

مقدمة للتقييم النفسي

لماذا نُجرِّه وماهيته

Introduction to Psychological Assessment

Why We Do It and What It Is

لماذا أحتاج إلى تعلم ما يتعلق بالقياس والتقييم؟

مخطط الفصل

نبذة تاريخية موجزة عن القياس

لغة التقييم

افتراضات التقييم النفسي

لماذا تُستخدم الاختبارات؟

الأهداف التعليمية:

بعد قراءة هذا الفصل ودراسته، ينبغي أن يكون الطلاب قادرين على:

1. وصف المعالم الرئيسية في تاريخ القياس.

2. تعريف الاختبار، والقياس، والتقييم.

3. وصف وإعطاء أمثلة للأنواع المختلفة من الاختبارات.

4. وصف وإعطاء أمثلة للأنواع المختلفة من تفسيرات الدرجات.

5. وصف وتوضيح الافتراضات التي ينطوي عليها التقييم النفسي.

تطبيقات شائعة للتقييمات النفسية

المشاركون في عملية التقييم

التقييم النفسي في القرن الحادي والعشرين

ملخص

6. وصف وتوضيح التطبيقات الرئيسية للتقييم النفسي.

7. توضيح لماذا يستخدم المشتغلون بعلم النفس الاختبارات.

8. وصف المشاركين الرئيسيين في عملية التقييم.

9. وصف بعض التوجُّهات الرئيسية في التقييم.

يلتحق معظم طلاب علم النفس لدراسة هذا العلم لأنهم يودون العمل مع الناس ومعاونتهم، أو لتحقيق فهم أفضل لأفكارهم الشخصية ومشاعرهم وسلوكياتهم. ويأمل كثير من هؤلاء الطلاب أن يصبحوا أخصائيين أو مرشدين نفسيين ويعملون في مواقف كإكلينيكية. ويهتم طلاب آخرون بدرجة أساسية بالبحث ويأملون العمل في الجامعة أو المؤسسات البحثية الأخرى. غير أن عدداً قليلاً يكون لديه الرغبة الشديدة في التخصص في الاختبارات والقياس النفسي. ونتيجة لذلك، فإننا نسمع الطلاب يسألون في أغلب الأحيان، "لماذا ينبغي عليّ أن أدرس مقررًا في الاختبارات والقياس؟" إن هذا يعد سؤالاً منطقيًا. لذلك عندما ندرّس مقررات الاختبارات والقياس، فإننا نستغرق بعض الوقت في توضيح سبب حاجة الطلاب لتعلم ما يتعلق بالقياس والتقييم. وهذا هو أحد الأهداف الرئيسية لهذا الفصل. ونأمل بأننا سوف نقنعك بأن ذلك يعد من الأمور الجديرة بالاهتمام.

يعد القياس والتقييم النفسي مهمًا في جميع جوانب المهن النفسية.

ويعد القياس والتقييم النفسي مهمًا في جميع جوانب المهن النفسية تقريباً. فهؤلاء منكم الذين يهتمون بعلم النفس الكلينيكي، أو علم النفس الإرشادي، أو علم النفس المدرسي، أو البحث، يتبين لهم أن التقييم قد تم تجاوزه فقط من قِبَل العلاج النفسي من حيث الأهمية المهنية (Norcross, Karg, & Prochaska, 1997; Phelp, Eisman, & Kohout, 1989).

ومع ذلك، فإن هذه النتيجة لا تُقدَّر تقديرًا كاملاً أهمية الدور الذي يلعبه التقييم في الممارسة المهنية. وكما ذكر (Meyer et al., 2001)، فإنه على العكس من العلاج النفسي، يعد التقييم الرسمي مظهرًا فريداً من مظاهر ممارسات علم النفس. أي أنه بينما يقدم عدد من الأخصائيين الآخرين في الصحة النفسية علاجاً نفسياً (مثل: المعالجين النفسيين، والمرشدين، والأخصائيين الاجتماعيين)، فإن المشتغلين بعلم النفس هم الوحيدون من بين أخصائيي الصحة النفسية الذي يجرون عادة تقييمات رسمية. ويقدم "موضوع ذو أهمية خاصة 1-1" معلومات تتعلق ببرامج علم النفس للدراسات العليا التي تدرَّب الأخصائيين الكلينيكيين، والمرشدين، وأخصائيي علم النفس المدرسي.

ولا يقتصر استخدام الاختبارات النفسية وغيرها من التقييمات على المواقف الكلينيكية أو الإرشادية. فمثلاً:

- يكرّس أخصائيو علم النفس الصناعي والتنظيمي قدرًا كبيراً من وقتهم المهني في تطوير، وتطبيق وتفسير درجات الاختبارات. وأحد المظاهر الرئيسة لعملهم هو تطوير تقييمات سوف تساعد في تحديد العاملين المحتملين الذين يمتلكون المهارات والخصائص الضرورية للنجاح في العمل.
- يحتاج أيضاً علماء النفس الباحثون إلى أن تكون لديهم الكفاءة في القياس والتقييم. ويتضمن معظم البحث السيكولوجي - بغض النظر عن موضع تركيزه (مثل: التفاعلات الاجتماعية، أو

نمو الأطفال، وعلم النفس المرصّي، أو سلوك الحيوان) - القياس و/أو التقييم. وسواء اهتم الباحث بزمن الاستجابة السلوكية، أو حدة البصر، أو الذكاء، أو المزاج المنقبض (وذلك على سبيل المثال)، فإنه (أو إنها) سوف يحتاج إلى الانغماس في القياس كجزء من البحث. والحقيقة أن جميع مجالات العلم يهتم ويعتمد اعتماداً كبيراً على القياس. فقبل أن يستطيع أحدنا دراسة أي متغير في أي علم من العلوم، من الضروري أن يستطيع التأكد من وجود ذلك المتغير وقياس خصائصه المهمة أو خصائص تكوين فرضى أو كينونة معينة.

● ينغمس المشتغلون بعلم النفس التربوي -الذين يركزون في أغلب الأحيان على تطبيق علم النفس في مواقف تربوية- في العمليات الاختبارية، والقياس، والتقييم. ويتراوح انغماسهم بين تطوير وتحليل الاختبارات التي تُستخدم في مواقف تربوية، وتدريب المعلمين حول كيفية تطوير تقييمات صفيّة أفضل.

موضوع ذو أهمية خاصة 1-1

كيف يختلف الأخصائيون الكلينيكيون، وأخصائيو الإرشاد النفسي، وأخصائيو علم النفس المدرسى عن بعضهم؟

ناقشنا في مقدمة هذا الفصل الدور المهم الذي يلعبه القياس والتقييم في المهن المتعلقة بعلم النفس. وقد اعتمدت الجمعية الأمريكية لعلم النفس برامج تدريبية مهنية في مجالات علم النفس الكلينيكي، وعلم النفس الإرشادي، وعلم النفس المدرسي. وعلى الرغم من وجود أنواع أخرى كثيرة من البرامج السيكلوجية (مثل: علم النفس الاجتماعي، وعلم النفس الكمي، وعلم النفس الفسيولوجي)، فإن معظم الأخصائيين النفسيين الذي يتعاملون مع المرضى تلقوا تدريباً في برامج علم النفس الكلينيكي أو الإرشادي، أو المدرسي. ويشير البعض إلى هذه البرامج بأنها "برامج موجهة للممارسة": نظراً لأن خريجها يتم تدريبهم لتقديم خدمات عناية صحية نفسية. وخريجو مستوى الدكتوراه في برامج علم النفس الكلينيكي أو الإرشادي أو المدرسي، يكونون عادة مؤهلين لمزايا مهنية مكافئة، مثل: الترخيص المهني، والممارسة المستقلة، وأحقية التعويض التأميني للأضرار (مثلاً: Norcross, 2000). غير أنه توجد بعض الفروق الجوهرية. فإذا كنت مهتماً بالالتحاق بمهنة معينة في علم النفس، وأردت أن تعمل في مواقف تطبيقية مع أفراد لديهم مشكلات انفعالية أو سلوكية، فمن المفيد أن تفهم الفرق بين هذه البرامج التدريبية.

وقبل أن نوضح كيفية اختلاف برامج علم النفس الكلينيكي والإرشادي والمدرسي عن بعضهم، من المفيد التمييز بين البرامج التي تمنح درجة الدكتوراه في الفلسفة (Ph.D.)، والبرامج التي تمنح الدكتوراه في علم النفس (PsyD). فبرامج الدكتوراه في الفلسفة تقدم تدريباً موسعاً في كل من التطبيقات الكلينيكية والبحثية (مثل: نموذج العالم - الممارس)، بينما برامج الدكتوراه في علم النفس تركز عادة بدرجة أساسية على التدريب في علم النفس الكلينيكي ويقل تأكيدها على البحث. وتوجد برامج دكتوراه الفلسفة ودكتوراه في علم النفس تتعلق بعلم النفس الكلينيكي والإرشادي والمدرسي، والمزايا النسبية للبرامج الأولى في مقابل الثانية نالت جدلاً شديداً في أغلب الأحيان. ولكننا لن نتعرض لهذا الجدل حالياً، غير أننا نحثك على السعي إلى النصيحة من أساتذة موثوق بهم فيما يتعلق بالدرجة العلمية التي تزيد من احتمال معاونتك في تحقيق أهدافك المهنية.

وهنا من الممكن أن يكون مفيداً التمييز بين برامج علم النفس الكلينيكي، وعلم النفس الإرشادي. ومن الوجهة التاريخية، فإن برامج علم النفس الكلينيكي تدرّب الأخصائيين النفسيين على التعامل مع الأفراد الذين يعانون من أشكال من الأمراض النفسية الأكثر حدة (مثل: الفصام، والاضطراب ثنائي القطب، واختلال العقل) بينما برامج علم النفس الإرشادي تدرّب الأخصائيين النفسيين على التعامل مع الأفراد ذوي مشكلات أقل حدة (مثل: اضطرابات التكيف، والإرشاد المهني والتربوي، وإرشاد الأزواج/الزواج). ولعلك لا تزال تلاحظ هذا التمايز بدرجة ما، ولكن الفروق قد تلاشت بمرور الأعوام. والحقيقة أن الجمعية الأمريكية لعلم النفس (APA) تخلت عن التمييز بين برامج علم النفس الكلينيكي وعلم النفس الإرشادي منذ أعوام ماضية كثيرة. وعلى الرغم من التشابه المتزايد بين هذه البرامج، إلا أن Nocross (2000) ذكر أنه لا تزال هناك فروق ملحوظة بينها، وهذه تشمل:

- عدد برامج علم النفس الكلينيكي كثيرة، ويتخرج منها عدد أكبر من الأخصائيين؛ إذ يوجد 194 برنامج علم نفس كينيكي معترف به من الجمعية الأمريكية لعلم النفس (APA)، 4 برامج علم نفس إرشادي معترف بها من هذه الجمعية.
- من حيث التقييم النفسي، فإن الأخصائيين الكلينيكيين يميلون إلى استخدام تقييمات إسقاطية للشخصية بدرجة أكبر، بينما يميل أخصائيو الإرشاد النفسي إلى استخدام تقييمات مهنية وتعلق بالعمل بدرجة أكبر.
- من حيث التوجّه النظري، فإن معظم الأخصائيين الكلينيكيين وأخصائيي الإرشاد النفسي يفضل المدخل الانتقائي التكاملي، أو المعرفي - السلوكي. غير أن الأخصائيين الكلينيكيين يُحتمل بدرجة أكبر أن يُصادقوا على توجه تحليلي نفسي أو سلوكي. بينما أخصائيو الإرشاد النفسي يميلون إلى تفضيل مداخل تركز على العميل أو المداخل الإنسانية.
- يُحتمل بدرجة أكبر أن يعمل الأخصائيون الكلينيكيون أعمالاً حرة، أو في مستشفيات، أو كليات الطب، بينما يُحتمل بدرجة أكبر أن يعمل أخصائيو الإرشاد النفسي في مراكز الإرشاد الجامعية، وفي مواقف الصحة النفسية المجتمعية.
- يتشابه الطلاب الذين يلتحقون بالبرامج الكلينيكية والإرشادية من حيث درجاتهم في اختبار سجل الدراسات العليا GRE، ومتوسط المعدل العام في المرحلة الجامعية GPA. غير أنه لوحظ أن عدد الحاصلين على درجة الماجستير الذين يلتحقون ببرامج الإرشاد النفسي أكبر ممن يلتحقون بالبرامج الكلينيكية (67%) في برامج الدكتوراه للإرشاد النفسي، في مقابل 21% في برامج الدكتوراه الكلينيكية.
- والخلاصة، أنه على الرغم من تزايد التشابه بين البرامج الكلينيكية وبرامج الإرشاد النفسي في الأعوام الأخيرة، إلا أنه لا يزال يوجد بينها فروق مهمة. والآن ننتقل إلى مجال علم النفس المدرسي. فبرامج علم النفس المدرسي تعد أخصائيين للعمل مع الأطفال والمراهقين في المواقف المدرسية. ويمكن أن تعد البرامج الكلينيكية وبرامج الإرشاد النفسي الخريجين للعمل مع الأطفال والمراهقين، ولكن هذا هو ما تركز عليه برامج علم النفس المدرسي. ومثل الأخصائيون الكلينيكيون، وأخصائيو الإرشاد النفسي، يتم تدريب أخصائيي علم النفس المدرسي على أساليب العلاج النفسي والإرشاد. كما يتلقون تدريباً مكثفاً على التقييم النفسي (المعرفي، والانفعالي، والسلوكي) ويتعلمون استشارة الآباء وغيرهم من الاختصاصيين لحفز نجاح عملائهم في المدرسة.

ويتوافر حالياً 57 من البرامج التدريبية في علم النفس المدرسي معتمدة من الجمعية الأمريكية لعلم النفس (APA).

وقد ركزت هذه المناقشة على البرامج التدريبية على مستوى الدكتوراه فقط. ويوجد أيضاً عدد أكبر من البرامج السيكلوجية على مستوى الماجستير، وهي تعد أيضاً الطلاب لمهن مثل أخصائي الصحة النفسية. ونظراً لأن منح التراخيص المهنية تتحكم فيها كل ولاية على حدة، فإن كلاً منها يحدد المحكات التربوية والتدريبية اللازمة للترخيص، وهذا يعد شيئاً آخر ينبغي مراعاته عند اختيار كلية دراسات عليا.

ويمكنك معرفة مزيد من التفاصيل المتعلقة بهذه المجالات من التخصصات السيكلوجية بزيارة موقع الإنترنت الخاص بالجمعية الأمريكية لعلم النفس، وهو: (<http://www.apa.org>) والتعريفات الرسمية لكل من هذه التخصصات مبينة على هذا الموقع، وكذلك وثائق أرشيفية أخرى تصف كل تخصص بالتفصيل الذي وافقت عليه لجنة APA للاعتراف بالتخصصات والكفاءات المتعلقة بمهن علم النفس (CRSPPP).

والخلاصة، إذا كنت مهتماً بالالتحاق بمهنة في علم النفس، فإنه من المحتمل أنك سوف تتخبط في التقييم بدرجة ما، أو تحتاج على الأقل أن تفهم نواتج أنشطة القياس والتقييم. وقبل التعمق في القضايا المعاصرة في القياس والتقييم، سوف نوضح بإيجاز تاريخ القياس النفسي.

نبذة تاريخية موجزة عن القياس BRIEF HISTORY OF TESTING

ذكرت أنستازى وإربينا (1997) أن "الأصول الأولى الفعلية للقياس قد فقدت في العصور السابقة للعصور الوسطى" (ص 32). ويقترح بعض الكتاب أن أول اختبار كان بالفعل "Apple test" الشهير، الذي أجرى على حواء في جنة عدن. غير أننا إذا استثنينا المصادر التوراتية، فإن القياس يعود عادة إلى الصينيين القدامى. ويلقى هذا القسم الضوء على بعض المعالم في تاريخ القياس.

القياس المبكر: حوالي عام 2200 قبل الميلاد Earliest Testing: Circa 2200 BC

يرجع الاستخدام المبكر الموثق للاختبارات عادة إلى الصينيين الذين اختبروا المسؤولين العموميين للتأكد من كفاءتهم (وهذا يماثل الامتحانات المعاصرة للخدمة المدنية). وقد نشأ برنامج الصين الاختباري عبر القرون وتضمن صيغاً متنوعة. فمثلاً، في أثناء سلالة Han الحاكمة، تم تضمين الامتحانات التحريرية لأول مرة، واشتملت على خمسة مجالات رئيسية، هي: الزراعة، والقانون المدني، والجغرافيا، والشؤون العسكرية، والدخل الحكومي.

وفي غضون القرن الرابع، تضمن البرنامج ثلاث مراحل شاقّة تتطلب من المفحوصين قضاء عدة أيام منعزلين في حجرات صغيرة بيتكرون مقالات وشعر (Gregory, 2004).

يرجع الاستخدام المبكر الموثق للاختبارات عادة إلى الصينيين الذين اختبروا المسؤولين العموميين للتأكد من كفاءتهم.

القياس في القرنين الثامن عشر والتاسع عشر Eighteenth - and Nineteenth - Century Testing

كارل فريدريك جاوس: يعد جاوس (1777-1855) أحد علماء الرياضيات ذائعي الصيت، وقدّم أيضاً إسهامات مهمة في علم الفلك ودراسة المغناطيسية. وفي متابعته لحركات النجوم وجد أن زملاءه توصلوا في أغلب الأحيان إلى مواقع لا تختلف عما توصل إليه إلا قليلاً. ووضح بيانياً تكرار المواقع الملاحظة بانتظام، ووجد أن المواقع تتخذ شكلاً منحنياً - وهو المنحنى الذي أصبحنا نعرفه باسم المنحنى الاعتدالي أو التوزيع الاعتدالي (ويُعرف أيضاً باسم منحنى جاوس). وحدد أن التقدير الأفضل للموقع الدقيق لنجم معين هو متوسط الملاحظات، وكل ملاحظة مستقلة تشتمل على قدر معين من الخطأ. وعلى الرغم من أن جاوس يُعرف عادة بأنه رائد القياس، إلا أننا نعتقد أن اعترافه الرسمي بأخطاء القياس وخصائص توزيعها هو الذي جعله شهيراً.

يعد جالتون مؤسس الاختبارات العقلية والقياس العقلي، وكان مسؤولاً عن أول جمع منظم للبيانات المتعلقة بالفروق الفردية على نطاق متسع.

امتحانات الخدمة المدنية: قُدِّمت اختبارات الخدمة المدنية المماثلة لتلك التي استُخدمت في الصين لانتقاء موظفي الحكومة في الدول الأوروبية. وبدأت لجنة الخدمة المدنية في الولايات المتحدة عام 1883 استخدام اختبارات تحصيلية مماثلة للمعاونة في انتقاء موظفي الحكومة (Aan-stasi & Urbina, 1997).

الأطباء والمعالجون النفسيون: طُوِّر الأطباء والمعالجون النفسيون في إنجلترا والولايات المتحدة في القرن التاسع عشر أنظمة تصنيفية للمعاونة في تصنيف الأفراد ذوي التخلف العقلي وغيرها من المشكلات العقلية. فمثلاً، خلال الأعوام التالية لعام 1830 كان الطبيب الفرنسي جين إسكويرول من أوائل الأطباء الذي ميَّز الحَبَل insanity (أي، الاضطرابات الانفعالية) عن العجز العقلي (أي، العجز العقلي الموجود منذ الميلاد). واعتقد أيضاً أن التخلف العقلي يوجد على متصل يتراوح بين المتوسط والشديد، ولاحظ أن المهارات اللفظية كانت أكثر الطرق اتساقاً لتحديد درجة التخلف العقلي. وفي خلال الأعوام التالية لعام 1890 حث إيميل كربلين وآخرون على استخدام اختبارات التداعي الحر في تشخيص المرضى النفسيين. وتتضمن اختبارات التداعي الحر عرض كلمات مثيرة يستجيب لها الفرد "بأول كلمة تخطر على عقله". وقد وسَّع فرويد فيما بعد هذا الأسلوب، وشجَّع المرضى على التعبير بحرية عن جميع الأفكار التي تخطر على عقولهم من أجل تحديد ما تتطوي عليه من أفكار ومشاعر.

الأدوات النحاسية ERA: أسهم علماء النفس التجريبيون المبكرون، مثل: ويلهالم فندت، وسير فرانسيس جالتون، وجيمس ماكين كاتل، وكلاارك وايلزر، إسهامات جوهرية في تطوير قياس القدرات المعرفية. وأحد أهم التطورات في هذه الفترة هو الانتقال إلى قياس القدرات الإنسانية باستخدام

إجراءات أكثر موضوعية ويمكن تكرارها بسهولة، وكانت عادة من النحاس، لقياس عمليات حسية وحركية بسيطة اعتماداً على فرضية أنها كانت مقاييس للذكاء العام مثلاً (Gregory, 2004).

وبعض الإسهامات المبكرة لهؤلاء العلماء كان مهماً بدرجة كبيرة مما يتطلب مزيداً من التوضيح.

سير فرانسيس جالتون: يعد جالتون في أغلب الأحيان مؤسس الاختبارات والقياس العقلي. وأحد إنجازاته الرئيسية هو إنشاء مختبر دراسة مقاييس الجسم الإنساني في المعرض الدولي للصحة في لندن عام 1884. وقد تحول هذا المعرض بعد ذلك إلى متحف ظل قائماً لأعوام عدة. وخلال هذا الزمن تم جمع بيانات تشتمل على قياسات جسمية (مثل: الطول، والوزن، ومحيط الرأس)، وحسّية (مثل: زمن الرجح، والتمييز الحسّي)، والحركية (مثل: سرعة الحركة، وقوة القبضة) من أفراد يزيد عددهم عن 17,000. وقد مثل ذلك أول جمع لبيانات منظمة للفروق الفردية على نطاق متسع (Anastasi & Urbina, 1997; Gregory, 2004).

جيمس ماكين كاتل: اتفق كاتل مع جالتون في اعتقاده بأنه يمكن استخدام الاختبارات الحسّية والحركية البسيطة في قياس القدرات العقلية. وكان كاتل عملياً في افتتاح مختبرات نفسية ونشر حركة القياس المتزايدة في الولايات المتحدة. ونظراً إليه على أنه أول من استخدم مصطلح "اختبار عقلي" في مقالة نشرها عام 1890. وبالإضافة إلى إسهاماته المهنية الشخصية، تخرّج على يديه العديد من الطلاب الذين التحقوا بمهن مثمرة في علم النفس من أمثال: هـ. ل. ثورنديك، ر. س. وودورث، هـ. ك. سترونج (Anastasi & Urbina, 1997; Gregory, 2004).

وقد أسهم أيضاً كل من جالتون وكاتل في تطوير إجراءات اختبارية، مثل: استبانات مقننة، وموازن تقدير، والتي أصبحت فيما بعد أساليب شائعة في تقييم الشخصية (Anastasi & Urbina, 1997).

كلارك وايزلر: يعد وايزلر أحد تلاميذ كاتل، وقد تخطت بحوثه بدرجة كبيرة البحوث التي أجراها أستاذه الشهير. فقد وجد وايزلر أن المقاييس الحسّية-الحركية المستخدمة بكثرة في تقييم الذكاء لا ترتبط أساساً بالتحصيل الدراسي. ووجد أيضاً أن الاختبارات الحسّية-الحركية ترتبط ارتباطاً ضعيفاً ببعضها بعضاً. وهذه النتائج المحبطة أدت إلى نهاية استخدام مقاييس الذكاء الحسّية-الحركية البسيطة، ومهدت الطريق لمدخل جديد للتقييم العقلي الذي يؤكد عمليات عقلية عليا أكثر تعقيداً. ومما يثير الدهشة وجود جوانب قصور منهجية جوهرية في بحوث وايزلر أعاقته عن الكشف عن معاملات ارتباط متوسطة توجد بالفعل بين بعض مقاييس الذكاء الحسّية-الحركية (ولمزيد من التفاصيل عن التحول المثير للأحداث، يمكنك الرجوع إلى (Francher, 1985; Sternberg, 1990). ومع ذلك، ربما يستغرق الباحثون عقوداً كثيرة لاكتشاف أنهم ربما لم يأخذوا بعين الاعتبار أهمية القياسات العقلية في بحوثهم المتعلقة بالذكاء، وقد مهدّ الطريق لمدخل ألفرد بينيه لقياس الذكاء الذي يؤكد القدرات العقلية العليا (Gregory, 2004).

القياس في القرن العشرين Twentieth - Century Testing

ألفرد-بينيه-مدخل القياس العقلي: قام بينيه مبدئياً بالبحث في القياسات الحسية-الحركية، مثل: زمن الرجح، وحدة الحس، ولكنه أصبح غير راض عنها، وكان رائداً في استخدام مقاييس العمليات المعرفية العليا لتقييم الذكاء (Gregory, 2004). وقد طلبت الحكومة الفرنسية في مستهل القرن العشرين من بينيه وزميله ثيودور سيمون تطوير اختبار للتنبؤ بالأداء الأكاديمي. وكانت نتائج جهودهما أول مقياس للذكاء حمل اسميهما، وتم نشره عام 1905.

نُشر الاختبار الأول لبينيه وسيمون عام 1905، ويعد أول اختبار ذكاء، وكان منبئاً جيداً بالنجاح الأكاديمي.

واشتمل الاختبار على بعض الاختبارات الحسية-الإدراكية، ولكنه ركز على المفردات اللفظية التي تقيس الفهم، والاستدلال، والحكم، والذاكرة قصيرة الأمد. وحقق بينيه وسيمون أهدافهما بتطوير اختبار كان منبئاً جيداً بالنجاح الأكاديمي. وتم نشر

التقييمات التالية لمقياس بينيه وسيمون عام 1908، 1911. وقد حظت هذه المقاييس بقبول واسع في فرنسا، وتمت ترجمتها وتقنينها على الفور في الولايات المتحدة بنجاح كبير بواسطة لويس تيرمان بجامعة ستانفورد. ونتج عن ذلك مقياس ستانفورد-بينيه، وتم تنقيحه مرات متعددة (الطبعة الخامسة المنقحة SB5 لا تزال تُستخدم في وقتنا الحاضر). ولعل النسخة المعدلة التي أعدها تيرمان لمقياس ستانفورد - بينيه أكثر شيوعاً في فرنسا وأماكن أخرى في أوروبا من المقياس الأصلي.

اختبارات الجيش ألفا وبيتا: نال قياس الذكاء زخماً آخر في الولايات المتحدة في أثناء الحرب العالمية الأولى. فقد احتاج الجيش الأمريكي إلى طريقة لتقييم وتصنيف المجندين للقبول في الخدمة العسكرية، وتصنيفهم للأعمال العسكرية. وقد شكّل روبرت م. بيركس - وهو رئيس سابق للجمعية الأمريكية لعلم النفس - مجموعة عمل قامت ببناء سلسلات من اختبارات الاستعدادات أصبحت تُعرف باسم اختبارات الجيش ألفا، وبيتا - أحدهما لفظي (ألفا) والآخر غير لفظي (بيتا). ومن خلال جهودهم وجهود الجيش في فحص المجندين أصبح ملايين الأمريكيين لديهم ألفة بمفهوم قياس الذكاء.

روبرت وودورث - مدخل قياس الشخصية: طوّر روبرت وودورث عام 1918 قائمة البيانات الشخصية لودورث، حيث اعتُبرت بدرجة كبيرة أول اختبار رسمي للشخصية. وقد صُممت هذه القائمة للمعاونة في جمع معلومات شخصية عن المجندين في الجيش. ومثل كثير من فوائد تطوير مقاييس بينيه في مجال قياس الذكاء، فإن تقديم قائمة البيانات الشخصية لودورث كان لها فوائدها في مجال تقييم الشخصية.

اختبار بقع الحبر لرورشاخ: طُوِّر رورشاخ بقع الحبر بدءاً من عام 1920. وكان هناك كثير من الجدل فيما يتعلق بالخصائص السيكمترية لاختبار الرورشاخ (وغيره من الأساليب الإسقاطية)، ولكنه ظل واحداً من أساليب تقييم الشخصية الأكثر شيوعاً واستخداماً في بداية القرن الحادي والعشرين.

اختبارات القبول بالكليات: تم بناء اختبار الالتحاق بالكليات (CEEB) في الأصل لتزويد الكليات والجامعات بمقياس موضوعي وصادق للقدرات الأكاديمية للطلاب، من أجل التخلص من المحاباة والتورث، وتأكيد الجدارة الأكاديمية. وأدت جهود المجلس الذي أعد الاختبار إلى تطوير أول اختبار استعداد دراسي (SAT) عام 1926 (يُطلق عليه الآن اختبار التقييم الدراسي). وبدأ بناء البرنامج الاختباري الأمريكي للكليات (ACT) عام 1959، وهو منافس رئيس للاختبار SAT. وقبل ابتكار هذه الاختبارات، كانت قرارات القبول بالكليات ذاتية بدرجة كبيرة، وتتأثر بشدة بخلفية الأسرة ومكانتها. ولذلك فإن غرضاً آخر لتطوير هذه الأدوات هو جعل عملية الانتقاء أكثر موضوعية.

مقاييس ويكسلر للذكاء: نال قياس الذكاء دَفْعَةً أُخْرَى بدءاً من عام 1930 عندما طُوِّر دافيد ويكسلر اختبار ذكاء اشتمل على مقاييس القدرة اللفظية وغير اللفظية في الاختبار نفسه. وقبل ويكسلر، ومقياس Wechsler-Bellevue I، كانت اختبارات الذكاء تقيس عادة الذكاء اللفظي أو غير اللفظي، وليس كلاهما. وأصبحت مقاييس ويكسلر أكثر اختبارات الذكاء شيوعاً واستخداماً في وقتنا الحاضر.

استبيان مينيسوتا للشخصية متعدد الأوجه (MMPI): نُشر هذا الاستبيان في بدايات أعوام 1940 للمعاونة في تشخيص الاضطرابات الطبية النفسية، وهو اختبار شخصية موضوعي (أي يمكن تقدير درجاته بطريقة موضوعية)، وخضع لقدر كبير من البحث. واستمر الاستبيان MMPI-2 من أكثر مقاييس الشخصية استخداماً بل وشيوعاً في وقتنا الحاضر.

القياس في القرن الحادي والعشرين Twenty - First - Century Testing

شهدت الستون عاماً الأخيرة توسعاً كبيراً من حيث تطوير الاختبارات، واستخدام الاختبارات النفسية والتربوية. فمثلاً، المسح الأخير "للكتاب السنوي للقياسات العقلية" أدّى إلى تحديد ما يزيد على 600 اختبار مدوناً في القسم المتعلق باختبارات الشخصية، وما يزيد عن 400 اختبار في الأقسام المتعلقة باختبارات الذكاء والاستعدادات. وقد خصّص كثير من فصول هذا المرجع لمناقشة تطوير واستخدام أنواع مختلفة من الاختبارات. وسوف نوضح فيما بعد في هذا الفصل بعض الاتجاهات الحديثة في التقييم، وبعض العوامل التي نتوقع أنها سوف تؤثر في مسار ممارسات التقييم في القرن الحادي والعشرين.

وقد لخصنا هذا التطور التاريخي في جدول 1-1.

جدول 1-1 معالم التطور التاريخي للقياس

حوالي عام 2200 قبل الميلاد	● اختبر الصينيون المسؤولين العاميين.
نهايات 1800، وبدايات 1900	● اكتشف كارل فريدريك جاوس التوزيع الاعتمالي عند تقييم أخطاء القياس. ● استخدمت امتحانات الخدمة المدنية في أوروبا.
القرن التاسع عشر	● قيّم الأطباء والمعالجون النفسيون المرضى النفسيين بأساليب جديدة. ● عصر الأدوات النحاسية - تأكيد قياس القدرات الحسية والحركية. ● بدأ استخدام امتحانات الخدمة المدنية في الولايات المتحدة. ● انتباه مبكر للاستبيانات وموازن التقييم بواسطة جالتون وكاتل.
عام 1905	● نشر مقياس بينيه وسيمون - مرشد في عصر القياس النفسي.
عام 1917	● نشر اختبار الجيش ألفا، وبيتا- أول اختبارات جماعية لاختبارات الذكاء.
عام 1918	● نُشرت قائمة البيانات الشخصية لودورث - بداية عصر قياس الشخصية.
خلال أعوام 1920	● تطوير اختبار الاستعداد الدراسي (SAT)، واختبار بقع الحبر لرورشاخ - توسّع تأثير القياس.
خلال أعوام 1930	● نشر دافيد ويكسلر مقياس Wechsler-Bellevue - استهلال سلسلة من اختبارات الذكاء ذات التأثير.
خلال أعوام 1940	● نشر استبيان مينيسوتا للشخصية متعدد الأوجه (MMPI) - أصبح نموذجاً لاستبيانات الشخصية الموضوعية.

لغة التقييم THE LANGUAGE OF ASSESSMENT

لقد استخدمنا عدداً من المصطلحات الشائعة ولكنها فنية إلى حد ما. وقبل الاسترسال، ربما يكون من المفيد تعريفها لك.

الاختبارات، والقياس، والتقييم Tests, Measurement, and Assessment

الاختبارات Tests: الاختبار، هو أداة أو إجراء يتم بواسطته الحصول على عينة من سلوك الفرد وتقييمها وتقدير درجاتها باستخدام إجراءات مقننة (الجمعية الأمريكية للبحث التربوي (AERA)، والجمعية الأمريكية لعلم النفس (APA)، والمجلس القومي الأمريكي للقياس التربوي (NCME)، 1999).

الاختبار هو إجراء يتم بواسطته الحصول على عينة من سلوك الفرد وتقييمها وتقدير درجاتها باستخدام إجراءات مقننة (AERA et al., 1999).

وهذا يعد تعريفاً متسعاً أو عاماً، ولكن ربما يخدمنا هذا التعريف المتسع بدرجة أفضل ونحن في مستهل مناقشتنا. وينبغي أن نطمئن لأننا سوف نقدم معلومات أكثر تحديداً عن الأنواع المختلفة من الاختبارات في الوقت المناسب. وقبل الاسترسال ينبغي أن نوضح بتفصيل أكثر أحد مظاهر تعريفنا للاختبار، أي: الاختبار هو عينة من السلوك. ونظراً لأن الاختبار يقتصر على عينة من السلوك، فمن المهم أن تعكس الاختبارات عينة ممثلة للسلوك الذي تهتم به، وسوف تصبح أهمية مفهوم العينة الممثلة في أثناء متابعتنا لدراسة القياس والتقييم أكثر وضوحاً، وسوف نلقي الضوء عليها بمزيد من التفصيل عندما نتناول الخصائص الفنية للاختبارات.

يُعرّف القياس بأنه مجموعة من القواعد لتعيين أعداد تمثل الأشياء، أو السمات، أو الخصائص، أو السلوك.

الاختبارات المقننة Standardized Tests: الاختبار المقنن هو اختبار يطبق، ويصحح، وتفسر درجاته بطريقة معيارية. ويُعد معظم الاختبارات المقننة بوساطة أخصائيين، أو شركات نشر الاختبارات. والغرض من التقنين ضمان أن تكون ظروف العملية الاختبارية متكافئة تقريباً قدر المستطاع بالنسبة لجميع الأفراد الذين يطبق عليهم الاختبار. وإذا تحقق ذلك، فإنه لا يكون لأي فرد ميزة عن غيره نتيجة لتباين إجراءات التطبيق، وتكون نتائج التقييم قابلة للمقارنة.

القياس Measurement: يمكن تعريف القياس بأنه مجموعة من القواعد لتعيين أعداد تمثل الأشياء، أو السمات، أو الخصائص، أو السلوك. والاختبار النفسي هو أداة قياس، ولهذا فإنه يتضمن قواعد (مثلاً، مرشد تطبيق، ومحكات تقدير الدرجات) لتعيين أعداد تمثل أداء الفرد. وبالتالي، تُفسر هذه الأعداد على أنها تعكس خصائص من يطبق عليه الاختبار، فمثلاً، ربما يُفسر عدد المفردات التي يجيب عليها الفرد إيجابياً (مثل، "صحيحة"، أو "تنطبق عليّ") في مقياس للاكتئاب بأنه يعكس خبرة الفرد فيما يتعلق بالمزاج المنقبض.

التقييم Assessment: يُعرّف التقييم بأنه إجراء منظم لجمع معلومات يمكن استخدامه للاستدلال على خصائص الناس أو الأشياء (AERA et al., 1999). وينبغي أن يؤدي إلى مزيد من الفهم لهذه الخصائص. والاختبارات بالطبع تُعد إحدى الطرق المنظمة لجمع معلومات، ولذلك فهي مجموعة من الأدوات للتقييم. وتعد أيضاً مراجعة السجلات التاريخية، والمقابلات، والملاحظة أساليب تقييم مشروعة، وجميعها مفيدة للغاية إذا تكاملت. والحقيقة، إن التقييم يشير عادة إلى عملية تتضمن تكامل المعلومات التي يتم الحصول عليها من مصادر متعددة باستخدام طرق متعددة. ولذلك، فإن التقييم عملية أكثر اتساعاً وأكثر شمولاً من القياس.

يُعرّف التقييم بأنه أي إجراء منظم لجمع معلومات يمكن استخدامه للاستدلال على خصائص الناس أو الأشياء (AERA et al., 1999).

وقد لاحظ (Meyer et al. 2001) في مقارنته بين القياس النفسي والتقييم النفسي أن القياس يعد عملية مباشرة نسبياً، حيث يُطبق اختبار معين للحصول على درجة معينة. وفي المقابل، فإن التقييم النفسي يُحدث تكاملاً بين درجات متعددة يتم الحصول عليها عادة باستخدام اختبارات متعددة، ومعلومات يتم جمعها بمراجعة السجلات، وإجراء مقابلات، وملاحظة. والغرض هو الفهم الأفضل للعميل، وإجابة أسئلة ذات مغزى (مثل: لماذا يؤدي هذا الطالب أداءً ضعيفاً في المدرسة؟)، والإعلام بهذه المعلومات للأفراد المناسبين.

وقد زاد (McFall and Trent 1999) على ذلك بتذكيرنا أن "الغرض من التقييم الكلينيكي هو جمع بيانات تسمح لنا بتقليل عدم التأكد فيما يتعلق باحتمالات الأحداث" (p.215). ومثل هذا "الحدث" في موقف كلينيكي معين ربما يتعلق باحتمال أن يكون المريض لديه اكتئاب في مقابل اضطراب ثنائي القطب، أو في مقابل الخبل، وجميعها تتطلب علاجات. وفي علم نفس التوظيف ربما يكون احتمال المتقدم للالتحاق بالشرطة سوف ينجح في إكمال دراسته في أكاديمية الشرطة، وهذا يتطلب تدريباً مرتفع الكلفة. وتفيد درجات الاختبارات في هذه العملية بالقدر الذي "تسمح بالتنبؤ بالأحداث أو التحكم فيها بدقة أكبر أو بخطأ أقل مما لو لم تستخدم الاختبارات" (McFall & Trent, 1999, p. 217). وبالإشارة إلى الحدثين السابقين - التشخيص والانتقاء للتوظيف - فإن الاختبارات تكون مفيدة في عملية التقييم بالقدر الذي تجعلنا أكثر دقة في اتخاذ قرارات تشخيصية أو توظيفية على الترتيب. والمهنيون يخطئون في جميع الحالات، وغرضنا تقليل هذه الأخطاء، من حيث العدد والمقدار. والقياس والتقييم في علم النفس يفيد فائدة كبيرة لتحقيق هذا الغرض إذا أُجرى بطرق مناسبة.

والآن وبعد تعريفنا للمصطلحات الشائعة، فإننا نعتزف -وبقليل من المعارضة- بأن كثيراً من المهنيين يستخدمون في ممارساتهم الفعلية مصطلحات "العمليات الاختبارية، والقياس، والتقييم" بمعنى واحد. واعترافاً من (Popham 2000) بذلك، فإنه ذكر أن مصطلح "التقييم" أصبح المصطلح المفضل لدى كثير من الاختصاصيين. فمصطلح "القياس" يبدو أكثر صرامة وجدياً عند تطبيقه على الناس، وهناك ميل لتجنبه. "والعمليات الاختبارية" لها تضميناتها السلبية الخاصة بها. فمثلاً، يصعب أن يمر أسبوع دون أن تشتمل الصحف على مقالات تتعلق "بالتدريس من أجل الاختبار" أو "الاختبارات ذات المخاطر العالية"، وتكون لهذه المقالات تضمينات سلبية. وبالإضافة إلى ذلك، فإن الناس عندما تسمع كلمة "اختبار"، فإنهم يفكرون عادة في اختبارات الورقة والقلم. ونتيجة للزيادة المطردة لعدم الرضا عن اختبارات الورقة والقلم التقليدية؛ تم في الأعوام الأخيرة تطوير إجراءات عمليات اختبارية بديلة alternative testing (مثل، تقييمات الأداء، والبورتفوليو). ونتيجة لذلك، فإن "العمليات الاختبارية testing لا يُنظر إليها على أنها تصف بخاصة الممارسات الحديثة. وهذا يجعل مصطلح "التقييم assessment" هو المصطلح المعاصر الشائع.

وهناك مصطلحات إضافية ينبغي أن تكون على ألفة بها. فالتقويم Evaluation هو مصطلح يُستخدم في أغلب الأحيان عند مناقشة التقييم، والعمليات الاختبارية، والقضايا المتعلقة بالقياس. والتقويم يعد نشاطاً يتضمن الحكم أو تقدير قيمة أو جودة شيء ما. فمثلاً، تعيين تقديرات رسمية

للطلاب تعكس أداءهم الأكاديمي يشار إليها بأنها تقويم ختامي Summative Evaluation. وعلم السيكومتري هو علم القياسي النفسي، "وعالم السيكومتري" هو أخصائي نفسي أو تربوي متخصص في مجال العمليات الاختبارية Testing، والقياس Measurement، والتقييم Assessment. وسوف يُحتمل أن تسمع الناس تشير إلى الخصائص السيكومترية لاختبار معين، وهم بذلك يقصدون قياس الخصائص الإحصائية للاختبار. وتشمل هذه الخصائص: الثبات، والصدق. ويشير "الثبات Re-Liability" إلى استقرار، واتساق، درجات الاختبار، ودقتها النسبية.

ويشير الثبات - على المستوى الأكثر تنظيراً - إلى درجة خلو درجات الاختبار من أخطاء القياس. فالدرجات الخالية نسبياً من أخطاء القياس سوف تكون مستقرة أو متسقة (أي، ثابتة)، وبذلك تكون أكثر دقة في تقدير قيمة معينة. و"الصدق Validity"، يشير - في عبارات بسيطة - إلى ملاءمة أو دقة تفسير درجات الاختبار. فإذا تم تفسير درجات الاختبار بأنها تعكس الذكاء؛ فهي تعكس بالفعل القدرة العقلية؟ وإذا استخدمت درجات الاختبار في التنبؤ بالنجاح في عمل معين، هل يمكنها بالفعل التنبؤ بمن سوف يكون ناجحاً في ذلك العمل؟

الثبات، يشير إلى استقرار أو اتساق درجات الاختبار.

الصدق، يشير إلى دقة تفسير درجات الاختبار.

أنواع الاختبارات Types of Tests

عرّفنا "الاختبار" بأنه أداة أو إجراء يتم عن طريقه الحصول على عينة من سلوك الفرد، وتقييمها، وتقدير درجاتها باستخدام إجراءات مقننة (AERA et al., 1999). ولعلك اختبرت بعدد كبير من الاختبارات طوال حياتك، ومن المحتمل أن تكون قد لاحظت أن الاختبارات ليست متشابهة. فمثلاً، يُختبر الطلاب في المدرسة للمعاونة في تحديد تقديراتهم، ويُختبرون للحصول على رخصة قيادة السيارة، ويُختبرون باستبيانات ميول للمعاونة في اتخاذ قرارات تربوية ومهنية، ويُختبرون باختبارات قبول قبل التحاقهم بكلية معينة. وهذه القائمة الموجزة ليست شاملة بالطبع!

وقد ذكر Cronbach (1990) أنه يمكن تصنيف الاختبارات بعامّة في مقاييس الأداء الأقصى أو مقاييس الأداء المميّز. ويشار أيضاً إلى مقاييس الأداء الأقصى بأنها اختبارات القدرة، ولكن اختبارات التحصيل تكون أيضاً متضمنة هنا. وربما تقدّر درجات مفردات اختبارات الأداء الأقصى على أساس أن إجابة المفردة إما "صحيحة"، أو "غير صحيحة"، ويُحَثُّ المختبرون على بذل أقصى قدرتهم في أدائهم.

وتصمم "اختبارات الأداء الأقصى" "Maximum performance tests" لتقييم الحدود العليا لمعارف الفرد وقدراته. فمثلاً، يمكن أن تصمم اختبارات الأداء الأقصى لتقييم مدى جودة أداء الطالب في مهام منتقاة (مثل، ضرب الأعداد التي تشتمل على ثلاثة أرقام)، أو إتقانه نطاق محتوى معين (مثل، التاريخ الأمريكي). وتعد اختبارات الذكاء والاختبارات التحصيلية الصفيّة أمثلة شائعة لاختبارات الأداء الأقصى. وعلى العكس من ذلك، فإن اختبارات الأداء المميز Typical Response tests تحاول

تُصمَّم اختبارات الأداء الأقصى لتقييم الحدود العليا لمعارف الطالب وقدراته.

تقيس الاختبارات التحصيلية المعارف والمهارات في مجال معين قدّم فيه تعليم.

قياس السلوك المعتاد للأفراد وخصائصهم. ويشار في أغلب الأحيان لاختبارات الأداء المميّز على أنها اختبارات الشخصية، وتُستخدم "الشخصية Personality" في هذا السياق استخداماً متسعاً لتعكس مجموعة كبيرة من الخصائص غير المعرفية، مثل: الاتجاهات، والسلوكيات، والانفعالات، والميول (Anastasi & Urbina, 1997). ويُقصر بعض الأفراد مصطلح "الاختبار test" على مقاييس الأداء الأقصى، بينما يُقصر استخدام مصطلحات، مثل "مقياس Scale، أو استبيان Inventory" على أدوات الاستجابة المميّزة. ونحن نستخدم في هذا المرجع مصطلح "الاختبار" بمفهومه المتسع، وتطبيقه في كل من الأداء الأقصى، وإجراءات الاستجابة المميّزة.

اختبارات الأداء الأقصى Maximum Performance Tests: تُصمم اختبارات الأداء الأقصى لتقييم الحدود العليا لمعارف الطالب وقدراته. ويوجد في إطار القسم المتسع لاختبارات الأداء الأقصى عدد من الأقسام الفرعية. أولاً: يمكن تصنيف اختبارات الأداء الأقصى على أنها إما اختبارات تحصيلية أو اختبارات استعدادات. ثانياً: يمكن تصنيف اختبارات الأداء الأقصى على أنها إما موضوعية أو ذاتية. وأخيراً: توصف اختبارات الأداء الأقصى في أغلب الأحيان بأنها إما اختبارات سرعة أو اختبارات قوة. وهذه التمايزات -على الرغم من أنها ليست مطلقة في طبيعتها- لها أساس تاريخي طويل، وتقدّم بعض المعلومات الوصفية المفيدة.

تقيس اختبارات الاستعدادات القدرات والمهارات المعرفية المتجمّعة نتيجة لخبرات الحياة بعامّة (AERA et al., 1999).

الاختبارات التحصيلية واختبارات الاستعدادات: تُصنّف اختبارات الأداء الأقصى في أغلب الأحيان على أنها إما اختبارات تحصيلية أو اختبارات استعدادات. وتُصمَّم "الاختبارات

التحصيلية achievement tests" لتقييم معارف ومهارات فرد معين في نطاق محتوى قدّم له فيه تعليم معين. وعلى العكس من ذلك، فإن "اختبارات الاستعدادات aptitude tests" تكون أكثر اتساعاً في منظورها، وتُصمم لقياس المهارات والقدرات المعرفية، والمعارف التي تجمّعت لدى الفرد نتيجة لخبراته مدى حياته (AERA et al., 1999). وبعبارة أخرى، فإن الاختبارات التحصيلية متعلقة أو مرتبطة ببرنامج أهداف تعليمية معينة، بينما اختبارات الاستعدادات تعكس التأثير التراكمي لخبرات الحياة ككل. غير أن هذا التمايز ليس مطلقاً، وهو أمر في الحقيقة يتعلق بالدرجة أو موضع التأكيد. ويتصوّر معظم خبراء القياس في وقتنا الحاضر كلاً من الاختبارات التحصيلية واختبارات الاستعدادات على أنها مقاييس للقدرات المعرفية النامية، ويمكن ترتيبها على متّصل من حيث درجة الصلة بين القدرات موضع التقييم، وخبرات تعلم معينة.

وتمايز آخر بين الاختبارات التحصيلية واختبارات الاستعدادات يتضمن الطريقة التي تُستخدم أو تُفسَّر بها نتائجها. فالاختبارات التحصيلية تُستخدم عادة في قياس ما تم تعلمه أو "تحصيله" في وقت معين. وعلى العكس من ذلك، فإن اختبارات الاستعدادات تُستخدم عادة في التنبؤ بالأداء المستقبلي أو تعكس إمكانية الفرد المتعلقة بالأداء الأكاديمي أو الأداء في العمل. وعلى الرغم من ذلك، فإن هذا التمايز ليس مطلقاً أيضاً. فعلى سبيل المثال، الاختبار الذي يُطبق في نهاية المرحلة الثانوية لتقييم التحصيل ربما يُستخدم أيضاً في التنبؤ بالنجاح في الكلية. وعلى الرغم من أننا نشعر أنه من المهم الاعتراف بأن التمايز بين الاختبارات التحصيلية والأكاديمية ليس مطلقاً، إلا أننا نشعر أيضاً بأن التمايز بين التحصيل والاستعدادات يكون مفيداً عند مناقشة الأنواع المختلفة من القدرات.

الاختبارات الموضوعية والذاتية: تتضمن عادة الموضوعية التجرد أو عدم وجود تحيزات شخصية. وقد ذكر Cronbach (1990) أنه كلما قلَّ تأثر درجات الاختبار بالأحكام الذاتية للشخص القائم بتقدير درجاته، كان الاختبار موضوعياً. وبعبارة أخرى، تشير الموضوعية objectivity إلى مدى اتفاق القائمين بتقدير درجات اختبار معين، وتقديرهم درجات الاستجابات بنفس الطريقة. والاختبارات التي تشتمل على مفردات تتطلب اختيار الاستجابة (مثل: الاختيار من متعدد، والصواب أو الخطأ، أو الموازنة) والتي يمكن تقدير درجاتها باستخدام مفتاح تصحيح، مما يقلل الذاتية في تقدير الدرجات إلى أقل حد ممكن يشار إليها في أغلب الأحيان بأنها اختبارات "موضوعية". وعلى العكس من ذلك، فإن الاختبارات الذاتية هي تلك التي تعتمد على الأحكام الشخصية للقائم بتقدير الدرجات. فمثلاً، اختبارات المقال تعد اختبارات ذاتية، نظراً لأن الشخص القائم بتقدير درجات الاختبار يعتمد إلى حد ما على أحكامه (أحكامها) الذاتية. ومعظم الطلاب يدركون جيداً أن معلمين مختلفين ربما يقدروا درجات مختلفة لنفس سؤال المقال، فاختبارات المقال وغيرها من الصيغ الاختبارية التي تتطلب من الشخص القائم بتقدير درجات الاختبار توظيف أحكامه (أحكامها) الشخصية الذاتية، يشار إليها في أغلب الأحيان بأنها اختبارات "ذاتية Subjective". ومن الشائع والمرغوب فيه، أنه ينبغي على القائمين بتطوير اختبارات ذاتية تقديم موازين تقدير وصفية rubrics صريحة من أجل تقليل تأثير الأحكام الذاتية للشخص القائم بتقدير درجات الاختبار.

يعكس الأداء في اختبارات القوة صعوبة المفردات التي يستطيع أن يجيب على الفرد إجابة صحيحة.

يعكس الأداء في اختبارات السرعة الفروق في سرعة الأداء.

اختبارات السرعة واختبارات القوة: تُصنَّف عادة اختبارات الأداء الأقصى على أنها اختبارات سرعة أو اختبارات قوة. ففي "اختبار السرعة البحت Speed test"، يعكس الأداء فقط الفروق في سرعة الأداء. ويشتمل اختبار السرعة بعامة على مفردات سهلة نسبياً، وزمن إجابتها محدود مما يمنع أي مستجيب من استكمال إجابة جميع المفردات بنجاح. وفي اختبار "القوة البحت Power test"، لا

تعد سرعة الأداء مشكلة، فكل فرد يتاح له زمن كافٍ لمحاولة الإجابة عن جميع المفردات، ولكن