنه إجراء مهمة معيارية على نحو ملائم لتقدير نمو الأطفال، الذين لديهم تأخر كامن فى النمو. قد يُطلب من كل الأطفال أن يوضحوا ما تعنيه بعض الكلمات، وأن يتعرفوا على بعض الألوان، وأن يرسموا بعض الأشكال البسيطة. قد ينمو لدى اخصائى الأطفال شعور شخصى وذاتى للأداء المتوسط فى الأعمار المختلفة. وقد يسأل اخصائى الأطفال، بينما يحاول الحفاظ على الأنساق، أحد الأطفال ما هو «الكلب»؛ «القط» و«الحمار» ولكن يسأل طفل ثانى عن «كلب صغير»، «قطة صغيرة» و«حمار صغير». وقد تشجع مثل هذه التساؤلات الطفل الرافض على أن يستجيب بعبارة مثل «أتعلم (كلمة لجذب الانتباه)، الكلب الصغير ـــ باو واو». وقد يُطلب من أحد الأطفال أن يشير إلى أشياء فى الحجرة التى ألوانها «أحمر»، «أزرق»، «أخضر» و«أصفر»، ولكن قد يصل طفل آخر فى يوم لا توجد فيه أشياء باللون الأزرق، وبذلك قد يُطلب منه أن يتعرف على شىء باللون البرتقالى بدلاً من الأزرق. وقد تحدث تضاربات بسيطة أخرى فى مهام الرسم.

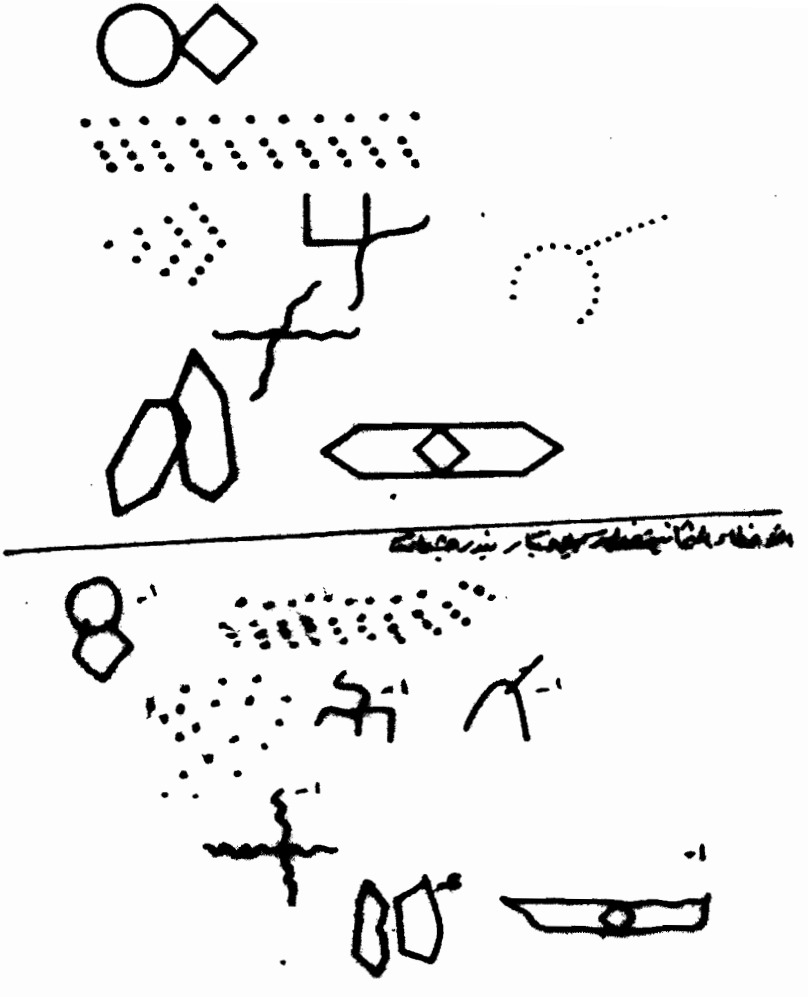
أفترض أن أخصائى الأطفال قد كشف عن مجموعة معقولة من المعايير العمرية للنمو من خلال الممارسة، عندئذ قد تكون إجراءاته ملائمة ليتبين معظم أشكال تأخر النمو الحادة، ولكن قد تغفل الكثير من أشكال التأخر المعتدلة.

فى المقابل، الإجراءات الأكثر ثباتًا وموضوعية سيكون لها ميزة التعرف على كل أشكال التأخر الحاد تقريبًا ومعظم أشكال التأخر المعتدل أيضًا.

وللطريقة غير الرسمية التى يستخدمها أخصائى الأطفال التى شرحناها آنفًا عيوب ؛لأن أشكال التأخر الدقيقة لا يمكن أن تكتشف بالإجراءات غير الرسمية. لايستطيع خبراء التشخيص غير الرسميين أن يعرفوا عما إذا كانت الفروق القليلة الملحوظة هى نتيجة تغايرات فى التطبيق والتصحيح، أم أنها فروق حقيقية فى نمو الطفل. ربما التشجيع الزائد («باو ـ واو») الذى أُعطى للطفل الرافض يمثل ميزة هامة لذلك الطفل وعيبًا لطفل آخر. فى المقابل، فإن الإخصائى النفسى الذى يستخدم القياس والتصحيح الرسميين يستطيع أن يكون واثقًا بأن حتى الفروق القليلة الملحوظة تعكس تغايرات فى نمو المفحوصين. وبالمصطلحات المتعلقة بالقياس النفسى، تحتوى الأساليب غير الرسمية على أخطاء فى القياس أكبر بكثير من الأخطاء التى تحتويها الأساليب الرسمية المقننة.

قابلية القياس

الاختبارات النفسية قابلة للقياس عمومًا؛ بمعنى أنها تسفر عن درجات رقمية. وقد تم تعريف القياس على أنه تخصيص منظم من الأعداد للأفراد، باستخدام إجراءات منضبطة ومقررة مسبقًا وقابلة للتكرار (الن ين، ١٩٧٩). على المستوى الأبسط، هذا قد يكون جدولة عدد البنود الصحيحة، التى تسمى ببساطة الدرجة الخام. تكون هذه الجدولة سهلة عندما تتكون الاختبار من عبارات موضوعية بنظام للتقدير «ناجح ـ راسب»، ولكن الأخصائيين النفسيين يذهبون إلى ما هو أبعد من ذلك ويجتهدون لقياس ما يبدو أنه اختبارات غير قابلة للقياس الكمى. وأحد أمثلة هذه الاختبار اختبار بندر جشتلت البصرى الحركى (بندر،١٩٤٦) الموضح فى الشكل ١ـ ١.



شكل (١ ـ ١)

نموذج بدنر جشطالت، مأخوذ من الاختبارات

البرصى لندر جشطالت، والذى طورته كوبتيز وفقًا للجمعية الأمريكية للطب النفسى

إذا بدا أى اختبار غير قابل للقياس الكمى، فإن اختبار بندر جشتلت كذلك يتكون الاختبار من تسعة مخططات هندسية، يقوم الطفل بنسخها فى ورقة بيضاء. الهدف من هذا الاختبار هو أن نقيس النضج البصرى الحركى الدقيق. عادة ، دون وجود نظام تصحيح قابل للتقدير الكمى، فإن الأخصائيين النفسيين يقيمون الرسوم بطريقة ذاتية للصفة العامة للنسخ المرسومة. حدث تقدم أساسى عندما طورت الأخصائية النفسية إليزابيث كوبيتز (١٩٦٣) نظامًا قابلاً للتقدير الكمى، والذى يسمح لكل صورة أداء للاختبار أن تُحسب كميّاً (وموضوعيّاً) لـ ٢٧ خطأ محتملاً. لماذا يعتبر التقدير الكمى ميزة؟. يوجد سببان: الأول، تقدم الأرقام وصفًا أكثر دقة من البيانات اللفظية، التى تؤدى إلى اتصال أفضل. إن وصفًا لصورة أداء اختبار بندرجشتلت بثمانية أخطاء يكشف للمُشخص المدرب بشأن مظهر صوره أداء الاختبار، أكثر مما أن تقول ببساطة إن رسومات بندر كانت متتالية ومتقاطعة ومرسومة على نحو رديء. السبب الثانى، والأكثر أهمية، يمكن أن تُعالج الدرجات الرقمية حسابيّاً لتسمح حتى باتصال أفضل، كما يُرى بالشكل (١ ـ ١).

بالنسبة للمراقب العرضى، فإن الموضوعية والتقدير الكمى هما الفرقان الوحيدان بين القياس النفسى الأصولى (الرسمى) والتقييم التقديرى. فى الواقع من السهل أن نرى أن الأخصائى النفسى يطبق كل بند على نحو صارم بالأسلوب نفسه تمامًا وينتهى بأرقام، بينما المشخص التقديرى (غير الرسمى) لا يفعل ذلك. إن الموضوعية والتقدير الكمى هما الميزتان الأكثر ووضوحًا وليسا الأكثر أهمية للقياس النفسى. إن القدرة الحقيقة للقياس النفسى تنشأ عن المقارنة بمجموعة معيارية.

المقارنة المعيارية والدرجات المشتقة:

تجتاز الاختبار النفسية المقننة مرحلة بحث وتطوير قبل أن يسمح بها للاستخدام الاكلينيكى. كجزء من هذه المرحلة، يُطبق اختبار توقعى على العديد من الأطفال. وتشكل درجات هؤلاء الأطفال مجموعة معيارية (ايكن، ١٩٩٤)، والتى يشار إليها أيضًا على أنها عينة التقنين. إن تكوين مجموعة معيارية هو الخاصية المميزة للاختبارات النفسية المقننة، حيث تساعد المجموعة المعيارية الأخصائيين النفسيين

على المراقبة الشديدة لما هو طبيعى بالفعل. عندما يؤذن للاختبار بالاستخدام العام، يمكن مقارنة درجة كل طفل بالمجموعة المعيارية لتحديد مرتبته بالنسبة إلى أقران، وسوف يناقش مبدأ المقارنة المعيارية فيما بعد فى هذا الفصل مناقشة مستفيضة فى الجزء المعنون «اختبار عينة ممثلة».

إن القدرة على إجراء مقارنات معيارية يعتبر مزية عظيمة للقياس النفسى الأصولى (الرسمى). دون مجموعة معيارية، يجب على المشخص أن يعتمد على الانطباع الذاتى وعلى التقدير الشخصى لما يمكن أن يكون عليه الأداء النموذجى للطفل. وبوجود مجموعة معيارية، لا يحتاج الإخصائيون النفسيون إلى تقدير عما إذا كان أداء أحد الأفراد متوسطًا؛ لأنهم ببساطة يقارنون أداء الفرد بالمجموعة المعيارية ويرون مدى مستواه.

يتجلى التأثير الكامل للمقارنة المعيارية عندما تستخدم الدرجات المشتقة. الدرجة المشتقة هى درجة مفردة توضح مدى مضاهاة أداء الطفل للمجموعة المعيارية، الذى يقدم طريقة لتلخيص الأداء. ويتم التوصل إلى الدرجات المشتقة بحساب الدرجة الخام أولاً. والدرجة الخام عادة هى عدد العبارات (البنود) الصحيحة، والخطوة التالية هى أن نرجع إلى الجداول التى تخبرنا عن كيفية أداء المجموعة المعيارية. وتسمح هذه الجداول للأخصائيين النفسيين أن يتوصلوا إلى الدرجة المشتقة، مثل تكافؤ السن، الرتبة المئوية أو الدرجة المعيارية. بهذا المعنى، تشتق كل هذه الدرجات بمقارنة أداء أحد الأفراد بأولئك فى مجموعة الأقران. والهدف من الدرجات المشتقة هو أن نحول الدرجة الخام التى بلا معنى إلى درجة يفهم الغرض منها (هلمستدلر، ١٩٦٤). إن معظم الدرجات التى يألفها القراء من خلال إجرائهم لاختبارات القبول (اختبار الاستعداد المدرسى العام، اختبار القبول بمدرسة القانون، اختبار القبول بكلية الطب) هى درجات مشتقة.

مكافئا العمر والفرقة الدراسية

تستخدم الدرجات المتكافئة لوصف السمات التى تنمو كما ينمو الطفل (مثلاً، مهارات القراءة، النضج الادراكى). تقرر الدرجات المتكافئة مستوى النمو أو التطور الذى وصل إليه الطفل، من خلال تبيين العمر أو الفرقة الدراسية التى كان أداء

الطفل مساويًا لها أو متكافئًا معها. على سبيل المثال، مكافىء القراءة فى الصف السابع يعنى أن درجة الطفل الخام (عدد بنود القراءة الصحيحة) كانت مماثلة لمتوسط الدرجة الخام بالنسبة لطلبة الصف السابع الآخرين. فى الممارسة الفعلية، يتم تقسيم مكافئ العمر والفرقة الدراسية إلى أجزاءالسنة. ومن ثم، يشير مكافئ العمر ١١ شهر ـ ٧ سنوات إلى أن أداء فرد ما كان متطابقًا مع متوسط أداء كل الأطفال الذين أعمارهم ٧ سنوات و١١ شهر فى المجموعة المعيارية (انظر الجدول ١ـ١).

جدول (١ ـ١): نموذج لمناقشة مكافىء العمر والنسبة المئوية ودرجات الصف

درجات الصف	مكافىء العمر (الشهر / السنة)	النسبة لعمر ٨ سنوات
١٧	الميلاد ـ ١٠	٩٩
١٦	٦ ـ ٩	٩٨
١٥	٠ ـ ٩	٩٦
١٤	٦ ـ ٨	٩١
١٣	٥ ـ ٨	٨٤
١٢	٢ ـ ٨	٧٥
١١	١ ـ ٨	٦٣
١٠	٠ ـ ٨	٥٠
٩	٦ ـ ٧	٣٧
٨	٣ ـ ٧	٢٥
٧	١ ـ ٧	١٦
٦	٠ ـ ٧	٨
٥	٨ ـ ٦	٥
٤	٢ ـ ٦	٢
٣	٠ ـ ٦	١
٢	١١ ـ ٥	١
١	٥ ـ ٥	١

وبالمثل، مكافىء الفرقة الدراسية ٣ ـ ٢ فى القراءة يشير إلى أن درجة قراءة الطفل كانت مـتطابقة مع متـوسط الدرجة التى حصل عليـها أطفال المجـموعة المعـيارية فى الشهر الثانى بالصف الثالث. إن الدرجات المتكافئة بكل أنواعها شائعة فى السياقات التعليمية، وبين المشخصين من غير الإخصائيين النفسيين.

الرتب المئوية

إن الرتب المئوية عادلة مثلهـا مثل الدرجات المتكافئة. تقرر الرتب المئـوية ما النسبة المئوية للمجموعة المعيارية التى تجـاوزتها درجة الفرد. وبذلك، تشير الرتبة المئوية ٢٣ إلى أن درجة هذا الفرد تخطت ٢٣ مـن بين كل مائة شخص فى المجموعـة المعيارية؛ وتشير الرتبة المئوية ٧٥ إلى أن درجة الـفرد تخطت درجات ٧٥ فرد من بين مائة فرد من أفراد المجمـوعة المعيارية، وهلم جرا. وكمـا فى مكافئى العمر والفرقـة الدراسية، تقدم هذه الدرجـات وسيلة لتفـسير أداء فرد مـعين مقارنة بأداءات أقـرانه. والطريقة الأخرى والأكثر تنقيحًا لقياس الأداء هى استخدام الدرجات المعيارية.

الدرجات المعيارية

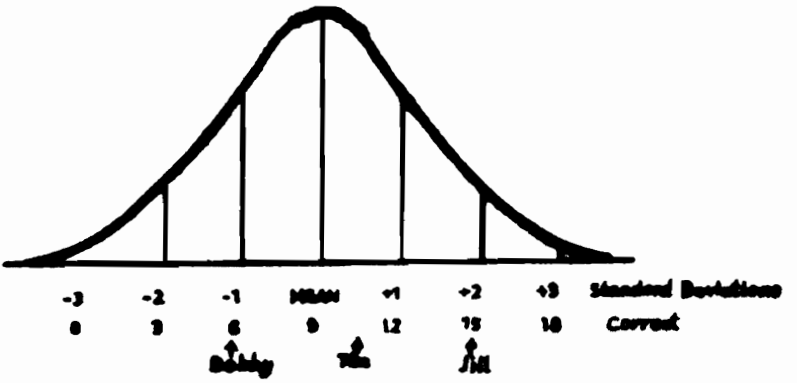
الدرجات المعيارية هى الدرجات المشتقة التى تقـدم طريقة دقيقة لتحديد أداء أحد الأفراد مـقارنة بمجمـوعة أقران محـددة. يتم إنجاز هذه المهمـة بتعيين عـدد الوحدات المعيارية التى أكـثر أو أقل من الدرجة التى عليها المـتوسط. ومن أجل الإيضاح، دعنا نبحث بعض المفاهيم والمصطلحات الأساسية.

أولاً: الكثيـر من السمات الجسـمية والنفسـية موزعة فى مـضلع تكرارى جرسى الشكل، مع تجمع معـظم الدرجات حول المتوسط، وبعد ذلك يتناقـص تدريجيًا على جانبى المتوسط (كرونبـاخ، ١٩٨٤). ويصف هذا المنحنى الجرسى المألوف أو التوزيع الاعتدالى. ومن السهل حساب متـوسط هذا المنحنى الجرسى ، بتقسيم كل الدرجات فى المجموعة المعيارية على نحو متناسب.

ثانيًا: تحـتاج هذه المهـمة إلى مقـياس معـيارى لجدولـة مدى انحراف درجـة أحد الأفراد عن المتوسط. وهكذا يكون الانحراف المعيارى مقـياساً معيارياً. إن الانحراف

المعيارى عملية إحصائية تقيس مقدار تشتت التوزيع، ويُحسب من درجات المجموعة المعيارية مثلما يمكن حساب المتوسط[1]. بمعرفة المتوسط والانحراف المعيارى، يستطيع الأخصائيون النفسيون أن يجدوا على نحو دقيق أين تقع درجة الفرد فى مضلع التكرار الجرسى.

أحد الأمثلة: نفترض أن أخصائيًا نفسياً يطبق اختبار مفردات مكونًا من عشرين كلمة على كل أطفال المدرسة الذين فى سن ٨ سنوات بمدينة تمب بولاية أويزونا. حُسب المتوسط ليكون تسع استجابات صحيحة، والانحراف المعيارى ليكون ثلاث استجابات صحيحة (انظر الشكل ١ ـ ٢).



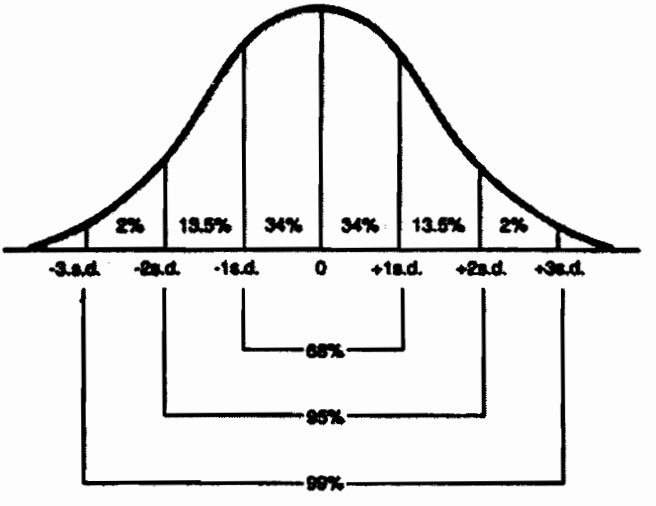
شكل (١ ـ ٢)
درجات اختبار المفردات (٢٠ كلمة)

بهذه المعلومات، يمكن تحديد موقع أى درجة لاحقة على أساس الموقع النسبى من المتوسط. حصل بوبى على ٦ استجابات صحيحة، وهو بذلك منحرف معيارياً بدرجة واحدة تحت المتوسط. حصلت چيل على ١٥ استجابة صحيحة، وهى بذلك منحرفة معيارياً بدرجتين فوق المتوسط. وحصل تيم على عشر استجابات صحيحة،

(١) الإنحراف المعيارى = $\sqrt{\frac{3\ (س - س-)^2}{ن - ١}}$

وهو بذلك منحرف معـياريًا بمقدار ٣٣ , ٠ فوق المتوسط. كل نظم الدرجـة المعيارية ـ التى يوجـد منها مـجموعـة متنوعـة ـ تفسـير أداء الفـرد، بملاحظة عدد الانحـرافات المعيارية فوق أو تحت متوسط درجة الفرد.

الانحـراف المعيـارى طريقة مـفيـدة لتبين مـدى اختـلاف درجة أحد الأفـراد عن المتوسط. فـعلى سبـيل المثال، التوزيع الاعـتدالى أو مضلع التـكرار الجرسى، يوضح التوزيع المتماثل للدرجات، حـيث يقع ٦٠٪ من الدرجات تقريبا فى نطاق الانحراف المعيارى ±١، وتقع ٩٥٪ من الدرجات فى نطاق الانحراف المعيارى ±٢، و٩٩٪ من كل الدرجات تقع فى نطاق الانحراف المعيارى ±٣ (انظر الشكل ١ ـ ٣).

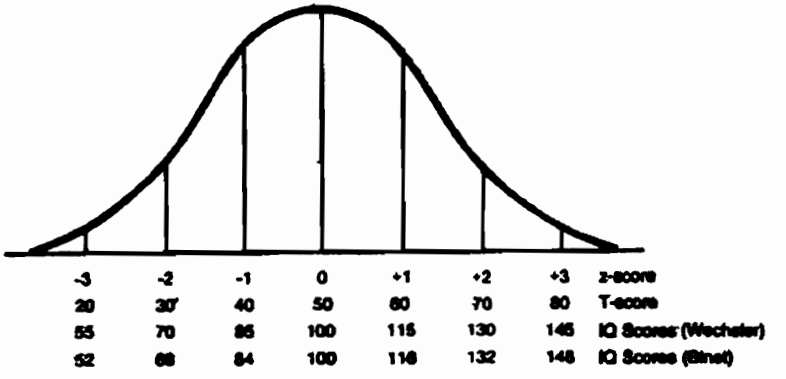


شكل (١ ـ ٣)

(الدرجات المعيارية، الدرجات التائية ودرجات نسبة الذكاء)

أدرك الأخصائيـون النفسيون ـ منذ عهـد بعيد ـ أنه سيكون من المفيـد تمثيل موقع الفرد فى مضلع تكرارى جرسى بدرجة مفردة، التى تعرف كتوزيع عينى. (الدلالة Z هى دلالة فرضيـة؛ ويمكن أن يسمى التوزيع بـ Y أو X). اصطلاحًـا، يطلق على كل الدرجات التى فى المـتوسط صفـر، كل الدرجات التى يزيد انحـرافها المعـيارى على

المتوسط بدرجة واحد + ١، وكل الدرجات التى يزيد انحرافها المعيارى عن المتوسط بدرجتين +٢، وكل الدرجات التى يقل انحرافها المعيارى عن المتوسط بدرجة ـ ١، وهلم جرا (انظر الشكل ١ـ٤).



شكل (١ ـ ٤)

وقد أنشأ خبراء إحصاء أخرون نظم تصحيح بديلة، التى يمكن أيضًا رؤيتها فى الشكل (١ـ٤). و من غير الواضح سبب ظهور العديد من نظم الأعداد المختلفة. وأحد الأنظمة الأخرى، هو نظام الدرجة التائية، والذى يطلق على كل الدرجات التى فى المتوسط ٥٠، والدرجات التى يزيد انحرافها المعيارى عن المتوسط بدرجة (+١) ٦٠، والدرجات التى يزيد انحرافها المعيارى عن المتوسط بدرجتين تساوى ٧٠ وهكذا.

إن اختبار منيسوتا للشخصية المتعدد الأوجه (هاثاواى مع ماكينلى، ١٩٦٧)، مقاييس مكارثى لقدرات الأطفال (مكارثى، ١٩٧٢) واختبار تصنيف الأطفال (بول، ١٩٩٣) تشتمل على درجات توصف كدرجات تائية. نصف بعض اختبارات التحصيل الجماعى أيضًا الدرجات فيها، كدرجات تائية، توصف درجات مقياس وكسلر للذكاء كدرجات معيارية بمتوسط ١٠٠ وانحراف معيارى ١٥.

وهذا يعنى أن الدرجات التى تزيد عن المتوسط بدرجة واحدة تساوى ١١٥،

والدرجات التى يقل انحرافها المعيارى عن المتوسط بدرجة واحدة تساوى ٨٥. وتشتمل درجات مقياس ستانفورد ـ بينيه للذكاء أيضًا على متوسط ١٠٠، ولكن انحرافها المعيارى ١٦ (ثورندايك، هاجن وساتلر، ١٩٨٦).

مزايا الدرجات المعيارية:

يفضل الأخصائيون النفسيون والآخرون الذين يعملون فى مجال الاختبار بشكل منتظم، الدرجات المعيارية على الرتب المئوية والدرجات المكافئة للعمر (أو درجات الأعمار الزمنية). ولكن لماذا تُفضل الدرجات المعيارية؟ إن الرتب المئوية والأعمار الزمنية هى بيانات ترتيبية (متدرجة) وليس بيانات ذات مستوى مئوى. فيما يتعلق بالبيانات المتدرجة مثل الرتب المئوية، لا يوجد تحديد للمسافة بين الرتب، وكل ما يمكن معرفته فقط هو أن أحد الأشخاص الذى درجته المعدلة ١٤ ـ أفضل من شخص آخر، درجته المعدلة ١٣. لا يوجد سبيل لمعرفة مقدار هذه الأفضلية؛ لأن معظم الأفراد تكون درجاتهم قريبة من المتوسط، وتتجمع معظم الرتب المئوية بالقرب من المتوسط؛ إذ تتم إضافة بضعة درجات خام فقط لدرجة شخص ما ستزيد رتبته المئوية جوهرياً.

وبالمثل، طرح بضعة درجات خام من درجة شخص ما قريب من المتوسط ستسفر عن انخفاض محسوس فى رتبته المئوية. وعلى العكس، فى النطاق الخارجى للتوزيع ـ حيث يسجل بضعة أشخاص فقط ـ فإن تغير بضعة درجات خام سيفضى إلى تغير فى بعضة رتب مئوية أو ـ فى بعض الحالات ـ لن يؤدى إلى أى تغير . بهذا المعنى، الرتب المئوية تكون غير متسقة عبر التوزيع. يُقدم الشكل التوضيحى (١ـ١) نموذجًا واقعياً لحدود الرتب المئوية.

	Raw score	Z-score	Percentile rank	
John	10	0.33	63	Initial score
(near mean)	11	0.66	75	Subsequent score
	1	0.33	12	Change
Pam	15	2.00	98	Initial score
(distant from mean)	16	2.33	99	Subsequent score
	1	0.33	1	Change

شكل (١ ـ ١)

اختبار عينة ممثلة

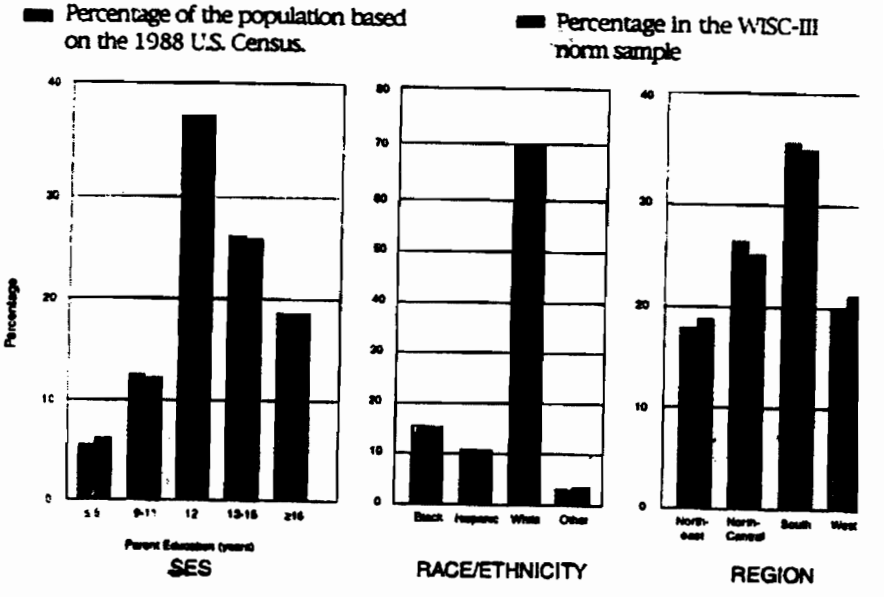
إن الدرجات المشتقة (الرتب المئوية أو الدرجات المعيارية) جديدة لتحديد موقع الطفل عند مقارنته بالمجموعة المعيارية. ولكن من الذى يشكل هذه المجموعة المعيارية؟ هل نستطيع أن نكون على وعى بتسلسل درجات أولئك الذين نختبرهم فى مكاتبنا؟ هل نستطيع أن نصنف درجات أولئك الذين يحضرون للتقييم فى مكاتب عديد من الأخصائيين النفسيين عبر الولايات المتحدة الأمريكية، وبعد ذلك يستخدمون هؤلاء الأفراد كمجموعة معيارية؟ إن مثل هذه الإجراءات ستكون غير مقبولة بالتأكيد لأن أولئك الذين يجيئون من أجل هذا القياس، من المرجح أن يكونوا مختلفين فى بعض النواحى الرئيسية عن مجموعة الأطفال العشوائية. إن مثل هذا النظام سيكون مشابهًا لتسجيل درجات حرارة كل المرضى، الذين وصلوا إلى مكتب أخصائى الأطفال، واستخدام هذا كأساس لتحديد ما هى درجات الحرارة الطبيعية وغير الطبيعية للأطفال (نصل الأطفال الذين يصلون إلى مثل هذه المكاتب قد يكونون مصابين بالحمى). إن المطلوب هو مجموعة ممثلة أو نموذجية للمقارنةبها، ويجب على المجموعة المعيارية الممثلة الملائمة أن تفى بشروط معيارين:

١ـ العشوائية أو التمثيل.

٢ـ الحجم.

إن المرضى الذين سُجلت درجات حرارتهم فى مكتب أخصائى الأطفال آنفًا لايشكلون عينة عشوائية ولا ممثلة. وقبلما أن يصنف بناة الاختبارات درجات الاختبار المتاحة، التى من المحتمل ألا تكون ممثلة، واستخدامها كمجموعة معيارية، يجب عليهم أن ينشأوا طريقة ما لأختيار الأفراد الذين يكونوا ممثلين؛ فدون عينة ممثلة، قد تكون المتوسطات غير دقيقة، وقد يعزو الطبيب السريرى السواء للدرجات التى تكون منحرفة مقارنة بكل الأطفال فى عمر معين، أو قد يعزو الانحراف للدرجات تكون طبيعية فى الواقع. إن استخدام العينات الممثلة عند اختيار مجموعة معيارية يكفل عندما تكون درجات الأطفال الفردية التى جمعت بالممارسة الإكلينيكية قد قورنت بالمجموعة المعيارية. عمليًا من النادر أن يكون ممكنًا للأطباء

السريريين أن يجمعوا عينة ممثلة من الدرجات فى منطقتهم الجغرافية. وبناء على ذلك، يعتمد الأطباء السريريون على اختبارات منشورة بعينات قومية ممثلة. انظر الشكل التوضيحى ١ـ٥ الذى يعرض مثال لكيفية اختيار مجموعة معيارية قومية ملائمة.



شكل (١ ـ ٥)
الخصائص الديموجرافية للعينات المعيارية وفقاً لمقياس ويكسلر النسخة الثالثة (١٩٩٠)

لا يجب أن تكون المجموعة المعيارية عشوائية وحسب، ولكن يجب أن تكون أيضًا كبيرة بقدر كاف. إذا تكونت المجموعة المعيارية من ثلاثة أطفال فقط يبلغون من العمر ثمانى سنوات، فإنه من الممكن أن يكون الأطفال الثلاثة المختارون نقطة فوق المتوسط أو نقطة تحت المتوسط، وهكذا يتسببون فى رؤية منحرفة لما يكون عليه المتوسط بالعقل. وإذا تم اختيار ٥٠ طفلاً فى سن الثامنة عشوائيًا، عندئذ ستقل فرصة سوء الحظ فى الأختيار بمقدار كبير، إن بعض الاختبارات المنشودة ـ كما سنرى فى فصول لاحقة ـ تحتوى على مجموعات معيارية صغيرة بدرجة غير مقبولة أو غير ممثلة. يحد هذا القصور بشدة من فائدة هذه الاختبارات الإكلينيكية.

أما أولئك الذين يرتابون فى المطابقة العملية لاختيار عينة ممثلة وذات تقنين معاصر، فقدأثار فلين (١٩٨٤) مسألة أن المجتمع الأمريكى يزيد ذكاءً عبر الزمن تقدم سببًا للتأمل. اقترح فلين أن نسبة الذكاء التى تكتسب عبر الزمن يبلغ معدلها ٠,٣ درجة تقريبًا لكل سنة. وقد قرر البعض أن يسمى هذه الدرجات المكتسبة «تأثير فلين» (نيزر وآخرون، ١٩٩٦)، وهو تأثير له مضامين واضحة، كما نرى فى شكل (١ ـ ٥). حتى الأدوات المقننة بدقة ـ مقاييس ذكاء وكسلر ـ أُعيد تقنينها كل ١٨ إلى ٥ سنة فقط. وكما يُظن، إعادة تقنين الاختبار تسفر عن درجات ستكون من ٥ إلى ٨ درجات أقل.

بمعنى، إذا أُستخدمت نسخة مقياس وكسلر التى أُذن لها بالنشر توًا وليس النسخة التى تم سحبها، عندئذ فإن الفرد الذى سيطبق عليه المقياس ستكون درجاته أقل بنسبة من ٥ إلى ٨ درجًات. إن هذه الفروق تكون ذات معنى عميق فى أطراف توزيع نسبة الذكاء حيث تتخذ قرارات التعرف، التى قد تؤثر فى التسكين المدرسى. فعلى سبيل المثال، استخدام اختبار مقنن حديثًا ستنقل العديد من الأفراد الذين أدخلتهم درجاتهم فى السابق من نطاق نسبة الذكاء ٧٠ ـ ٦٤ إلى نطاق نسبة الذكاء ٦٥ ـ ٦٩، حيث سترتفع فرص تشخيصهم على أنهم مصابون بالتخلف العقلى. وعلى الطرف الآخر من النطاق، قد يجد الأفراد الذين تتراوح درجات نسبة ذكائهم بين ١٣٠ ـ ١٣٤ أن درجاتهم قد هبطت إلى ١٢٥ ـ ١٢٩، عندما اختبروا بالنسخة الجديدة. إن مثل هذا الهبوط فى درجة نسبة الذكاء قد يستبعد أفرادًا من المشاركة فى برنامج للمتعلمين الموهوبين أو المتقدمين.

تحرى الاختبارات النفسية

إن الاختبارات النفسية الموضوعية القابلة للقياس الكمى لها مزايا أخرى على القياس غير النظامى. ويمكن أن يكون للاختبارات النفسية المنشورة قيمتها التى قُدّرت بالطرق الأمبيريقية والعلمية، بقياس ثباتها وصدقها.

الثبات

يمكن أن تعتبر الاختبارات النفسية ـ بوجه عام ـ على أنها محاولة لقياس السمات

(مثل: الذكاء، الاستعداد الرياضى (الحسابى). أحد أهداف القياس هو أن يقيس هذه السمات بدقة قدر الإمكان. استنبط أصحاب نظريات القياس العقلى ـ منذ عهد بعيد ـ منطقًا بشأن الكيفية التى يمكن بها قياس السمات بإتقان. (من أجل الإطلاع على مناقشة رائعة، انظر كابلان ساكوزو، ١٩٨٢). فقد حاولت هذه المناقشة أن تبرهن على أن الاختبارات النفسية لا تقيس السمات بدقة تامة، حيث يكتنفها دائمًا خطأ ما. وهذا الخطأ يمنع درجة الاختبار من أن تكون انعكاسًا صحيحًا للدرجة الحقيقية للسمة التى تحت البحث. الدرجة الحقيقية الفرضية موجودة، ولكن لا يتم الحصول عليها من خلال القياس. يهتم الثبات بمدى الخطأ الذى اشتمل عليه القياس، أو مدى بُعد الدرجات الناتجة عن الدرجة الحقيقية.

وإحدى الطرق لتلافى مشكلة الخطأ هو أن نجرى عدة قياسات أو قياسات متكررة للسمة نفسها، وليس قياسًا واحدًا وحسب. إذا كانت عوامل مثل التعب والدافعية والألفة بالممتحن تسبب فى الأخطاء، عندئذ ينبغى أن تسمح القياسات المتكررة لهذه العوامل أن تبطل تأثير كل منها الآخر. وفى حالة مثل هذه، يمكن اعتبار متوسط قياسى التقدير الأفضل للدرجة الحقيقية. إن القياسات المتكررة سوف لا تسمح فقط بتقدير للدرجة أكثر دقة، ولكن أيضًا ستقدم دليلاً لإجراءات قياس الثبات.

الخطأ المعيارى للقياس

إن القياس مرة وإعادة القياس بالأداة نفسها بعد ذلك (يُسمى الثبات بإعادة الاختبار)، يمكن من استنباط عمليتين إحصائيتين هامتين: العملية الإحصائية الأولى هى الخطأ المعيارى للقياس، والعملية الإحصائية الثانية هى معامل الثبات. ينتج الخطأ المعيارى للقياس بإجراء الاختبار وإعادة إجراء الاختبار وبعد ذلك حساب مدى قرب اتفاق الدرجتين . (بالمصطلحات النظرية، الخطأ المعيارى للقياس هو الانحراف المعيارى للدرجات المستخلصة بهذا الإجراء لشخص، واحد إذا تم اختبار ذلك الشخص عدة مرات). سيكون الخطأ المعيارى صغيراً إذا كانت الدرجات متماثلة فى التطبيقات المتكررة. إن هذا سيدل ضمنًا على أن كل قياس يجب أن يكون قريبًا تمامًا من الدرجة الحقيقية، ومن ثم دون خطأ كثير، ومن ثم يكون الاختبار ثابتًا. ويشير

الخطأ المعيارى الكبير إلى أن ا لدرجات اختلفت إلى حد بعيد فى التطبيقات المتكررة. وفى هذه الحالة لا يكون أى من القياسات قريبًا من الدرجة الحقيقية؛ ومن ثم يتم قياس خطأ كثير، ويكون الثبات ضعيفًا. وللخطأ المعيارى للقياس استعمال عملى. إذا استخدام المنطق الذى استنبط رياضياً من قبل منظرى القياس، تسمح إحصائية الخطأ المعيارى للأخصائين النفسيين ليكتشفوا مدى قرب الدرجة المرصودة بالدرجة الحقيقية، «فى نطاق حدود معينة من الثقة» على سبيل المثال، مقياس وكسلر لذكاء الأطفال ـ النسخة الثالثة يعتبر أداة ثباتها عاليًا. والخطأ المعيارى للقياسى لنسبة ذكاء المقياس الكامل هو ٣ تقريبا، وهذا يعني أن ـ باستخدام المنطق المستنبط سابقًا ـ الفرص هى وجود اثنتين من ثلاث مرات، تكون فيها الدرجة الحقيقية فى نطاق ±٣ درجات من الدرجة النائجة. وباستخدام اختبار أقل ثباتًا، والذى يحتوى على خطأ أكثر، سيكون الخطأ المعيارى أكبر.. فعلى سبيل المثال، اختبار بيبودى للصور والمفردات ـ المعدل (دون & دون، ١٩٨١) يحتوى على خطأ معيارى للقياس بمقدار ٧، و هذا يعنى أن الفرص هى اثنين من ثلاثة، وأن الدرجة الحقيقية هى فى نطاق ±٧ من الدرجة المرصودة باستخدام هذا الاختبار. إن الخطأ المعيارى ـ بذلك ـ هو عملية إحصائية مفيدة للغاية لاتخاذ قرارات عملية بشأن مدى الثقة التى توضع فى درجة ما

معامل الثبات

الدليل العام للثبات هو معامل الثبات، و يمكن أن يُحسب معامل الثبات باستخدام بيانات إعادة الاختبار نفسها، التى استخدمت لحساب الخطأ المعيارى للقياس الذى ذكر آنفًا.

ويشتمل تحديد الثبات على ربط درجة كل فرد فى التطبيق الأول لاختبار ما بالدرجة فى التطبيق الثانى. إذا اتفق التطبيقان، عندئذ يُفترض أن أداة القياس تستبين الدرجة الحقيقية، ويكون الخطأ زهيداً. غير أنه إذا لم يتفق التطبيقان، عندئذ يُفترض أنه سيتم قياس خطأ كبير؛ وبناء على ذلك، تعانى أداة القياس مشكلات فىّ الثبات؛ لأن معامل الثبات يُعبر عنه كمعامل ارتباط، فإن الدليل يمكن أنه يتراوح من ١، مشيراً إلى اتفاق تام بين التطبيقين، إلى صفر مشيراً إلى عدم وجود علاقة. تقع معظم معاملات الثبات فى مكان ما بين هاتين القيمتين، بدرجات تتراوح من ٠,٨٠ إلى

٩٠ , ٠ وهى تعتبر عمومًا مقبولة للاستخدام العملى. بالإضافة إلى ذلك، توجد طرق أخرى لحساب معاملات الثبات. تشتمل إحدى الطرق على ربط صيغتين منفصلتين لنفس الاختبار كل منهما بالآخر؛ وتشتمل طريقة أخرى على ربط نصف بنود اختبار بالنصف الآخر من نفس الاختبار. تفهرس هذه الطرق رقميًا كيف يتم قياس السمات بدقة، بإيجاد معامل الثبات (يتراوح من ١ إلى صفر) مثلما حدث فى طريقة إعادة الاختبار.

توجد مسألة أخرى يجب أن نطرحها بشأن الثبات: عندما تتساوي كل العوامل الأخرى، تكون الاختبار الطويلة أكثر ثباتًا من الاختبارات القصيرة (ليمان، ١٩٧١). تحتوى ا لاختبارات الطويلة على بنود كافية بحيث إذا أغفل الطفل بندا واحد لا يؤثر على درجة الطفل الكلية. الاختبار الأطول يحتوى على بنود أكثر، والذى يكون مكافىء لإجراء قياسات متكررة وإيجاد متوسط الدرجات. تقدم هذه الممارسة للقياسات المتكررة ـ كما نوقشت سابقًا ـ تقدير أفضل للدرجة الحقيقية وبذلك ثبات أفضل. كما سنرى فى الفصول اللاحقة، توجد فروق أساسية فى ثبات الاختبار المنشورة. يكون ثبات بعض الاختبار الطويلة جيد، بينما بعض الاختبار القصيرة كانت ناقصة جدًا لدرجة أنها تدعى من قبل أهل المهنة «سريعة وكثيرة الأخطاء».

الصدق

على الرغم من أن الثبات هام، إلا أن فحصه وحده لا يعتبر دليلاً على قيمة الاختبار. فعلى سبيل المثال، إدعاء أن اختبار يقيس الذكاء يمكن أن يكون اختبار ثابت (يقيس سمة ما باستمرار ودون خطأ) لكن ليس له علاقة على الإطلاق بالذكاء، وهو السمة المفترض أن يتم قياسها. يريد الاخصائيون النفسيون أن تكون اختباراتهم ثابتة وصادقة. الصدق يجيب عن السؤال «هل يقيس الاختبار ما يدعى أن يقيسه؟».

توجد عدة أنواع من الصدق.

(صدق المحتوى) تتكون هذه الطريقة لقياس الصدق من فحص بنود الاختبار وإصدار حكم بشأن عما إذا كانت بنود الاختبار تقيس ما تدعى أنها تقيسه، فعلى سبيل المثال، افترض أن باحثين ابتكروا اختبار للذكاء العام، والذى يتكون كلية من بنود بالأستدلال الرياضى، فعلى الرغم من أن الاستدلال الرياضى هو أحد جوانب

الذكاء، إلا أن الاختبار لا يحتوى على بنود مختلفة بالقدر الكافى ليعتبر اختباراً للذكاء العام. سيكون صدق المحتوى لهذا الاختبار منخفضًا كقياس للذكاء. وقد يكون صدق المحتوى فيه ملائمًا كقياس للاستدلال الرياضى الأساسى، أو قد يصبح قياسًا ملائمًا للذكاء اللفظى إذا كانت بعض البنود تتعامل مع المتشابهات اللفظية، وأن تضاف إليه بنود خاصة بالمفردات. إن صدق المحتوى هام على نحو خاص لاختبارات التحصيل، وكمقابل لاختبارات القدرات ـ التى يُفترض أنها الأساس للفرد الذى يُختبر تقدم استعداداً ما لبنود الاختبار ـ يكون لاختبارات التحصيل معنى فقط، عندما يتعرض الأفراد الذين يتم أختيارهم للتعلم عن موضوع الاختبار. إن طلبة الصف السابع الذين درس لهم فى مقرر العلوم عن الطقس، والمركبات الكيميائية الأولية، والحشرات سيكونون فى ظروف معاكسة إذا تعاملت بنود الاختبار مع علم تشريح الإنسان والجيولوچيا. بالنسبة لهؤلاء التلاميذ، فإن صدق محتوى الاختبار المؤلف من بنود عن علم التشريح والجيولوجيا سيكون هزيلاً؛ لأن الاختبار لن يقيس المعلومات التى تعلمها الطالب. وقد يكون صدق محتوى الاختبار نفسه ملائمًا بالنسبة لمدرسة بها مقرر علوم مختلف، يركز على علم التشريح والجيولوچيا.

لأن اختبارات التحصيل المنشورة مُصممة للاستخدام القومى، لا يمكن أن تلائم تمامًا مقررات المدارس المحلية. مع ذلك، يستطيع الأخصائيون النفسيون بالمدارس (أو المشخصون الآخرون الذين يقيمون المهارات الأكاديمية) أن يُقْدموا على أن يتحققوا من أن اختبارات التحصيل المنشورة والتاريخ التعليمى لكل طفل فى حالة اتفاق، باستعراض الأهداف التعليمية ومحتوى اختبارات التحصيل. إذا لم يكن من الممكن إيجاد مطابقة ملائمة للتاريخ التعليمى بمحتوى الاختبار، عندئذ يتضاءل صدق درجات اختبار التحصيل بالنسبة لطفل معين. إن الدرجات ـ مع ذلك ـ مفيدة (يوجد منهج محورى يمكن إدراكه على نحو كاف فى الولايات المتحدة ليسمح للاختبارات القومية أن تُعد بصدق ما). وحتى إذا أُستخدمت اختبارات تحصيل مقننة، فإن المعلمين يستخدمون فى الغالب القياسى غير الرسمى القائم على أساس

المنهج؛ ليضمنوا اختبار عينة ملائمة لتعليم المدارس المحلية. وقد يحدث مثل هذا القياس فى مستهل العملية التشخيصية، وقد تستمر طالما يدرس الطفل. إن هذا القياس غير الرسمى هو قياس تكميلى للقياس الرسمى المتعلق بالقياس النفسى ويؤدى وظيفة مختلفة بعض الشىء لتوجيه ومراقبة التعليم. والمناقشة التفصيلية لموضوع القياس التعليمى غير الرسمى والقائم على أساس المنهج تتجاوز الغرض من هذا الكتاب. (من أجل المزيد من المعلومات بهذا الشأن، انظر شين: ١٩٨٩).

الصدق التكوينى

إن الصدق التكوينى ـ مثل صدق المحتوى ـ يصدر حكمًا؛ ولكنه مخالف لصدق المحتوى، فى أنه يعتمد على منطق أكثر تعقيداً وملاحظات، وليس على الفحص البسيط للبند. يبدأ الصدق التكوينى بمتغير استكشافى، أو مفهوم، مُصمم لتفسير السلوك. وبعض الأمثلة لمثل هذه المفاهيم: الذكاء، الاكتئاب والقلق. وكل من هذه المفاهيم ـ طبقًا للمشتغل بالأمور النظرية الذى نشرها ـ يُعتقد أنها تتضمن خصائص معينة: إن الذكاء الذى يتنبأ بالتحصيل فى مجموعة متنوعة من المواقف، ليس كلية نتيجة للخبرة، وثابت؛ يرتبط القلق بدرجة من الضغط وغير ثابت عبر الزمن. ويمكن فحص الصدق التكوينى للأداة بتكوين تجارب علمية لرؤية كيف تتغير الدرجات تحت ظروف مختلفة. فعلى سبيل المثال، يُتوقع أن ترتفع الدرجات فى اختبار قلق، إذا وضع المفحوصين فى مواقف ضاغطة (مثلاً، إذا قيل لهم إنهم على وشك إجراء امتحان صعب فوراً). بالمثل، يُتوقع أن تقل الدرجات فى اختبار قلق إذا تناول المفحوصين عقاراً من خواصه المعروفة أن يخفض القلق. إذا وجد الباحثون أن هذه الاستجابات لم تحدث تحت ظروف محددة، عندئذ سيتحسن الصدق التكوينى لاختبار القلق؛ لأنه سيكون من الواضح عندئذ أن الاختبار فى الواقع يقيس القلق. بناء على ذلك، يشتمل الصدق التكوينى على حكم، لكن الأحكام ليست عادلة التى استخدمت لتقييم صدق المحتوى.

الصدق التنبؤى والتلازمى:

هما طريقتان لإثبات الصدق من خلال ربط الاختبارات بمعايير مستقلة. ويفترض أن الاختبارات النفسية تتنبأ بالمعايير المستقلة، الخارجية التى قد تتضمن سلوكيات معينة (مثلاً، محاولات الانتحار)، السمات العامة (مثلاً، درجات فى اختبارات نفسية أخرى)، أو نتائج مستقبلية معينة (مثلاً، التخرج من الكلية). وتعتبر اختبارات نسبة الذكاء مثالاً كلاسيكياً للأدوات التى من المتوقع أن تتنبأ بالمستقبل. وقد تم تصميم اختبارات نسبة الذكاء فى الأصل لتحديد أى من التلاميذ المبتدئين من المرجح أن يكونوا ناجحين (أولئك الذين ينبغى أن يستمروا فى المنهج الأكاديمى العام) وأولئك الذين يُتوقع لهم الفشل (أولئك الذين سيوجهون إلى منهج بديل). ويمكن إظهار قيمة هذه الاختبارات التى تقيس نسبة الذكاء إمبيريقيًا. بإثبات أنها تتنبأ بالتحصيل الدراسى. ويمكن أن يتم جمع الأدلة بسهولة وذلك بحساب معامل يرتبط بدرجات نسبة الذكاء للتلاميذ، عندما يبدأون الدراسة، بتحصيلهم اللاحق الذي يقاس بدرجات اختبار التحصيل، درجات الفصل، أو تقديرات المدرس. وكلما كان الارتباط أعلى، كانت قدرة الاختبار التنبؤية أكبر، وكان تصنيف الاختبار أكثر صدقًا. إن جوهر الصدق التنبؤى هو أن درجات الاختبار المتداول تتنبأ عن معيار مستقل ما فى المستقبل.

يتطابق الصدق التلازمى تقريبًا مع الصدق التنبؤى، فيما عدا أن معيارًا خارجيًا يُقاس فى الوقت نفسه، الذى يُطبق فيه الاختبار. يُستخدم الاختبار الموجود قبلاً أو الاختبار المصمم على نحو جيد فى الغالب كمعيار (محك) يرتبط به الاختبار الجديد. فعلى سبيل المثال، قد يُعد أحد الباحثين مقياس ذكاء خاصاً به، ولكى يثبت الباحث قيمة هذا الاختبار، يقوم بتطبيقه على مجموعة من تلاميذ المرحلة الابتدائية بجانب اختبار معيارى، مسلم به كمقياس صادق للذكاء، مثل مقياس وكسلر. إذا ارتبط مقياس الذكاء الجديد بالاختبار المعيارى المعترف به بدرجة عالية، يكون الباحث قد أثبت قيمة الاختبار بالصدق التلازمى. من الواضح أن الصدق التلازمى يعتمد على مطابقة محك خارجى مناسب، مثل مقاييس وكسلر. من الوجهة التاريخية، كان البحث فى فائدة الاختبارات النفسية من هذا النوع ـ الارتباطات بين

الاختبارات النفسية ومحك ما آخر، والذى يمكن أن يُقاس كميًا على طول سلسلة متصلة، مثل متوسط درجة الفرقة الدراسية أو الدرجات فى اختبار نفسى آخر. فى الثمانينيات والتسعينيات، ثم استخدمت الاختبارات النفسية فى مجالات أخرى من البحث العلمى؛ حيث تم تصنيف الحالات إما موجودة أو غير موجودة، وليس على طول سلسلة متصلة (بمصطلحات القياس يشار إلى الأول كمتغير مستقل، وإلى الأخير كمتغير تابع). إن اضطراب نقص الانتباه / فرط النشاط، هو حالة تلقى مقداراً كبيراً من الأهتمام بين المعلمين والجمهور العام، هى حالة مثل هذه. وبعد تقييم يقوم به شخص خبير، يمكن أن يحكم على الأطفال أنهم مصابون بالاضطراب أو غير مصابين به بطريقة نعم / لا. ابتداء من عام ١٩٩٧، تم تحديد اضطراب نقص الانتباه/ فرط النشاط من خلال رأى خبير، بناء على قائمة واضحة من الأعراض؛ الاختباراتِ النفسية ليست الأساس لتشخيص اضطراب نقص الانتباه/ فرط النشاط. وليس من المستغرب، أن يبحث الباحثون لتحديد عما إذا كانت الاختبارات النفسية للانتباه أوْ ضَبط الاندفاع لها دور فى مساعدة الأطباء الإكلينيكيين على تقرير تشخيص مَا لاضطراب نقص الانتباه / فرط النشاط. تظهر دراسة أجراها باركلى وجرودزينسكى فى عام ١٩٩٤ كيف تستخدم الاختبارات فى هذا الشأن، وكيف يمكن تقديم هذا النوع من معلومات القياس لمساعدة المُشخصين فى قدرتهم على اتخاذ القرارات اليومية.

وقد درس باركلى وجرودزينسكى ٤٧ طفلاً ـ منهم ٢٤ مصابون باضطراب نقص الانتباه / فرط النشاط والـ ٢٣ الأخرين غير مصابين بهذا الاضطراب ـ الذين تم تقييمهم من خلال عدة اختبارات قادرة على تحديد اضطراب نقص الانتباه / فرط النشاط. وقد أجملت نتائج الاختبارات فى نوعين من الاحتمالات لتيسير المقارنات بين الاختبارات: النوع الأول من الأحتمال هو عما إذا كانت درجة غير نمطية فى اختبار نفسى ترتبط على نحو ثابت بوجود اضطراب نقص الانتباه / فرط النشاط، ويُسمى هذا النوع القدرة التنبؤية الإيجابية، ويُحسب بتعيين نسبة قائمة على أساس عدد الأطفال، الذين حصلوا على درجة غير نمطية فى اختبار نفسى كقيمة للمقام، وعدد هؤلاء الأطفال الذين تم تشخيصهم بأنهم مصابون باضطراب نقص الانتباه / فرط النشاط كقيمة للبسط.

هذا هو احتمال أن أحد الأطفال مصاب بالفعل باضطراب نقص الانتباه / فرط النشاط، إذا أشارت نتائج الاختبار أن اضطراب نقص الانتباه / فرط النشاط موجود. والاحتمال الثانى يُسمى القدرة التنبؤية السلبية، الذى مؤداه أنه إذا كانت درجة اختبار عادية مرتبطة على نحو ثابت بعدم وجود حالة ما، ففى هذه الحالة ينعدم وجود اضطراب نقص الانتباه/ فرط النشاط. وتحسب القدرة التنبؤية السلبية بتقرير نسبة قائمة على أساس عدد الأطفال، الذين حصلوا على درجة عادية فى القياس النفسى كقيمة للمقام، وعدد أولئك الأطفال الذين تم تشخيصهم على أنهم غير مصابين باضطراب نقص الانتباه/ فرط النشاط كقيمة للبسط.

والنتيجة هى احتمال عدم الإصابة باضطراب نقص الانتباه/ فرط النشاط عندما تشير درجات الاختبار إلى عدم وجود اضطراب نقص الانتباه/ فرط النشاط. إن الدرجات المنظمة بهذه الطريقة (أى، كقدرة تنبؤية إيجابية وقدرة تنبؤية سلبية) تكشف عن أنماط لا يمكن أن تكتشف من عمليات إحصائية أخرى، تستخدم فى الغالب بالاشتراك مع الاختبارات النفسية، مثل الارتباط بين درجات الاختبار، وتشخيصات اضطراب نقص الانتباه/ فرط النشاط أو متوسط قيم الاختبارات النفسية المختلفة للأطفال المصابين وغير المصابين باضطراب نقص الانتباه/ فرط النشاط. فى دراسة باركلى وجرودزينسكى، كل الاختبارات تقريبًا لم تكن ذات قدرة تنبؤية إيجابية أو قدرة تنبؤية سلبية؛ لتكون ذات فائدة إكلينيكية. أما عن الاختبارات التى نجحت فى أداء هذا العمل، يجب أن تستخدم نتائجها بحذر؛ لأن الدرجة غير النمطية والدرجة النمطية لا تساعد المُشخص على حد سواء.

وقد وجد باركلى وجرودزينسكى أن اختباراً ما لتركيز وللإنتباه وضبط الاندفاع، طريقة جهاز جوردون التشخيصة لمهمة الانتباه (انظر الفصل ٩)، كانت له قدرة تنبؤية إيجابية رائعة، تعنى احتمال إصابة الأطفال ذوى الدرجات غير العادية فى أحد أبعاد نظام جوردون التشخيصى ووجود احتمال إصابتهم باضطراب نقص الانتباه مع فرط النشاط بنسبة ٩٢٪. وحتى النتائج الأقوى أرتبطت ببعد آخر (أخطاء العمل) من أبعاد نظام جوردون التشخيصى. فى عينة أطفال باركلى وجرودزينسكى، كان هناك احتمال بنسبة ١٠٠٪ للإصابة باضطراب نقص الانتباه/ فرط النشاط، إذا

وجدت درجة غير عادية فى بعد أخطاء العمل. وقد يفترض المرء على الفور أن نظام جوردون التشخيصى لمهمة الانتباه هو اختبار قيم للغاية لاكتشاف اضطراب نقص الانتباه/ فرط النشاط.

غير أنه قد يثبت أن قيمة القدرة التنبؤية السلبية لنظام جوردون التشخيصى لمهمة الانتباه أقل نفعًا. فعلى سبيل المثال، وجد باركلى وجرودزينسكى أنه إذا تلقى طفل درجة عادية فى إجمالى الصحيح، فإنه لديه فرصة بنسبة ٦٣٪ فى أنه لا يصاب باضطراب نقص الانتباه / فرط النشاط (ولكن توجد فرصة بنسبة ٣٧٪ أن يصاب باضطراب نقص الانتباه/ فرط النشاط). أما البعد الآخر لهذا الاختبار ـ أخطاء العمل ـ فقد تم تحديد القدرة التنبؤية السلبية بنسبة ٥٩٪ (٤١٪ من ذوى درجات اختبار عادية مصابين باضطراب نقص الانتباه/ فرط النشاط).

إن اختبار مثل نظام جوردون التشخيصى ـ فى الممارسة الإكلينيكية ـ له فائدة بالتأكيد، ولكن الاختبار يكون له قيمة فقط، إذا فهم القائم بالتشخيص مزايا وعيوب الاختبار. تنبىء الدرجة غير النمطية أن اضطراب نقص الانتباه / فرط النشاط مرجح بدرجة كبيرة؛ الدرجة النمطية تخبر عن معلومات أقل بكثير. غير أن القضية هى أن الاختبارات النفسية يمكن أن تكون مفيدة للغاية فى اتخاذ القرارات اليومية بل أن بحث صدق اختبار ما ـ محدود باستخدام مقصود ـ هو ضرورته عمومًا للتطبيق الأكلينيكى.

حتى نتائج مثل هذه التى تم الحصول عليها تحت ظروف بحث صيغت بدقة، قد لاتكون قابلة للتطبيق مباشرة، تحت ظروف حياة حقيقية أخرى مختلفة قليلاً. (إلوود، ١٩٩٣). فى هذه الدراسة كان نصف المفحوصين على وجه التقريب مصابين باضطراب نقص الانتباه/ فرط النشاط، والنصف الأخر من المفحوصين من غير المصابين بهذا الاضطراب. وبهذا، سيكون احتمال سحب أى طفل مصاب باضطراب نقص الانتباه/ فرط النشاط عشوائيًا فى الدراسة ٥٠٪. هذا الاحتمال القبلى، أو المعدل الأساسى للاضطراب بين أولئك الذين تحت البحث، هو أيضًا احتمال حرج فى تحديد مدى الثقة التى توضع فى نتائج الاختبار (وفى الواقع فى كل أنواع المعلومات التشخيصية). وعلى الرغم من أنه قد توجد قدرة تنبؤية إيجابية

بنسبة ١٠٠٪ بين الأطفال، عندما يُقسم تكوينهم على نحو متعادل بين الأطفال المصابين باضطراب نقص الانتباه/ فرط النشاط والأطفال غير المصابين بهذا الاضطراب (المعدل الأساسى لاضطراب نقص الانتباه/ فرط النشاط ٥٠٪)، فماذا سيحدث إذا تم فحص مجموعة غير منتقاة من أطفال فى سن ٨ سنوات، مثل كل الطلبة فى سن ٨ سنوات من مدرسة حكومية نمطية، باستخدام أسلوب جوردون التشخيصى/ تحت هذه الظروف، سيكون الاحتمال القبلى للإصابة باضطراب نقص الانتباه/ فرط النشاط ٣٪ تقريبًا (المعدل الأساسى للإصابة باضطراب نقص الانتباه/ فرط النشاط فى المجتمع العام)، وليس ٥٠٪. تحت هذه الظروف من السهل أن ندرك أن النتيجة الإيجابية ـ فى كل الاحتمالات ـ فى أسلوب جوردون التشخيصى ستكون أقل كثيراً فى الثقة بأن الطفل مصاب باضطراب نقص الانتباه/ فرط النشاط مما هو حقيقي فى دراسة البحث. ومع كل ذلك، معظم الأطفال غير مصابين باضطراب نقص الانتباه/ فرط النشاط، وقد تكون النتيجة الإيجابية قد تكون خطأ. إن قضية الاحتمال القبلى ـ أو المعدل الأساسى ـ قد نوقشت منذ زمن طويل (مثلاً، ميهل مع روزن، ١٩٥٥) ولكنه مُورس على نحو غير كامل لأن الأخصائيين النفسين كانوا دائمًا يفتقرون إلى المعلومات عن الاحتمال القبلى لنتيجة معينة. ولا يزال الأخصائى النفسى الحكيم يتأمل هذه الأنواع من المعلومات عندما يحاول أن يقرر تشخيصات. كما رأينا فى قضية التنبؤ بالانتحار التى نوقشت فى الفصل الثانى، لا تكون كل النتائج مقبولة على حد سواء، وبذلك يجب إضافة الحكم الإكلينيكى والقيم الإنسانية إلى الاعتبارات الإحصائية التى نوقشت آنفًا.

عناوين الاختبارات ومضامينها

تزعم الاختبارات النفسية أنها تقيس مصفوفة مدهشة من السمات والمهارات والقدرات. وللاختبارات عناوين تتراوح من «التميز السمعى» إلى «اللغة المكتوبة». لسوء الحظ، عنوان الاختبار يكون فى الغالب مرشداً غثًا فيما يتصل بما يقيسه ذلك الاختبار أو جزء منه. يحدث فى الغالب خطآن فى التصور، عندما يُعتقد أن محتوى الاختبار يُعرف بعنوانه. الخطأ الأول فى التصور هو الخلط (Jingle Fallacy) ـ الفكرة الخاطئة بأن الاختبارات ذات العنوان المتماثل تقيس الشىء نفسه ـ والخطأ

الثانى هو خطأ الارتباك (JangleFlallacy) ـ الاختبارات ذات الأسماء المختلفة تقيس أشياء مختلفة (ميسيك، ١٩٨٣). فعلى سبيل ا لمثال، كل من اختبار كوفمان للتحصيل التعليمى (كوفمان وكوفمان، ١٩٨٥) واختبار بيبودى المعدل للتحصيل الفردى يحتويان على اختبارات فرعية للتهجى. هذه الاختبارات الفرعية التى لها العنوان نفسه توضح خطأ الخلط.

فى اختبار كوفمان للتحصيل التعليمى، يُطلب من الطفل أن يتهجى الكلمات التى تُملى عليه باستخدام قلم رصاص وورقة. فى اختبار بيبودى المعدل للتحصيل الفردى، يختار الطفل التهجئة الصحيحة للكلمة من أربعة احتمالات يقوم الممتحن بنطقها. وقد يحاول المرء أن يثبت سواء كان كل من هذين الاختبارين يقيسان التهجئة؛ القليل سيوافق على أن كلاً منهما يقيس المهارة نفسها ، برغم عنوان التهجئة الملصق بهما. وقد تفهم مغالطة الارتباك بتأمل ثلاثة اختبارات، لها عناوين تدل على أنها تقيس أشياء مختلفة. يمكن أن يقاس الذكاء، التحصيل واحتباس الكلام بأسئلة اختبار متشابهة ـ إن لم تكن فى الواقع متطابقة ـ اعتماداً على نوع الاختبار واعتماداً على قصد الباحث. فعلى سبيل المثال، الاختبار الفرعى «المفردات التعبيرية» من اختبار كوفمان المختصر للذكاء (كوفمان وكوفمان، ١٩٩٠)، والاختبار الفرعى «التعبير الشفوى» من اختبار وكسلر للتحصيل الفردى (الجمعية النفسية، ١٩٩٢)، «التسمية البصرية» من فحص احتباس الكلام المتعدد اللغة (بنتون، هامشر مع سيان، ١٩٩٤). كل من هذه الاختبارات يحتوى على بنود تقتضى من الطفل ـ جزئيًا ـ أنه يسمى ما يراه فى الصورة. فطبقًا لعناوين الاختبار وحدها، قد يعتقد المرء ـ على نحو خاطئ ـ أن الاختبارات تقيس ثلاث خصائص مختلفة.

لأن الاخصائيين النفسيين يستخدمون الاختبارات ليتخذوا قرارات بشأن الأطفال، فإنه غنى عن القول أنه يجب على الاخصائيين النفسيين أن يكونوا على وعى بخصائص الاختبار الهامة، مثل الثبات والصدق. ومن المهم بالقدر نفسه بالنسبة للأخصائين النفسيين أن يكونوا على وعى بما يقيسه كل اختبار بدقة. و يجب على مستهلكى معلومات الاختبارات النفسية أن يدركوا أن تفسير الاختبار ليس دائمًا سهلاً بالقدر الذى يبدو عليه. إن أخذ أسماء الاختبارات وحسب على القيمة

الظاهرية يؤدى فى الغالب إلى الوقوع ضحية للأخطاء، كما سنرى فى الفصل التاسع، لكى يؤدى الاخصائيون النفسيون مسئولياتهم المهنية، يجب عليهم أن يقوموا بما هو أكثر من تطبيق الاختبارات وإعلان الدرجات. وعملية التفسير مطلوبة إذا كنا سنستخدم نتائج الاختبارات لمساعدة الأطفال. يتحير أحيانًا أولئك الذين يتلقون تقارير نفسية مكتوبة؛ لأن أسماء الاختبارات والقيم الرقمية المرتبطة بها لا تتفق مع الافتراضات الحدسية. ربما فى هذه النماذج يقوم الأخصائى النفسى بتفسير يتجاوز أسماء الاختبارات البسيطة، ويتجنب أخطار Jangle jingle Fallacy.

القياس النفسى كعلم

على الرغم من أن المناقشة سالفة الذكر تحدد فقط مبادئ القياس النفسى الأساسية، إلا أنه ينبغى أن تظل المضامين واضحة. إن معظم القياس النفسى هو علم. تركز الاختبارات على السلوك القابل للملاحظة، الموضوعى والذى يمكن قياسه كميًا. إنهم يجدون للحصول على نتائج يمكن أن تخضع للفحص من قبل الأقران. تنشر البحوث فى القياس النفسى على شكل تقارير فى مجلات متخصصة محكمة ذات معايير موضوعية. المبادئ والإجراءات المرتبطة بالقياس راسخة فى نظام مدروس بدقة، يُسمى النظرية السيكومترية (نونالى، ١٩٧٨). يجمع البحث فى الغالب بين الملاحظة، والمنطق، والطرق الإحصائية الفعالة ـ المواكبة للتطور العلمى.

كما سترى فى الفصول اللاحقة ، تستخدم بعض أدوات القياس كل أفضل الإجراءات التى حددت حتى الآن. فعلى سبيل المثال، عديد من مقاييس القدرات، التحصيل والشخصية المقننة لها خصائص سيكومترية جيدة، مثل المجموعات المعيارية الكبيرة والمتقاة على نحو جيد، التطبيق والتصحيح الموضوعى للاختبارات، نظم نقل الدرجة المعيارية المعدلة، والثبات الظاهر وبيانات الصدق. إن المزايا المرتبطة بهذه التنقيحات هى مزايا واضحة منطقيًا ومتجسدة إمبيريقيًا، كما هو مبين فى بقية الكتاب. مع ذلك، بعض مقاييس الشخصية ـ خاصة تلك المسماة بالأساليب الإسقاطية ـ تفتقر إلى عديد من هذه التنقيحات. لأن الأساليب الإسقاطية هى جزء من مجال واسع من القياس النفسى، فإن هذه الأساليب قد تم إمعان النظر فيها بالمثل

للوصول إلى دليل إمبيريقى ذى أهمية. كنتيجة لذلك، أُجبر المناصرون للأساليب الإسقاطية غير المنقحة نسبيًا أن يشحذوا تركيز بحوثهم ويطوروا أدواتهم ليتهيأوا لنقد الأقران، الذى يؤدى إلى معرفة إضافية لملائمة هذه الأساليب.

أخيراً، إن علم النفس كعلم يتصل بعلم النفس كمهنة، عندما تستخدم الاختبارات النفسية. واستجابة لجذوره العلمية،استطاع تكوين معايير لأدوات القياس ولممارسى« المهنة. فعلى سبيل المثال، وضعت معايير القياس التعليمى والنفسى التى نشرتها الجمعية النفسية الأمريكية (١٩٨٥) إرشادات للتقنين، والمعايير، الثبات، الصدق، تفسير الدرجات، وحتى نظم استخدام الكومبيوتر فى التصحيح.

وقد أعلنت معايير تدريب علم النفس المهنى لتكفل معرفة مبادئ علم النفس الأساسية، ونظرية القياس، والتدريب النوعى بأساليب القياس الفردى.

كل هذا يدل على أن إعداد اختبار ـ كعلم ـ يقتضى أكثر من وضع بنود اختبار معًا وتسميتها اختباراً. بالمثل، يشتمل الاستخدام الملائم للاختبار على ما هو أكثر من النقاط دليل الاختبار وقراءة تعليمات التطبيق. مع تزايد أدوات القياس وظهور بيانات الاختبارات فى مجالات الحياة اليومية التى تتسع دائمًا، فإن مستخدم القياس النفسى يجب أن يتسلح بالمعرفة. إن فهم المبادئ الموجودة فى هذا الفصل، والفصول اللاحقة سيساعد مستخدم الاختبارات النفسية على اكتساب هذه المعرفة.

ملخص

إن الاختبارات النفسية، أو القياسات العقلية، تتشارك فى مجموعة من الإجراءات والمبادئ، التى تقدم مزايا على القياس الأقل تنظيمًا. تؤكد الاختبارات النفسية على التطبيق المنتظم والتصحيح الموضوعى. بإزالة عدم الاتساق والنزعة الذاتية من القياس، يستطيع المشخصون أن يزعموا على نحو أكثر ثقة أن الفروق المنظورة بين درجات الأطفال تعكس فروقًا حقيقية فى السمة التى يتم قياسها (مثلاً، مستوى النمو).

معظم الاختبارات النفسية تصحح عديداً. من الأرقام، التى يمكن معالجتها إحصائيًا لتبين بدقة مكان أى طفل، مقارنة بمجموعة الأطفال المعيارية الممثلة الذين طبق عليهم الاختبار. إن الدرجات المعيارية (مثل: الدرجات التائية، درجات نسبة الذكاء)، الرتب المئوية، ومكافئ العمر ومكافئ الفرقة الدراسية هي أكثر الطرق شيوعًا لتقديم تقرير بالدرجات.

هذه الدرجات يشار إليها إجمالاً كدرجات مشتقة؛ لأن الاختبارات النفسية يمكن قياسها كميًا، لذلك يتم بحثها بسهولة. وقد بلور الباحثون طرقًا أمبيريقية لقياس ثبات الاختبار؛ ليضمنوا أن تقاس السمات بأقل الأخطاء. و يمكن أيضًا أن يتم قياس صدق الاختبار إمبيريقيًا، بفحص بنود الاختبار (صدق المحتوى)، بدراسة تغيرات درجات الاختبار فى المواقف المختلفة وملاحظة اتفاقها مع النظرية (الصدق التكوينى)، أو يربط الاختبارات بمعايير خارجية (الصدق التلازمى والتنبؤى).