٢

المبادئ والطرق السايكومترية

لعل أحد أهم خصائص الإدراك البشري القدرة على تشريح وتحليل التجارب وعلى تفصيل أوجه وعناصر الإدراكات الحسية والمفاهيم بغرض التعامل معها بشكل فعال. إلا أن المشكلة في تحليل الأشياء إلى عناصرها أو مكوناتها المختلفة تكمن في اختلاف طريقة التحليل ونتائجه من شخصٍ لآخر. ويتطلب التواصل الفعال والعمل التعاوني بين الأفراد نوعاً من الإجماع على الإجراءات والمتغيرات التي نحلل بواسطتها تجاربنا الحياتية. وحالما نتمكن من وضع نظامٍ موحد للقواعد التي تنظم عمليات تحليل وتصنيف الظواهر التي نراقبها ونفكر ونتحكم بها، سيصبح استخدام عملياتنا العقلية على أكمل وجه ممكن. وسيمكننا هذا من أن نتحدث عن تجاربنا مع الآخرين وأن نستخدم هذه التجارب بطرق قديمة ومعروفة أوحديثة ومبتكرة. كما سيمكننا من أن نبيع هذه التجارب أو نشتري خلاصة تجارب الآخرين أو أن نجمع تجاربنا مع أشياء أخرى أو أن نتنبأ بالتجارب المقبلة. وإن إيجاد نظام للقياس يسهل كافة العمليات الآنفة الذكر.

obeikandl.com

القياس والإحصاء

القياس

يُعرّف القياس على أنه ربط الأشياء والأحداث بقيم عددية وهو قديمٌ قدم الإنسانية إن لم يكن أقدم[١]. ولقد كانت أولى أشكال القياس معنية بالخصائص الفيزيائية للأشياء مثل الطول والعرض والإتساع والزمن التي استخدمت في مجالات عملية وتطبيقية مثل الانتاج والبناء والتجارة وتوزيع الموارد. فعلى سبيل المثال، ساهمت الحسابات الفلكية التي قام بها البابليون وغيرهم من الشعوب القديمة في التخطيط لزراعة المحاصيل في الأوقات الأكثر خصوبة. كما طور الإغريق نظاماً رياضياً تجاوز حدود الاستخدامات العملية للقياس وشكل أساساً للعلوم النظرية والتطبيقية. ومن أحد العوامل الأخلاقية واللاأخلاقية التي ساهمت في التطورات التي حققها العالم القديم هو نظام العبودية الذي حرر بعض اليونانيين الأذكياء من تعب العمل لكسب المعيشة وبالتالي منحهم الوقت للتفكير في أمور أبعد مدىً من اعتبارات الحياة العملية اليومية لديهم.

يمكن للعلم، إلى حدٍ ما، أن يستمر دون تحديد المتغيرات بكميات (أو أعداد) معينة، إذ اعتمد شكل التفكير العلمي الذي أستخدمه أرسطو كلياً على الوصف والتصنيف اللاعدديين، كما استمرت العلوم وبالذات علم الأحياء لقرون بعد أرسطو في الوصف والتصنيف ووصلت هذه الطريقة إلى قمة تطورها في نظام لينييوس للتصنيف العلمي. يعتمد هذا النظام على وضع الأشياء ذات الخصائص المتشابهة ضمن الفئة ذاتها، وتتصف هذه الطريقة بكونها طبقية كما يوضح المقياس التاريخي العرقي الذي يبدأ بالمملكة ثم الشعبة أو الفصيلة ثم الصف ثم الطبقة ثم العائلة ثم النوع ثم الصنف ثم العرق. وتتشابه الأحياء المنتمية إلى

---

(١) تمتلك بعض الحيوانات والطيور قدرات حسابية بسيطة تمكنها من معرفة ما إذا نقص عدد بيوضها أم لا وتسمى هذه القدرة "subitism".

الصف ذاته أكثر من تشابه الأحياء المنتمية إلى نفس الشعبة وتتشابه تلك التي تنتمي إلى الصنف ذاته أكثر من تلك المنتمية إلى نفس النوع. وبعد أن تصنف النباتات أو الحيوانات وفق المقياس التاريخي العرقي وتعطى اسماً مناسباً (محدداً من حيث الصنف)،يتم التنبؤ بردود أفعالها وسلوكها بمقارنتها مع غيرها من الكائنات المنتمية إلى نفس الفئة. ويبقى الاتجاه الأرسطي (نسبةً إلى أرسطو) اتجاهاً محدوداً بالوصف لا يوفر أي تفسيرات.

وعلى شكل مغاير للأبحاث الأرسطي الوصفي/التصنيفي للعلوم، يتعامل الاتجاه الغاليلي (نسبة إلى غاليليو) الوظائفي مع المتغيرات والمبادئ التي تنطبق على الأشياء والأحداث ومع صياغة هذه المبادئ بأسلوب قانوني ورياضي. ومن الأمثلة على هذه المعادلات والصيغ هي القانون $s = {}^1/_2\, gf^2$ أو المسافة التي يقطعها جسم يسقط سقوطاً حراً في الفراغ يساوي نصف ثابت تسارع الجاذبية (g) مضروباً بمربع الوقت محسوباً بالثانية (t).

الرياضيات وعلم النفس

على عكس العلوم الفيزيائية، تميزت الجهود العلمية في مجال علم النفس وغيرها من العلوم الاجتماعية بكونها وصفية لم تتم صياغة مبادئها وقوانينها ضمن معادلات رياضية. ويستثنى من هذا التعميم القانون النفسي الفيزيائي الذي اقترحه غوستاف فنتشرمنذ أكثر من مئة عام وهو: $S = k\log R + c$ أو مقدارالإحساس(S) هو تابع لوغاريتمي لشدة عامل الإثارة (R).

إن مصطلح القياس النفسي، من وجهة نظر الفيلسوف إيمانيويل كانت، هو تناقضٌ في المصطلحات بحد ذاته، إذ يرى كانت أن القياس الموضوعي هو من خصائص العلوم وأن علم النفس هو علم ذاتي غير موضوعي يقتصر عمله على الظواهر الذاتية التي لا يمكن قياسها. إلا أن الدراسات التي قام بها

فينتشروبينيه وغيرهما في نهاية القرن التاسع عشر وبداية القرن العشرين أثبتت أنه يمكن التعبير عددياً وكمياً عن التجارب الواعية والسلوك البشري الملاحظ. وعلى الرغم من أن هذا القياس العددي ليس على نفس الدرجة من الدقة الموجودة في الأحداث الفيزيائية، إلا أنه يبقى نوعاً من القياس.

يُعرف القياس النفسي كما يستخدم في مجالات الاختبارات النفسية والتعليمية بأنه "عملية إعطاء قيم عددية (مثل درجات الاختبارات) للأشخاص بطريقة تسمح لخصائص هذه الأعداد أن تعكس وبشكل دقيق صفات هؤلاء الأشخاص" (مورفي وديفيدشوفر، ١٩٩٤، ص. ٣٤). ويجب الملاحظة هنا أن هذا التعريف لا يدّعي بأن علم النفس قادر على قياس كافة أفكار الفرد ومشاعره وأنماط سلوكه إذ أن أهداف القياس النفسي لا تطمح إلى كل هذا بل تكتفي برسم لوحاتٍ تجريدية تعبّر عن الاختلافات في ردور أفعال الأفراد الداخلية والخارجية.

وكما ذكرنا في الفصل الأول، بدأ التركيز العلمي على قياس الاختلافات بين الأفراد في أعمال فرانسيس غالتون ولا زال هذا التركيز هدفاً أساساً للقائمين على عمليات القياس النفسي. ويجب على المقاييس النفسية لكي تحقق أهداف التنبؤ بطبيعة الأشخاص والسيطرة على بعض طرق سلوكهم أن تستخدم مقاييس ثابتة، ولتحقيق هذا الثبات يجب أن تكون أدوات القياس مستقرة تحت الظروف ذاتها (هذا في حال أن بيئة الفرد لم تتغير بشكل ملحوظ). وعليه فإن المقاييس الثابتة هي تلك التي تتطلب تطبيق إجراءات القياس ذاتها، أي تتطلب قياس الاختبارات. وإضافة إلى ذلك يشترط في المقاييس أن تكون صادقة: أي يجب أن توفر قيم عددية صحيحة تقيس الخاصية أو المتغير المراد قياسه.

"العملياتية" Operationalism

تتصف جميع أنظمة القياس بكونها افتراضية بعض الشيء، إذ أن المؤشر الذي تقع عليه المقاييس لا وجود له كصفاتٍ ملموسة في الواقع، وإن الأرقام الموجودة على المؤشر هي فقط طريقة عملية وسهلة لتمثيل نوعية أو شدة الشيء أو الحدث موضع القياس. وإن المتغيرات التي يتم قياسها ثم تحديد قيم عددية موافقة لها ضمن ظروف معينة أو ضمن إطار زمني معين هي ليست إلا ابتداعات خيالية ذات فائدة عملية تساعدنا في تبسيط ملاحظاتنا وتعاملاتنا وفهمنا لأنفسنا وللعالم من حولنا. ويسمى الميل إلى التصرف على أساس أن هذه المفاهيم والمجردات هي أشياء ملموسة لها وجود في الواقع وتدعى Reification (أي تحويل المجردات إلى أشياء مادية ملموسة)، وقد كان هذا المحفِّز الأساسي لنظرية بريدجمان العملياتية.

تركز نظرية العملياتية على أن إلتباس معاني المتغيرات أو المفاهيم كمفهوم القوة مثلاً يمكن تفاديه عن طريق تعريف هذه المفاهيم من حيث العمليات الضرورية والتي يمكن ملاحظتها التي تؤدي إلى شعورنا أو مرورنا بتجربة تتجلى فيها هذه المفاهيم. وببساطة يعد التعريف العملياتي تعريفاً يدلنا على ما يتوجب علينا فعله أو على الإجراءات التي يجب أن نتبعها بغرض تحفيز الشعور المراد تعريفه في داخلنا. فعلى سبيل المثال، يتطلب التعريف العملياتي للطول مثلاً إعطاء شخصٍ ما تعليمات بكيفية قياس ما يسمى بالطول، ويتطلب التعريف العملياتي للقلق وصفاً دقيقاً لتصرفات الفرد القلق وللأحداث أو المحفزات التي أثارت الشعور بالقلق عند هذا الشخص.

وتساعد التعريفات العملياتية للمفاهيم في تسهيل عملية التواصل وتوحيد معاني المفاهيم. إلا أن المبالغة في استخدام هذه التعريفات قد تؤدي إلى عقم في عملية البحث العلمي، إذ أن المفاهيم العلمية لا يمكن تعريفها دائماً بطريقة

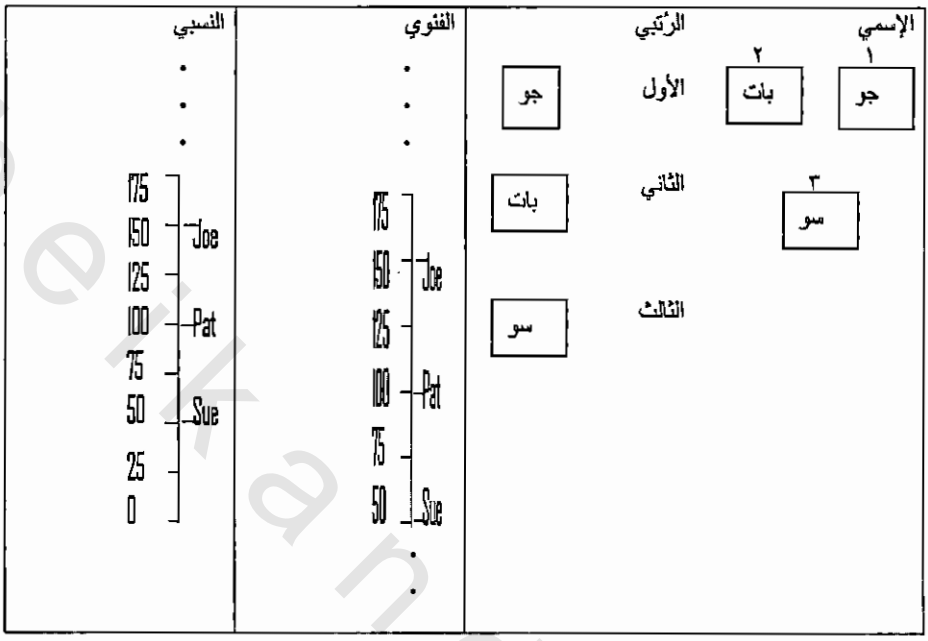
عملياتية، على الأقل ليس فوراً، وإن التأكيد الزائد على ربط المتغيرات بما يمكن ملاحظته ومراقبته وتعريف هذه المتغيرات على أساس الدلالة (أو الإشارة) يمكن أن يثبط التخمين المبدع والتفكير التجريدي. وقد أكدت الحركة العملياتية في علم النفس على أن مفاهيماً كالقلق والتعلم والإحباط لا تعني الشيء ذاته للأفراد المختلفين ولذلك فإن هذه المفاهيم لا تعبر عن التجارب ذاتها. وبما أن العلم الجيد يرتكز إلى توثيق الملاحظات والتنبؤات على أرض الواقع،يتوجب على العلماء أن يصفوا ما تمت ملاحظته أو ما قد حصل في موقف ما بأوضح شكل ممكن. ولعل أفضل طريقة لفعل هذا، وذلك عندما تكون الظاهرة موضع الدراسة يمكن تحديد مقدارها، هي بالإشارة بدقة متناهية إلى كيفية قياس المتغيرات المراد قياسها.

مؤشرات القياس Scales of Measurements

كما أشرنا سابقاً في هذا الفصل، إن أبسط تعريف للقياس هو الإشارة إلى الأحداث بواسطة الأرقام. واستناداً إلى هذا التعريف تعتبر الأرقام التي يحملها الرياضيون والتي تساعدنا على رصد ومتابعة أدائهم نوعاً من القياس. وبما أن جمع أو طرح أو تقسيم أو ضرب هذه الأرقام عملية لا تعني شيئاً، فإن هذا النوع من القياس يمثل أدنى درجات القياس أو ما يسمى بالمؤشر الإسمي. وبالنظر إلى الحقل الأول من الشكل ٢-١ يمكن القول أن جو وبات وسو ينتمون إلى فئات مختلفة من المتغير المحدد أعطيت الأرقام ١ و٢ و٣، وإن هذه الأرقام هي افتراضية تماماً ولا يمكن تحليلها بشكل ذو معنى إلا بإعتبارها أدوات مناسبة للتعرف على هؤلاء الأفراد.

الشكل ٢-١ أربع مؤشرات للقياس

مستويات القياس



يعتبر المقياس الرُتبي إلى حد ما مستوىً أعلى من مستوى القياس الإسمي لأنه يرتب ما يقاس من حيث الأول فالثاني فالثالث... إلخ. ومن الممكن أن تمثل هذه الأعداد الترتيبية تسلسل النجاح في مسابقة ما أو غيرها من المتغيرات التي ينال المشتركون فيها رتب مختلفة. ومن الملاحظ في الشكل ٢-١ أن جو أحتل المركز الأول على حين أحتل بات وسو المركزين الثاني والثالث على التوالي. وإن عملية الجمع بين هذه الأعداد بواسطة عمليات رياضية هي عملية لا معنى لها. وبما أن المسافة بين الأعداد غير معروفة فإنه لا يمكن القول بأن الفرق بين جو وبات لا يساوي الفرق بين بات وسو، ولذلك فإن كل ما يمكن الخلاص إليه في هذه الحالة هو أن جو أفضل من بات وبات أفضل من سو.

obeikandl.com

تصبح الأعداد أكثر دقة ووضوحاً عند إنتقالنا لمستوى القياس الفئوي، فكما نلاحظ في العمود الثالث من الشكل ٢-١، لا تتحدد مؤشرات القياس المقسمة بالتساوي بحدود عليا وحدود دنيا. وعلى عكس المقياس الرتبي، تتوافق الفوارق بين الأعداد الموجودة على المؤشر مع الفروقات المتساوية في أداء الممتحنين في المتغير موضع القياس. وبما أن معدل درجات جو هو ١٥٠ ومعدل درجات بات هو ١٠٠ ومعدل درجات سو هو ٥٠، فإن الفرق بين جو وبات يساوي الفرق بين بات و سو في الصفة التي تم قياسها. وتعد مقاييس الفهرنهايت والسيلسيوس لدرجات الحرارة أمثلة شائعة للقياس الفئوي. وفي مجال القياس النفسي يمكن إعتبار مؤشرات الدرجات المعيارية التي تعبر عن الأداء في الاختبارات الإدراكية مقاييساً فئوية.

وعلى حين يمكن طرح وجمع الأعداد الموجودة على المؤشر الفئوي، لا يسمح بإستخدام عمليتيي القسمة والضرب في هذا المستوى من القياس إذ يتوجب على الأعداد أن توضع على مؤشر نسبي حتى يمكننا أن نقول أن معدل درجات بات (١٠٠) هو ضعف معدل درجات سو (٥٠) وأن معدل درجات جو (١٥٠) هو ثلاثة أضعاف معدل درجات سو.وكما يوضح العمود الرابع من الشكل ٢-١ فإن المؤشر النسبي له صفر مطلق أو أصل حقيقي تكون قيمة المتغير أو الصفة المقاسة فيه تساوي الصفر. وعليه، يمكن التعبير عن قيمتين عدديتين بعيدتين بمسافة محددة عن الصفر على مقياسي الطول (أمتار) والوزن (غرامات) كمعدل نسبة، كالقول مثلاً إن الدرجة ١٠٠ تمثل ضعف كمية الصفة التي تم قياسها بالدرجة ٥٠ وهي أيضاً تمثل ثلثي الدرجة ١٥٠، ومن الملاحظ أن السقف أو الحد الأعلى لمؤشر القياس النسبي هو لامحدود ولكن هذا لا يؤثر على خصائصه المتعلقة بالنسب. لاتُستخدم المؤشرات النسبية على نحو واسع في مجال القياس النفسي

باستثناء مقياس "سون" Sone لقياس علو الصوت ومقياس "مل" Mel لقياس طبقة الصوت.

الطرق الإحصائية

يقدم الجزءان الأول والثاني من هذه السلسة وبالتحديد الفصل الرابع من كتاب سلالم التقدير وقوائم الرصد (أيكين، ١٩٩٦) والفصلان ٤ و٥ من كتاب الاستبانات ومقاييس السمات (أيكين، ١٩٩٧) ملخصاً لابأس به عن طرق الإحصاء الأولية التي استخدمت لوصف عينات من المعلومات واستنتاج المؤشرات العامة لباقي السكان. وتعد معظم المفاهيم والإجراءات الإحصائية التي تناولتها الفصول المذكورة أعلاه مناسبة لتحليل البيانات الناتجة عن الاختبارات وغيرها من أدوات التقييم، كما توفر العديد من كتب الإحصاء الأولي (مثل: كايس، ١٩٩٦؛ باغانو، ١٩٩٤؛ شافيلسون، ١٩٩٦) ملخصات جيدة عن طرق الإحصاء البسيطة. وتتضمن أدلة المستخدم التي ترافق العديد من برامج الحاسب الإحصائية (مثل: نوروسيس، a١٩٩٢، b١٩٩٢، c١٩٩٢ على معلومات وصفية عن عمليات إحصائية متقدمة. ولن ندخل في تفاصيل الإحصاء هنا نظراً لوفرة المصادر حول هذا الموضوع، ولكن يبقى التعرف على إجراءات إحصائية مختارة لتحليل البيانات الناتجة عن الاختبارات النفسية وإعطاء لمحة عنها أمراً مفيداً وضرورياً.

إن أحد أول الخطوات المستخدمة في عملية تحليل بيانات الاختبارات النفسية هي التوزيع التكراري للدرجات التي حصل عليها الأفراد المختلفون، ويمكن تحقيق هذا من خلال العديد من برامج الحاسب كبرنامج A-1 الموجود على القرص المرن المرفق بهذا الكتاب. ويعطينا هذا البرنامج تمثيلاً تصورياً

للتوزيع التكراري للنتائج عن طريق تحويل النتائج إلى مدرج تكراري أو مضلع للتوزيع التكراري. ويمكن فحص شكل هذه الرسوم البيانية بما في ذلك عدد الحد الأعلى للنقاط (المنوالات) ومدى تباعد أو اختلاف الدرجات وما إذا كان التوزيع متناظراً أم منحرفاً أو مسطحاً أم منحدراً. ويمكن لحساب الوسط الحسابي والوسيط والتباين والانحراف المعياري ومعاملات الالتواء والتفرطح عن طريق برنامج A-2 أن تكمّل المعلومات الناتجة عن طريق الفحص البصري للتوزيع التكراري.

ولعل معاملات الارتباط هي من أكثر الإحصائيات المتوفرة في الدليل المستخدم في الاختبارات النفسية بالإضافة إلى جداول الدرجات المعيارية والمئوية. وتوجد أنواع عديدة لمعاملات الإرتباط يمكن حسابها بواسطة برامج الفئة Cالموجودة على القرص المرن المرفق بكتاب درجات التقدير وقوائم الرصد وبرامج الفئة E في كتاب الاستبانات وقوائم السمات. ويساعد برنامج A-3 الذي يحتويه القرص المرافق لهذا الكتاب في رسوم بيانية تبعثرية أو خطوط للارتداد الخطي لواحد من متغيرين بالمقارنة مع الآخر. كما يحسب برنامج A-4: (١) معامل الإرتباط المولد للعزوم بين متغيرين و(٢) الوسيط الحسابي والانحراف المعياري للمتغيرات و(٣) معادلة التراجع الخطي للتنبؤ بالمتغير التابع (Y) من المتغير المستقل (X) و(٤) عدد الأخطاء الإيجابية غير الحقيقية والأخطاء السلبية غير الحقيقية، التوقعات الصائبة وعدد حالات الرفض الصحيحة المرتبطة بحد للدرجات الفاصلة على المتغير (المعيار) التابع. ويمكن للبرنامج أن يساعد في دراسة صدق الاختبار في التنبؤ بمتغير معياري ما.

يمكن إجراء عمليات أخرى على مجموعة من معدلات الدرجات الناتجة عن عدد من الاختبارات مثل: تحليل الإنحدار المتعدد والتحليل العنقودي والتحليل

العاملي والتحليل التمييزي. ومن العمليات الأقل شيوعاً والتي قد تكون مفيدة عند وجود عدة متغيرات يمكن ذكر عملية تحليل الارتباط القانوني وتحليل المسار وتحليل البنية الكامنة. تقود النتيجة النهائية لتحليل الإنحدار المتعدد إلى معادلة إنحدارية يتم استخدامها في توقع درجات الأفراد في متغير تابع (معيار) بالاعتماد على درجاتهم في متغيرين مستقلين (أداوات التنبؤ) أو أكثر. وتتوفر عدة إجراءات لاختيار المتغيرات المستقلة التي سيتم استخدامها ضمن معادلة الإنحدار التعددي النهائية (كالاختيار الأمامي والحذف التراجعي والاختيارالمرحلي) (راجع نوروسيس، ١٩٩٢a). ويتم إدخال جميع المتغيرات المستقلة (١-٣) بشكل متزامن في معادلة الإنحدار باستخدام برنامجA-5 الموجود على القرص المرافق لهذا الكتاب. كما يفيد برنامج A-5 في حساب التالي: قيم الانحدار القياسية وغير القياسية والأخطاء المعيارية لقيم الإنحدار والنتائج الإحصائية للاختبارات (t) لدلالات قيم الإنحدارومعامل الإرتباط المتعدد (R).

وعلى عكس تحليل الإنحدار المتعدد الذي يحتوي على متغير تابع واحد، يقوم تحليل الإرتباط القانوني على عدد من المتغيرات المستقلة والتابعة وذلك بهدف تحديد التركيب الخطي الأمثل لعدد من المتغيرات المستقلة الموزونة بغرض تحقيق الصحة الأمثل للتنبؤ بمجموعة من المتغيرات التابعة الموزونة عددياً. ومن الجدير بالذكر أن عملية تصنيف المتغيرات كمتغيرات مستقلة أو تابعةفي تحليل الإرتباط القانوني بين مجموعتين من المتغيرات عملية افتراضية تماماً.

تتطلب الإجراءات الإحصائية المتعددة المتغيرات كالتحليل العاملي وتحليل التمييزي جهداً حسابياً هائلاً، إلا أن مجموعة برامج الحاسب +SPSS/PC تنجز هذه العمليات بسهولة وسرعة (نوروسيس،١٩٩٢b)، وقد صممت إجراءات التحليل العاملي لتحديد العدد الأدنى من العوامل (الأبعاد) المطلوبة لتفسير

مجموعة معدلات درجات حصل عليها عدد nمن الأفراد في عدد من اختبارات m. وتعد الطرق التالية من أكثر طرق التحليل العاملي استخداماً: طرق المحور الرئيسي والحد الأدنى للمربعات غير الموزونة والاحتمالية القصوى وإجراءات ألفا وطرق التحليل العاملي للصور[٢]. وتنقسم عملية التحليل العاملي إلى مرحلتين، تتألف أولاهما من عملية استخلاص العوامل من مصفوفة بيانات الدرجات أومعاملات الإرتباط، بينما تنطوي المرحلة التالية على تدوير العوامل المستخلصة لتصبح بنى بسيطة يمكن تفسيرها بشكل ذو معنى. وتعد طرق فاريماكس وإيكواماكس وكواتريماكس أكثر طرق التدوير شيوعاً.

ويشتمل الجدول ٢-١ على قائمة لأحمال ثلاثة معاملات تم تدويرها استخلصت عن طريق تحليل معدلات درجات ٢٠٠٠ شخص على مقياس ويكسلر المعدل لذكاء البالغين (WAIS-R) وهذه الأحمال هي بالضرورة إرتباطات بين ١١ اختباراً فرعياً لمقياس WAIS-R مع معاملاتها، إذ ترتفع أهمية معامل ما في بنية الإختبار الفرعي بارتفاع أحماله. ويطلق على العامل رقم ١ اسم العامل اللغوي بسبب ارتفاع أحمال المعلومات والمفردات والإستيعاب والمتشابهات (وهي اختبارات فرعية لغوية) فيه. ونظراً لارتفاع أحمال إكمال الصور وترتيب الصور وتصميم المجسمات وتجميع الأجسام في العامل ٢، ونظراً لطبيعة هذه المهام يجب أن تطلق تسمية العامل الفراغي -الإدراكي - الحسي أو عامل الصور الفراغية على هذا العامل. وأخيراً، ترتفع أحمال المدى الرقمي والحساب والرموز العددية على العامل ٣، مما يعني أن اسم عامل المهارة العددية هو اسم مناسب لطبيعة هذا العامل.

---

(٢) وهي أحد طرق استخراج العوامل من المعلومات المعطاة. وتعتمد هذه الطريقة على استخراج العوامل من مصفوفة الإرتباط للمتغيرات المتنبأ بها وليس للمتغيرات الفعلية. (ملاحظة المترجم).

يهدف التحليل التمييزي لدرجات مجموعات من الأشخاص p في مهام معينة m إلى تحديد التركيبة الخطية للدرجات في الاختبارات m التي تساعدنا على التمييز بين المجموعات p. ويمكن مثلاً إجراء تحليل تمييزي لدرجات أفراد من أربع فئات عرقية (بيض، سود، آسيويون، إسبانيو الأصل) في بطارية اختبار للاستعدادات المتعددة وذلك بهدف تحديد التركيبة المرجحة الأمثل لمعدلات الدرجات القادرة على التنبؤ بأعضاء كل جماعة عرقية.

الجدول ٢-١

مصفوفة العامل المدور للاختبارات الفرعية لمقياس WAIS-R

الاختبار الفرعي	العامل ١	العامل ٢	العامل ٣
المعلومات	٠,٨١	٠,٢١	٠,٢٩
المدى الرقمي	٠,٢٧	٠,١١	٠,٨٣
المفردات	٠,٨٢	٠,٢٢	٠,٣٤
الحساب	٠,٤٩	٠,٢٨	٠,٥٩
الاستيعاب	٠,٧٩	٠,٢٣	٠,٢٦
المتشابهات	٠,٧٥	٠,٣٠	٠,٢٥
إكمال الصور	٠,٤٨	٠,٦٤	٠,١١
ترتيب الصور	٠,٥١	٠,٤٩	٠,١٣
تصميم المجسمات	٠,٢٦	٠,٧٢	٠,٣٨
تجميع الأجسام	٠,١٤	٠,٨٥	٠,١٩
الرمز العددي	٠,٢٣	٠,٣٧	٠,٦١

توفر عمليات التحليل التمييزي عدداً من المعلومات الإحصائية مثل معاملات الدالة التمييزية لعدة اختبارات وللدرجات التمييزية التي يتم حسابها من هذه المعاملات. ويساعد جدول تصنيفي ذو اتجاهين يبين عدد الأشخاص الذين تم تصنيفهم ضمن مجموعة معينة بشكل صحيح بناءً على درجاتهم الدالة التمييزية وعدد الأشخاص الذين تم تصنيفهم بشكلٍ خاطئ في إعطاء ملخص جيد عن فعالية درجات الاختبارات في التمييز بين المجموعات.

وكما هو الحال في تحليل العامل، يُعنى تحليل البنية الكامنة بتحديد درجة إمكانية تفسير العلاقات الملاحظة بين نتائج اختبارات p من خلال عدد من

المتغيرات الكامنة (أو غير الملاحظة). ويتعامل إجراء تحليل المسار الإحصائي مع العلاقات بين المتغيرات أيضاً، إلا أن التركيز في هذا الإجراء هو على تحليل الإرتباط بين مجموعة من المتغيرات بهدف تحديد أي من هذه المتغيرات تكون محتملة المساهمة في غيرها من المتغيرات أو مسببات لغيرها من المتغيرات. وتوفر حسابات تحليل المسار معلومات عن المسار الذي يأخذه متغير ما عند مساهمته السببية في متغيرات أخرى، كما توفر معلومات عن أهمية هذه المساهمة.

نظرية وطرق الاختبار التقليديَة

يشترط في الاختبارات النفسية احتواؤها على عدد من الصفات، إلا أن أهم هذه الصفات هي صفة الصدق، إذ يمكن للاختبار أن يكون على درجة عالية من الكفاءة والثبات في القياس إلا أن هذا لا يعني شيئاً في حال فشل الاختبار في قياس ما قد صُمم لقياسه. وعندما يتم تصميم اختبار ما لتحقيق عدة أهداف يمكن أن يكون الاختبار صادقاً في قياس بعض هذه الأهداف ولكن ليس لأهداف أخرى. كما من الممكن أن يكون الاختبار مؤشراً دقيقاً وحساساً عن قدرة إدراكية معينة تحت ظروف معينة دون أخرى، أوأن يكون مقياساً أو أداة تنبؤ صادقة مع عينات معينة من الناس دون أخرى.

أنواع الصدق في الاختبارات:

يمكن الحصول على دلائل على صدق اختبار ما بالطرق التالية:

١- بواسطة تحليل محتوى الاختبار (صدق المحتوى).

٢- عن طريق دراسات تجريبية على درجة الإرتباط بين درجات الاختبار والدرجات على مقياس للمعيار تم الحصول عليه في نفس الوقت (الصدق التلازمي) أو يمكن الحصول عليه في المستقبل (الصدق التنبؤي).

٣- عن طريق تحليل درجة التطابق بين قدرة الاختبارعلى التمييز بين الأشخاص مع غيره من المقاييس لذات المفهوم العقلي أو تطابقه مع المواقف التي يكون فيها هذا المفهوم فعالاً (صدق المفهوم). وتستخدم الطرق الثلاث المذكورة أعلاه في إثبات قدرة الاختبار على قياس أو التنبؤ بالقدرات الإدراكية والسلوك في مجالات التعليم والتوظيف والطب وغيرها.

ركزت عمليات إثبات صدق الاختبارات التحصيلية المعايَرة بشكل رئيس على صدق المحتوى، إذ يتطلب وضع اختبار تحصيلي جيد إعداداً مسبقاً لجدول بالمواصفات أو مخطط موجز عن المحتوى المراد اختباره. وتعد مقارنة المحتوى النهائي للاختبار مع هذا الجدول الخطوة الأولى في عملية إثبات صدق محتوى الاختبار. كما يمكن الإستعانة بخبراء ومتخصصين في المادة موضع الاختبار لإبداء الرأي في درجة تطابق محتوى الاختبار مع أهدافه. ويمكن لعملية إثبات صدق المحتوى أن تقتصر على إصدار أحكام ذاتية تقرر ما إذا كانت أسئلة الاختبار تمثل المادة التعليمية موضع الاختبار بشكل جيد.

وعلى الرغم من أهمية تحليل محتوى الاختبار في تقييم مختلف أنواع الأدوات النفسية، يعد صدق المحك (والذي يتضمن الصدق التلازمي والصدق التنبؤي) الصفة الأساسية للاختبارات الجيدة التي تقيس الاستعدادات العامة والخاصة. ويجب تقديم دلائل على الصدق التنبؤي لاختبار ما قبل السماح باستخدام هذا الاختبار كأداة لاختيار أو تنسيب الموظفين والطلاب. وفي حال أثبتت الدراسات أن الاختبار تمكن من المساهمة بإزدياد فعالية التنبؤ بمحك الأداء المقبول بشكل اقتصادي، أي أن الاختبار يتميز بالصدق المتزايد، يتم تحديد معدل الدرجات الفاصلة للاختبار ولا يتم قبول أي ممتحن لايحقق هذا المعدل.

تستخدم أساليب الارتباط والإنحدار بشكل واسع في تحليل الصدق التنبؤي للاختبار، إذ غالباً ما ترتبط درجات الاختبارات التنبؤية بدرجات مقياس المحك أو تُستخلص منها. ويؤدي ارتفاع معامل الإرتباط بين أداة التنبؤ ومتغيرات المحك إلى ارتفاع دقة التصورالمستقبلي لدرجات المحك بالإستناد إلى درجات التنبؤ. ويمكن حساب، بالإضافة إلى الإرتباط بين الاختبار ومقياس المحك (r)، معادلة الإنحدار الخطي للتنبؤ بدرجات المحك بناءً على درجات الاختبار والخطأ المعياري للتقدير. ويمكن حساب الخطأ المعياري للتقدير (Sset) عن طريق المعادلة التالية:

$$s_{est} = s_y\sqrt{1-r^2} \qquad (٢,١)$$

حيث Sy هي الإنحراف المعياري لمتغير المحك. ويمكن استخدام قيمة Sset في حساب مجال الخطأ في الدرجات التنبؤية للمحك والتي يمكن، وبدرجة معينة من الثقة، القول بأن الدرجات الناتجة للمحك تقع ضمنها حتماً.

ونظراً لأن الاختلاف بين درجات المحك التنبؤية والفعلية ممكن جداً فمن الجائز أن تقع أحدى الدرجات تحت حد الدرجات الفاصلة بينما تقع الثانية فوقه، وفي كلتا الحالتين يكون هناك خطأ في التنبؤ. ففي حال وقوع الدرجات التنبؤية للمحك فوق حد الدرجات الفاصلة ووقوع الدرجات الفعلية تحته ينتج ما يسمى بالخطأ الإيجابي غير الحقيقي، وفي حال وقوع الدرجات التنبؤية للمحك تحت حد الدرجات الفاصلة ووقوع الدرجات الفعلية فوقه ينتج ما يسمى بالخطأ السلبي غير الحقيقي. ومما لا شك فيه أن نقطة الحد الفاصل لمحك ما تؤثر على عدد الأخطاء الإيجابية والأخطاء السلبية الناتجة، إذ أن تحديد نقطة متدنية للحد الفاصل سوف يؤدي إلى عدد أكبر من الأخطاء الإيجابية غير الحقيقية بينما يؤدي تحديد نقطة مرتفعة للحد الفاصل للدرجات إلى عدد أكبر من الأخطاء السلبية غير الحقيقية. وتلعب عوامل أخرى دوراً في تحديد

الأخطاء السلبية والإيجابية غير الحقيقية. وتتمثل هذه العوامل في كل من: صدق الاختبار ونسبة المتقدمين الذين يجب اختيارهم (نسبة الاختيار)، والنسبة الدنيا (وهي معدل عدد المقبولين من مجمل تعداد العينة).

تساعد برامج الكمبيوتر A-3 وA-4 وA-5 الموجودة على القرص المرافق لهذا الكتاب في عملية تحليل الصدق التنبؤي للاختبارات إذ تحتوي هذه البرامج على إحصائيات للإرتباط والإنحدار تساعد على التحليل الإنحداري الخطي البسيط والمتعدد. وبالإضافة إلى تحديد معاملات الإرتباط ومعادلة الإنحدار الخطي، تساعد برامج A-4 وA-5 في حساب الخطأ المعياري للتقدير والاختبارات الإحصائية المناسبة لتحديد أهمية القيم المرجحة للإنحدار. ويحسب برنامج A-4 عدد الأخطاء السلبية والإيجابية غير الحقيقية المرتبطة بحد الدرجات الفاصلة.

يعد صدق المفهوم Canstruct Validity من أهم أنواع الصدق المرتبطة باختبارات الشخصية وغيرها من الاختبارات. ويعرّف صدق المفهوم على أنه قدرة الاختبار أو أداة التقييم على قياس سمة أوخاصية معينة تعتبر أساساً لتحديد السلوك البشري في مواقف معينة. فعلى سبيل المثال، تميز اختبارات القدرة اللغوية بين الأفراد الذين يمكن اعتبارهم مؤلفين ناجحين أو غير ناجحين بناءً على الاعتقاد بأن المؤلف الناجح يتمتع بقدرات لغوية أكثر تطوراً من المؤلف غير الناجح. إلا أن اثبات صدق المفهوم هو عملية أكثر تعقيداً من ذلك إذ أنها تتطلب عدداً من المواقف وعدداً من الدلائل إذ يعتمد صدق المفهوم على نتائج مجموعة من التحريات هدفها اثبات أن الأشخاص الذين حصلوا على درجات عالية في الاختبار تتوافق تصرفاتهم في عدد من المواقف مع ما هو متوقع من الأفراد الذين "يمتلكون" كمية كبيرة من هذا المفهوم. وفي الوقت ذاته تتوافق تصرفات أولئك الذين حصلوا على معدلات درجات متدنية في الاختبار مع ما هو متوقع من أولئك الذين يمتلكون كمية ضئيلة من المفهوم. ويشير صدق المفهوم إلى أن الارتباط بين

مقاييس المفهوم ذاته التي تم الحصول عليها بطرق مختلفة (المصداقية التقاربية) أعلى من الإرتباط بين عدة مفاهيم تم قياسها بالطريقة ذاتها (المصداقية التمييزية).

الثبات Reliability:

لا يمكن لاختبار أن يكون صادقاً إذا لم يقس المتغير ذاته في كل مرة.لا يعد ثبات الاختبار (أو توازن وثبات القياس)، شأنه كشأن الصدق، خاصية إما أن تكون موجودة في كافة أقسام الاختبار أو أن لا يكون لها وجود على الإطلاق، إذ يمكن للاختبار أن يكون أداة قياس ثابتة تحت ظروف معينة دون أخرى. وفي حال عدم كون الاختبار مقياساً ثابتاً لما يراد قياسه في الوقت الذي يجب أن يتم فيه القياس، لا يمكن اعتبار هذا المقياس صادقاً للخاصية أو الصفة أو السلوك موضع القياس. الثبات، على أية حال، هو شرط هام ولكن غير كافٍ لاثبات صدق الاختبار.

يرتبط ثبات الاختبار بكون معدلات الدرجات الناتجة عن هذا الاختبار معدلات ثابتة نوعاً ما على مر الزمن وتحت ظروف قياس متنوعة، وهذا لا يعني أن نتائج الاختبار لا تختلف بازدياد علم أو نضج أو مهارة الأفراد. وتجدر الإشارة إلى أن التركيز في عملية إثبات ثبات الاختبار يقع على الثبات ضمن مرحلة زمنية معينة في حين أنه في حالة الدراسات التطورية ينصب الاهتمام على مراقبة تطور الأفراد مع مرور الزمن. ويفترض في هذه الحالة أن نتائج الفرد بين فترة اختبار وأخرى تبقى ثابتة إلى حد ما وأنها خالية نسبياً من أخطاء القياس.

تفترض نظرية الاختبار التقليدية أن الدرجات الملاحظة للفرد (O) هي مجموع الدرجات الفعلية (T) وخطأ القياس (E)[٣] كما تفترض كذلك عدم وجود

(٣) تفترض نظرية الثبات التقليدية أن العلاقة بين الدرجات الفعلية وأخطاء القياس هي علاقة بين شيئين منفصلين يمكن إضافة الواحد منهما للآخر. وهذا في واقع الأمر قد =

ارتباط بين الدرجات الفعلية والأخطاء مما يعني أن تباين الدرجات الملاحظة يساوي مجموع تباين الدرجات الفعلية (الحقيقية) وتباين خطأ القياس:

$$s_o^2 = s_t^2 + s_e^2 \qquad (٢-٢)$$

تعرّف الثقة (r_{11}) بناءً على النموذج أعلاه على أنها نسبة تباين الدرجات الفعلية إلى تباين الدرجات الملاحظة:

$$r_{11} = s_t^2 / s_o^2 \qquad (٢-٣)$$

ومن الجدير بالذكر أن معظم المعادلات في نظرية الثبات التقليدية تستند إلى المعادلتين أعلاه.

لا يمكن حساب معامل الثقة بشكل مباشرة باستخدام المعادلة ٢-٣ بسبب جهل التباين الفعلي لدرجات الاختبار، إلا أن عدداً من طرق الحساب غير المباشرة قد صممت لتقدير ثبات الاختبارات. ويمكن تحديد أربعة أنواع من تقديرات الثبات هي: معاملات الثبات، والتكافؤ، والثبات والتكافؤ، والاتساق الداخلي. يُحسب معامل الثبات أو معامل الاختبار - إعادة الاختبار عن طريق إجراء الاختبار لذات المجموعة من الأشخاص مرتين مختلفتين ثم إيجاد الإرتباط بين الدرجات في المرتين. أما معامل التكافؤ أو معامل الصور المتكافئة فيتم

= يكون افتراضاً غير صحيح إذ يمكن للعلاقة بين الدرجات الفعلية وأخطاء القياس أن تكون علاقة بين شيئين تابعين أو مترابطين يمكن مضاعفتهما. فعلى سبيل المثال، يمكن للدرجات الملاحظة لعدد من الأفراد يتمتعون بقدرة إدراكية عالية أن تحتوي على نسبة خطأ أعلى من معدلات الأفراد ذوي القدرات الإدراكية الضعيفة. وتساهم عوامل كالدافع الداخلي أو الرغبة في أخذ الاختبار والسرعة وتوزيع الوقت وغيرها من العوامل اللاإدراكية، التي تتنوع بتنوع مستويات أداء الأفراد في أخطاء القياس الإدراكي. وتعطي الافتراضات التجميعية والمستقلة لنظرية الثبات التقليدية أبسط تصور لمفهوم نموذج العلاقة بين الدرجات الفعلية وأخطاء القياس، وبالرغم من توافق هذا التصور مع النموذج الخطي العام في نظرية الإحصاء إلا أنه ليس بالضرورة أن يكون التصور الأكثر صحة.

obeikandl.com

حسابه عن طريق تصميم صورتين متشابهتين من الاختبار واستخدامهما لاختبار المجموعة نفسها من الممتحنين ثم حساب الارتباط بين درجاتهما. كما يتم حساب معامل الثبات والتكافؤ عن طريق إجراء صورتين متشابهتين من الاختبار للمجموعة ذاتها من الأفراد في توقيتين مختلفين. وللتحكم في تأثيرات نظام إدارة الاختبار يتم تقسيم المتقدمين للاختبار إلى مجموعتين: المجموعة الأولى تأخذ الاختبار (أ) في الوقت (١) والاختبار (ب) في الوقت (٢)؛ أما المجموعة الثانية فتعطى الاختبار (ب) في الوقت (١) والاختبار (أ) في الوقت (٢). ويتم بعد ذلك حساب الإرتباط بين درجات نموذجي الاختبار للمجموعتين.

ولا تخلو الطرق الثلاثة المذكورة أعلاه من العلل، إذ يمكن لطريقة الاختبار وإعادة الاختبار أن تعطي بعض الممتحنين (وليس جميعهم) خبرة في الاختبار مما قد يؤدي إلى تحسن درجاتهم في المرة الثانية مما يؤثرعلى حساب الإرتباط بين الدرجات في كلتا المرتين. وأما في الطريقتين الثانية والثالثة فيعتبر إعداد نموذج مشابه أو صورة مكافئة للاختبار عملية معقدة وشاقة ولا تضمن التكافؤ الحقيقي بين نموذجي الاختبار. ولهذه الأسباب تم بذل جهود منذ بدايات عملية القياس النفسي إلى ترشيد عملية تقدير ثبات الاختبارات. ومن ضمن هذه الجهود محاولة إيجاد طرق لتحديد الاتساق الداخلي لاختبارٍ ما - أو درجة قياس العناصرالتي يتألف منها الاختبار للمتغير ذاته.

يفترض في الاختبارات التي تحتوي على عدد كبير من الفقرات التي تقيس الشيء ذاته أن تكون على درجة أعلى من الثبات بسبب كون الاختبارات ذات عدد الفقرات القليل أكثر عرضة لأخطاء القياس من الاختبارات الطويلة. وبالإضافة إلى ذلك يمكن تقدير الثبات في النماذج المتكافئة من الاختبار عن طريق وضع درجات الاختبار على مرحلتين، المرحلة الأولى للأسئلة ذات الأرقام الزوجية والمرحلة الثانية للأسئلة ذات الأرقام الفردية. يتم بعد ذلك حساب الإرتباط بين

درجات نصفي الاختبار $r_{1/2 1/2}$ ثم تصحح على أساس طول الاختبار الكامل (وهو ضعف طول كل نصف) وفق معادلة سبيرمان – براون:

$$r_{1_{/2}1_{/2}} = 2r_{1_{/2}1_{/2}} /(1 + r_{1_{/2}1_{/2}}) \qquad (٤-٢)$$

ويعد حساب معامل ثبات التجزئة النصفية الذي يتم الحصول عليه وفق المعادلة أعلاه مقياساً للثبات الداخلي للاختبار أو لدرجة قياس فقرات الاختبار للمتغير ذاته.

طور كودر وريتشاردسون إجراءاً أكثر شمولية لحساب معاملات الاتساق الداخلي لفقرات الاختبار ثنائية الدرجة التي تحتمل الدرجات ١-٠ (١ في حال صحة الإجابة و٠ في حال خطئها)، ويطلق على هذا الإجراء اسم معادلات كودر- ريتشاردسون. وقد عمم كرونباخ إجراءات كودر- ريتشاردسون لتشمل الفقرات الاختبارية ذات الدرجات المتعددة (أكثر من ١ أو صفر) مما أنتج المعامل ألفا:

$$a = k(1 - \sum s_i^{\,2} / s_o^{\,2})/(k-1) \qquad (٥-٢)$$

حيث k هي عدد الفقرات وs_i^2 هي تباين الفقرة i وs_o^2 هي تباين مجموع درجات الاختبار؛ والمجموع ($\sum$) هو لجميع الفقرات k.

ويمكن حساب معاملات سبيرمان -براون وكودر- ريتشاردسون ومعامل كرونباخ ألفا بواسطة برامج الكمبيوتر E-1 و E-2 و E-3 الموجودة على القرص المرافق لهذا الكتاب. كما يساعد برنامج E-6 في حساب الخطأ المعياري للقياس و نسب ٩٠٪ و٩٥٪ و٩٨٪ و٩٩٪ لمجال ثقة الدرجات الفعلية الموافقة للدرجات الملاحظة المعطاة لاختبار ما. ويمكن الحصول على معادلة الخطأ المعياري للقياس (وهي إجراء إحصائي عام يطبق على درجات جميع المتحنين) بالطريقة

التالية: إن حل s_t^2 في المعادلة (٢-٥) يعطينا $s_t^2 = r_{11}s_o^2$. وبتعويض هذا الناتج بـ s_t^2 في المعادلة (٢-٢) وحل s_e^2 نحصل على: $s_e^2 = s_e^2 - r_{11}s_o^2$.

وأخيراً، يعطينا تبسيط طرفي المعادلة وأخذ الجذر التربيعي للطرفين الخطأ المعياري للقياس:

$$s_e = s_o\sqrt{1-r_{11'}} \qquad (٦-٢)$$

ويطبق هذا الإجراء الإحصائي بالطريقة التالية: بفرض أن ثبات الاختبار يساوي٩٠٪ وأن الإنحراف المعياري هو ١٠، فإن الخطأ المعياري للقياس هو:

$10\sqrt{1-.9} = 3.16$. ولحساب المجال الذي تقع ضمنه، وبمقدار ٩٥٪من الثقة، الدرجات الفعلية لفرد ما هو عضو في مجموعة درجاتها الملاحظة، نقوم بالعملية الحسابية التالية: $O \pm 1.96(s_e) = 43.81,56.19$.

إن مفهوم الثبات ليس محدوداً بالمتغيرات الفردية إذ يمكن تحديد مدى ثبات التباين بين درجات متغيرين أو ثبات مركّب يتألف من متغيرين أو اكثر. ويمكن حساب الثبات والخطأ المعياري لقياس درجات التباين عن طريق برنامج الكمبيوتر E-7. تساعد هذه الإحصائيات في تحديد ما إذا كانت درجات فردٍ ما في اختبارين مختلفين أو في اختبارين فرعيين من نفس الاختبار (كبطارية الاستعدادات المتعددة) تختلف عن بعضها بشكل كبير. وتؤكد النتائج على أن ثبات التباين بين درجات اختبارين هي أقل من ثبات كل اختبار على حدة، وعليه فإن الخطأ المعياري لقياس تباين الدرجات هو أعلى بكثير من الخطأ المعياري للقياس في كل من الاختبارين.

يمكن استخدام برنامج E-8 في تقديرثبات اختبار مؤلف من عدة أقسام لكل قسم منها ثباته وإنحرافه المعياري وثقله العددي. ويمكن للبرنامج حساب مجموعة الأثقال المثلى التي تعطي أقصى درجة من الثبات المركّب.

obeikandl.com

نظرية التعميم

يساعد التحليل العاملي على تحليل تباين الدرجات الفعلية في المعادلة (٢-٢) إلى تباينات عاملية عامة وتباينات عاملية خاصة، إلا أن هذه المعادلة لا تميز بين مصادر الخطأ المختلفة التي تساهم في تباين الخطأ (S_e^2). وقد أثبت خبراء القياس النفسي على مر السنين أن للاختبارات أنواع متعددة من الثبات تعتمد على مصادر الخطأ في القياس التي تُأخذ بعين الاعتبار عند حساب معاملات الثبات.

تؤثر العوامل التالية على تقدير الثبات: الفقرات المحددة التي يتكون منها الاختبار وتعليمات الاختبار والظروف المحيطة بالممتحنين ومواصفات عينة الاختبار. كما تعطينا الطرق المختلفة في حساب الثبات (طريقة الاختبار وإعادة الاختبار والنماذج المكافئة والاتساق الداخلي) نتائج مختلفة.

وقد اعتمدت الإتجاهات الحديثة لحساب الثبات على الإنطلاق من حقيقة أن عدداً من مصادر الخطأ المختلفة (المناسبات أو النماذج، أو المصححون، أو الإداريون) يمكن أن تؤثر على ثبات الاختبار، إذ يمكن للاختبار أن يكون ثابتاً في عدد من النواحي حسب تأثير هذه المصادر أو المظاهر. وتفيد نظرية التعميم، وهي نظرية جديدة لحساب صدق وثبات المقاييس الإدراكية، بأن الدرجات الملاحظة ما هي إلا تقدير لمعدل درجات عام (Universe Score). وإن مدى صحة درجات اختبار ما في تقدير قيمة عامة ما هي إلاَّ مقياس لثبات الاختبار. ويعتمد مدى إمكانية التعميم من درجات الاختبار الملاحظة إلى القيمة العامة على طبيعة الوسط المراد التعميم عليه أي على المظاهر التي تعرّف هذا الوسط.

يمكن تعميم درجات اختبار ما على عدد من الأوساط العامة يعرّف كل واحد منها تبعاً للمظاهر التي تدخل في تركيبه، وقد تختلف جذرياً من وسط لآخر كل

من العناصر التي تشخِّص وسطاً عاماً معيناً مثل ظروف إدارة الاختبار، وتركيبة الفقرات أو نماذج الاختبار أو ما شابه ذلك. وفيما لا تؤثر بعض هذه المظاهر على تعميم درجات الاختبار، يلعب البعض الآخر دوراً شديد الأهمية في عملية التعميم. كما يمكن الحصول على معاملات التعميم أوالثبات عند تعميم الدرجات على أوساط متعددة المظاهر.

وتتطلب عمليات الحساب الإحصائية للتعميم تطبيق تحليل لطرق تباين مصممة لتحديد دلالة عناصر معينة وتأثيرها على معاملات التعميم الناتجة. وبتأكيدها على أهمية الظروف المحيطة بالاختبار من حيث طريقة إجراء الاختبار والأهداف التي صمم من أجلها، ساهمت نظرية التعميم في إزاحة الهاجس التقليدي لمستخدمي الاختبار الذي كان الشغل الشاغل للعاملين في مجال الاختبارات والمتعلق بكون الاختبار اختباراً جيداً أو سيئاً بهاجسٍ جديد وهو: "ما هو الغرض الذي يعتبر هذا الاختبار جيد أو سيئ له؟".

قياس الاختبارات والمعايير

إن درجات الاختبارات وغيرها من المقاييس النفسية لا تعني شيئاً بمفردها، إذ يتوجب تفسير هذه الدرجات وفق المقاييس الذاتية الخاصة بالأداء الجيد أوالسيئ أو الطبيعي أو غير الطبيعي التي يمتلكها أفراد مؤهلون لإصدار مثل هذه الأحكام. ويتم التوصل إلى هذه المقاييس الذاتية عن طريق الخبرة في مادة الاختبار وعن طريق توقع ما يمكن اعتباره مستوى الأداء المقبول. فعلى سبيل المثال، يمتلك المدرسون ذوي الخبرة الخاصة بالمادة والطلاب، وعليه فقد أنشؤوا إطاراً داخلياً مرجعياً والذي بواسطته أصبح بإمكانهم مقارنة الإجابات المختلفة التي يعطيها الطلاب ودرجات الاختبار لمجموعة طلاب. إلا أن هؤلاء المدرسون الخبراء يختلفون مع الأسف فيما بينهم بخصوص تفسير الدرجات ووضع الدرجات النهائية ووضع درجات الاختبارات المقالية وغيرها من مقاييس الأداء ذات الطبيعة غير الموضوعية.

اتضحت أهمية إيجاد طرق موضوعية لتقييم درجات الاختبارات منذ بدايات تاريخ الاختبارات النفسية والتعليمية. ونظراً لعدم صلاحية الاعتماد على خبرة المصحح التي لا تقترن بالوثائق في تقييم الأداء الاختباري لمجموعة كبيرة من الأفراد، تم اعتماد عملية إجراء الاختبارات لعينة كبيرة من الأشخاص تمثل المجموعة السكانية المراد اختبارها. ويتم انتقاء أفراد العينة عشوائياً بشكل يتناسب مع التمثيل الديموغرافي للأفراد من الأعمار والصفوف الدراسية والأعراق المختلفة المكونة للمجتمع (وذلك لأن هذه المتغيرات تؤثر على درجات الاختبارات). ويساعد برنامج D-I الموجود على القرص المرافق لهذا الكتاب في اختيار عينات عشوائية طبقية أو متعددة الطبقات والذين سيستخدمون كأداة لقياس الاختبار، ويتم إجراء الاختبارات القياسية تحت ظروف موحدة ومُسيطر عليها وذلك لأن شخصية وأسلوب القائمين على الاختبارات والظروف المحيطة بالاختبار قد يكون لها آثار واضحة على نتائج الاختبارات، وتتم عملية التصحيح بعد الفراغ من إجراء الاختبار ثم يتم إعداد التوزيع التكراري للدرجات. بعد ذلك، يتم تحويل الدرجات إلى أنواع متعددة من المعايير تستخدم لاحقاً كمقياس تُقارن معه درجات الذين سيتم اختبارهم في المستقبل. ومن الضروري التأكد من أن عملية تقرير درجات ممتحن ما تستند إلى مقارنة درجاته مع جدول المعايير الناتج عن إجراء الاختبار لعينة من الأفراد تشابه خصائص أفرادها خصائص هذا الممتحن، وذلك بالنظر إلى الخصائص السكانية. فعلى سبيل المثال يقارن الأداء الاختباري لطالب من الصف الخامس مع المعايير التي تم الحصول عليها من عينة من طلاب الصف الخامس وليس الصف الرابع أو السابع، ويقارن الأداء الاختباري لطفل في العاشرة من عمره مع أداء أطفال في العاشرة أو مع أداء طلاب في الصف الخامس.

وتعد معايير العمر والسنة الدراسية (الصف الدراسي) والمعايير المئينية ومعايير الدرجات القياسية من أهم المعايير التي تحتويها دلائل المستخدم

للاختبارات المختلفةً. ويتم حساب معيار العمر عن طريق حساب وسيط درجات لأطفال من العمر الزمني المعطى. فإذا كان وسيط الدرجات لأطفال عمر كل منهم ١٠ سنوات و٥ أشهر هو ٦٠، فإن معيار العمر لفرد يحصل على معدل الدرجات ٦٠ في هذا الاختبار هو ١٠ سنوات و٥ أشهر. وتحتوي اختبارات الذكاء القديمة على نوع آخر من المعايير العمرية هو معيار العمر العقلي.

وكما هو الحال في المعيار العمري يتم حساب معيار الصف الدراسي عن طريق حساب وسيط درجات طلاب سنة دراسية معينة. فإذا كان الوسيط لدرجات طلاب في الشهر السادس من الصف الرابع يساوي ٥٥، فإن معيار الصف لفرد يحصل على معدل الدرجات ٥٥ في هذا الاختبار هو ٤,٦، ومن الجدير بالذكر أن التعبير عن معيار الصف الدراسي هو بالسنوات الدراسية وأجزاء من العشرة، مما يعني أفتراض عدم تطور الأطفال في أشهر العطلة الصيفية.

وعلى الرغم من شيوع استخدام طريقتي معيار العمر ومعيار الصف في تفسير درجات الاختبارات القياسية، يبقى هنالك خطر التعامل مع هذا المستوى الرتبي للقياس على أنه مستوى فئوي. إذ أن القول بأن الفرق بين مقياس صف ٤,٦ و مقياس آخر يساوي ٦,٦ هو مساوٍ للفرق بين المقياسين ٦,٦ و٨,٦ هو استنتاج خاطئ لسبب بسيط وهو أن القدرات الأكاديمية تتطور بشكل أكبر وأسرع في السنوات الأولى من عمر الطالب وفي سنوات الدراسة الأولى بالمقارنة مع السنوات اللاحقة. وعليه لا تتساوى قيم وحدات معيار العمر ومعيار الفصل الدراسي في مختلف سنوات العمر والمراحل الدراسية.

تتصف المعايير المئينية (وهي نوع آخر من الدرجات النهائية للاختبارات) بعدم وجود علاقة خطية (منتظمة) بين الدرجات فيها والزمن. والمعيار المئيني أو معيار الرتب المئيني لاختبار ما هو النسبة المئوية لعدد أفراد عينة القياس الذين حصلوا على معدل الدرجات هذا أو على معدل أدنى منه. وعليه فإذا حصل ٦٠٪ من أفراد عينة القياس على معدل درجات ٧٥ أو أقل فإن الرتبة المئينية لفرد

معدل درجاته ٧٥ هو ٦٠، ويتوجب عند تحويل معدلات الدرجات إلى رتب مئينية مراعاة تناسب عينة القياس التي تم تحديد المعايير على أساسها مع الممتحن المراد تفسير درجاته.

ولعل أحد أسباب شيوع استخدام معايير العمر والصف والمعايير المئينية كون هذه المعايير سهلة الفهم لغير المختصين في مجال القياس النفسي. أما قواعد الدرجات المعيارية فهي مقاييس أكثر دقة من وجهة النظر السايكومترية إلا أنها أكثر تعقيداً. وتختلف قواعد الدرجات المعيارية عن قواعد العمر والصف والمعايير المئينية بكونها مقاييس فئوية، إلا أنها لم ترقى إلى المرتبة النِسَبية لقياس. تستند جميع الدرجات المعيارية إلى معدل درجات z على الشكل التالي:

$$z = (X - \bar{X})/s \qquad (٧-٢)$$

حيث X هي معدل الدرجات الخام و$\bar{X}$ هي الوسط الحسابي و s هي الإنحراف المعياري. وبما أن معدلات الدرجات z يمكن أن تكون أعداداً عشرية إما موجبة أو سالبة يتم ضربها بعدد صحيح ثم يضاف عدد صحيح آخر إلى الناتج. فعلى سبيل المثال، يتم حساب معدلات الدرجات T بضرب معدلات الدرجات z ب ١٠ ثم إضافة ٥٠ للناتج. ويتم حساب انحراف معدلات ذكاء ويكسلر بضرب معدلات الدرجات z ب ١٥ ثم إضافة ١٠٠، وتحسب معدلات درجات مجلس امتحانات القبول الجامعية (CEEB) بضرب معدلات الدرجات z ب ١٠٠ ثم إضافة ٥٠٠، وتحسب درجات الرتب من تسعة (Stanine) بضرب معدلات الدرجات z ب ٢ ثم إضافة ٥. وهنالك عدة أنواع من الدرجات المعيارية مثل NCE (الدرجات المكافئة للمنحنى الطبيعي) والدلتا. تفترض العديد من هذه الدرجات المشتقة أن التوزيع التكراري للدرجات هو توزيع طبيعي، ولذلك فإنه من المناسب استخدام الدرجات القياسية المطبّعة Normalized للتعبير عن هذه الدرجات.

تقيس البرامج ٢ و٣ و٤ من الفئة D الموجودة على القرص المرفق بهذا الكتاب المعايير المئينية ومعايير الدرجات القياسية ومعايير الدرجات القياسية المطبّعة. ويعد برنامجا D-2 وD-4 برنامجين متعددي المزايا إذ يمكن استخدامهما في اتجاهين، كاستخدام برنامج D-2 لتحديد الرتبة المئينية لمعدل درجات ما أو معدل الدرجات الموافق لرتبة مئينية ما. ويحدد البرنامج D-4 إما الاحتمالية الطبيعية لقيمة z أو القيم z و T و CEEB وستانين ومعدلات درجات NCEالموافقة لاحتمالية طبيعية تراكمية.

تحليل فقرات الاختبار

نظراً لأن درجات الاختبارات التحصيلية واختبارات الاستعداد المختلفة غالباً ما تعتمد على حساب مجموع أسئلة أو فقرات الاختبار التي أجاب عليها الممتحن بشكل صحيح، فإن صدق وثبات مجموع درجات الاختبارات تعتمد على دلالات الفقرات أو العناصر المكونة لهذه الاختبارات. وفي حالة الاختبارات التحصيلية حيث يتم التركيز على صدق المحتوى وعلى الإتساق الداخلي للاختبار، فإن تقييم دلالات الفقرات الاختبارية يتم عن طريق تحديد الإرتباط بين أداء الفرد في هذه الفقرة وأداؤه في الاختبار بشكل عام. ويمكن الحصول على معلومات تحليل فقرات الاختبار عن طريق استخدام مؤشر صعوبة الفقرات ومؤشر تمييز الفقرات. ويتم حساب مؤشر صعوبة الفقرات (p)، والذي يعبرعن نسبة الأفراد ضمن عينة الاختبار الذين أعطوا إجابة صحيحة على سؤال ما، عن طريق المعادلة $p = n/N$ حيث n هي عدد الأفراد الذين أعطوا إجابة صحيحة وN هي عدد الأفراد الذين تم اختبارهم. ويمكن حساب مؤشر الصعوبة بطريقة أخرى وهي تقسيم مجموعة المتحنين إلى مجموعتين فرعيتين هما مجموعة تضم أفراداً معدلات درجاتهم تضع في أفضل ٢٧٪ من المعدلات (U) ومجموعة تضم أفراد معدلات درجاتهم تقع في أسوأ ٢٧٪ من المعدلات (L)

obeikandl.com

وذلك تبعاً للتوزيع التكراري التالي لمجموع معدلات الدرجات. ويتم حساب p كما يلي:

$$p = (U_c + L_c)/(U + L) \qquad (٨-٢)$$

حيث U_c هو عدد الأفراد في المجموعة الأولى الذين أعطوا إجابة صحيحة على السؤال وL_c هو عدد الأفراد من المجموعة الثانية الذين أجابوا على السؤال بشكل صحيح. وبالإضافة إلى مؤشر صعوبة فقرات الاختبار، يتم حساب مؤشر التمييز للفقرات الاختبارية (D) وهو إجراء للتمييز بين المجموعتين الأولى والثانية من حيث إجابات أفرادهم على فقرة امتحانية ما - كما يلي:

$$p = (U_p + L_p)/(U + L) \qquad (٩-٢)$$

وبشكل عام، تكون مؤشرات الفقرات الاختبارية الفعالة كالتالي: يتراوح مؤشر p بين ٢٠ و٨٠ وتكون مؤشرات D٣٠ وما فوق. يساعد برنامج C-1 الموجود على القرص المرفق بهذا الكتاب في حساب مؤشرات الصعوبة والتمييز لأعداد كبيرة من الفقرات الامتحانية.

تقوم الاختبارات ذات المرجعية المحكّية Criterion - referenced tests، وعلى عكس الاختبارات ذات المرجعية المعيارية Norm - referenced tests، على أساس مقارنة الإجابات مع مجموعة من الأهداف التعليمية المراد تحقيقها وليس على أساس مقارنة الإجابات بعضها مع بعضها. وبينما يتم حساب مؤشر الصعوبة لفقرات الاختبارات ذات المرجعية المحكّية بنفس الطريقة المستخدمة مع الأنواع الأخرى من الاختبارات، يحسب مؤشر التمييز وفق المعادلة التالية:

$$D = U_c/U - L_c/L \qquad (١٠-٢)$$

حيث U هي عدد الممتحنين الذين حصلوا على معدلات درجات أعلى من مستوى المحك، وL هي عدد الممتحنين الذين حصلوا على معدلات درجات أدنى من

مستوى المحك، وهو عدد الأفراد في المجموعة U الذين نجحوا في الفقرة الاختبارية، وL_c هو عدد الأفراد في المجموعة L الذين نجحوا في الفقرة الاختبارية. ويساعد برنامج C-2 الموجود على القرص المرفق بهذا الكتاب في حساب مؤشرات الصعوبة والتمييز لفقرات للاختبارات ذات المرجعية المحكية.

ويستخدم المعامل الثنائي الأصيل Point - biserial coefficient (r_{pb}) بدلاً عن مؤشر تمييز الفقرات عندما يكون المحك الذي يقيِّم فعالية الفقرة الاختبارية محكاً خارجياً كمقاييس الأداء الوظيفي أو الأكاديمي. ويتم حساب هذا المؤشركالتالي:

$$r_{pb} = (\overline{Y}_c - \overline{Y})\sqrt{Nn_c/[(N-n_c)(N-1)]}/s. \qquad (٢-١١)$$

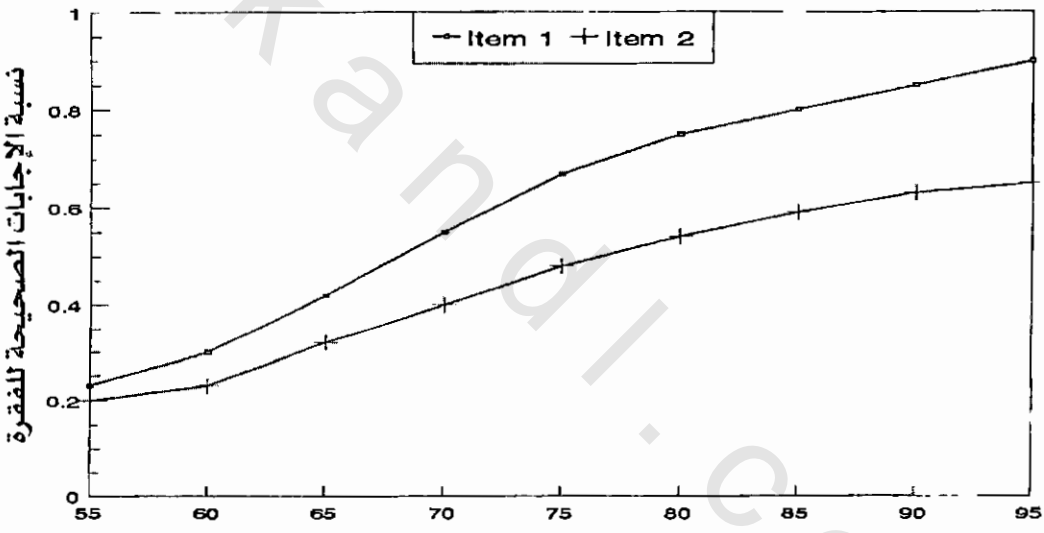
حيث N تساوي مجمل عدد المتحنين وn_c هي عدد المتحنين الذين أجابوا على السؤال بشكل صحيح و$\overline{Y}_c$ هي متوسط الدرجات على مقياس المحك للممتحنين الذين أجابوا على الفقرة الامتحانية بشكل صحيح و$\overline{Y}$ هي متوسط درجات المحك لجميع المتحنين و s هي الإنحراف المعياري لدرجات المحك لكافة المتحنين. وعلى الرغم من عدم وجود قيمة محددة لـ (r_{pb}) يتم على أساسها استبعاد الفقرات الاختبارية التي لا تصل إلى هذه القيمة، يمكن القول أن عدد نقاط متدنٍ جداً لـ (r_{pb}) يمكن أن يؤدي إلى مراجعة الفقرة الاختبارية أو الاستغناء عنها. ويساعد برنامج C-2 الموجود على القرص المرفق بهذا الكتاب في حساب المعامل الثنائي الأصيل.

يمكن الحصول على معلومات أكثر فائدة وتفصيلاً عن تحليل الفقرة الاختبارية وذلك عن طريق رسم منحنى لخصائص الفقرة (ICC) يعبر عن نسبة الإجابات الصحيحة إلى الفقرة معبراً عنها بمجموع درجات الاختبار (الشكل ٢-٢). نلاحظ في الشكل ٢-٢ توافُق قيمة ٠,٥ على المحور الرأسي مع مجموع

درجات مقداره ٦٨ لفقرة الاختبار ١ و٧٧ لفقرة الاختبار ٢، وهذا يعني أن الفقرة ٢ أكثر صعوبة من الفقرة ١، إلا أن منحنى الفقرة ١ يتصف بانحدار أشد من الفقرة ٢ مما يشير إلى أن الفقرة ١ أقدر على التمييز بين الممتحنين. ويساعد برنامج C-4 الموجود على القرص المرفق بهذا الكتاب في حساب منحنيات ICC بناءاً على المعلومات الأولية المأخوذة من نموذج الإجابة على فقرات الاختبار ذات المتغيرات الأحادية، أو الثنائية أو الثلاثية (راجع الفصل التالي).

الشكل ٢-٢

منحنيان لخصائص الفقرات لهما مؤشران مختلفان للصعوبة والتمييز



وبالإضافة إلى حساب p وإما قيم D أو (r_{pb}) لكل فقرة من فقرات الاختبار، يجب النظر في التوزيع التكراري للمموهات itemdisracters (الخيارات غير الصحيحة) في أسئلة الاختيار من المتعدد ضمن المجموعتين U و L. تعد المموهات في أسئلة الاختيار المتعدد عنصراً مهماً من عناصر فعالية الفقرة الاختبارية، وقد

تؤثر هذه المموهات تأثيراً هداماً على الفقرة الاختبارية إذا أدت إلى الغموض أو كانت غير فعالة. إذ يمكن أن تقود الأخطاء أو التناقضات أو غيرها في محتوى المموهات إلى اختيار نسب كبيرة (وغير متوقعة) من الممتحنين لإجابة مموهة ما بدلاً عن الإجابات الأخرى مما يؤدي إلى سوء في توزيع اختيار المموهات وإلى إعادة النظر في الفقرة الامتحانية المعنية أو إلى حذف هذه الفقرة بشكل كامل. ويساعد برنامج C-3 الموجود على القرص المرفق بهذا الكتاب في إجراء اختبار إحصائي حول توزيع الإجابات على مموهات أسئلة الاختيار المتعدد.

وأخيراً؛ إن كون مؤشري الصعوبة والتمييز لاختبار ما مقبولين، لا يعني أن هذا الاختبار يعمل بالشكل نفسه مع كافة المجموعات البشرية (من عرق أو جنس أو عمر معين) التي يتم اختبارها. ولهذا السبب يتطلب وضع الاختبارات التحصيلية القياسية وغيرها من اختبارات القدرة الحصول على معلومات تحيز Biasedness فقرات الاختبار أو ما يسمى بفعل الفقرة التباييني (DIF) Differential item functioning. يحتوي برنامج C-7 على مقاربات للحصول على معلومات فعل الفقرة الاختبارية التباييني عن طريق: الرسم البياني لدلتا الفقرة وطريقة كاي تربيع التي استخدمها شيونمان وكاميللي ومانتل وهاينتسل. وتعد طريقة الرسم البياني لدلتا الفقرة الاختبارية أقدم هذه الطرق إلا أن طريقة كاي تربيع التي استخدمها مانتل وهاينتسل كانت الأكثر استخداماً خلال السنوات الأخيرة.

درج صانعوا الاختبارات على الاحتفاظ بنتائج تحليل الفقرات الاختبارية على بطاقات مفهرسة ترفق بالفقرة الاختبارية وتخزن بغرض استخدامها لاحقاً. أما في الوقت الحالي ومع وجود أجهزة كمبيوتر ذات سرعة المعالجة العالية فيتم تخزين الفقرات الاختبارية ضمن بنك تصنف فيه تبعاً للموضوع أو درجة الصعوبة أو مستوى التمييز أو بحسب العمليات الإدراكية المطلوبة للإجابة عليها. ولا يقتصر استخدام "بنوك" فقرات الاختبار هذه على واضعي الاختبارات

المحترفين إذ غالباً ما ترفق هذه "البنوك" بالعديد من الكتب الدراسية بغرض استخدامها كأسئلة تدريبية أو كمصدر لإعداد الاختبارات المدرسية. ويمكن لمدرسي المقررات أن يستخدموا حواسيبهم المكتبية لاختيار فقرات اختبارية متعلقة بعدد من المضامين ثم تجمع هذه الفقرات معاً لتشكل اختبارات وحدات دراسية معينة أو امتحانات عامة. وتفيد برامج الحاسب في تسهيل عملية إختيار فقرات اختبارية مرتبطة بمادة دراسية معينة وتمتلك عدداً من المواصفات الإحصائية المحددة.

نظرية الاستجابة لفقرة الاختبار IRT Item response theory

تختلف نظرية الإستجابة للفقرة عن غيرها من نظريات وطرق الاختبار التقليدية التي لا تعير اهتماماً كبيراً للإجابات على الفقرات الاختبارية، إذ أن التركيز الأساسي لنظرية IRT هو على منهجية الإجابة على فقرات الاختبار. وتستند هذه النظرية التي تعرف باسم IRT إلى العلاقة الوظيفية النظرية بين سلسلة للقدرات الكامنة عند الفرد وبين الإجابات على فقرات الاختبار. ويتم تمثيل نموذج IRT عن طريق تابع لوجستي ذو مؤشر واحد أو مؤشرين أو ثلاثة مؤشرات. ويعبر عن النموذج ذو المؤشرات الثلاثة كالتالي:

$$P(\theta) = c + (1-c)/[1+e^{-a(\theta-b)}] \qquad (١٢-٢)$$

حيث e هي قاعدة اللوغاريتمات الطبيعية (٢,٧١٨٢٨٢) وb هي مؤشر صعوبة الفقرة الاختبارية و c هي مؤشر التخمين الكاذب وθ هي مستوى قدرة المجيب على السؤال (الممتحن) بالنسبة لمقياس الدرجات المعيارية وP(θ) هي احتمالية إجابة فرد ما مستوى قدرته هو θ بشكل صحيح على السؤال. وبفرض أن c=0 نحصل على النموذج ثنائي المؤشرات:

$$P(\theta) = 1/[1+e^{-a(\theta-b)}] \qquad (١٣-٢)$$

ويؤدي الافتراض بأن جميع الفقرات الاختبارية لها درجة التمييز ذاتها إلى النموذج أحادي المؤشر أو ما يسمى نموذج راش:

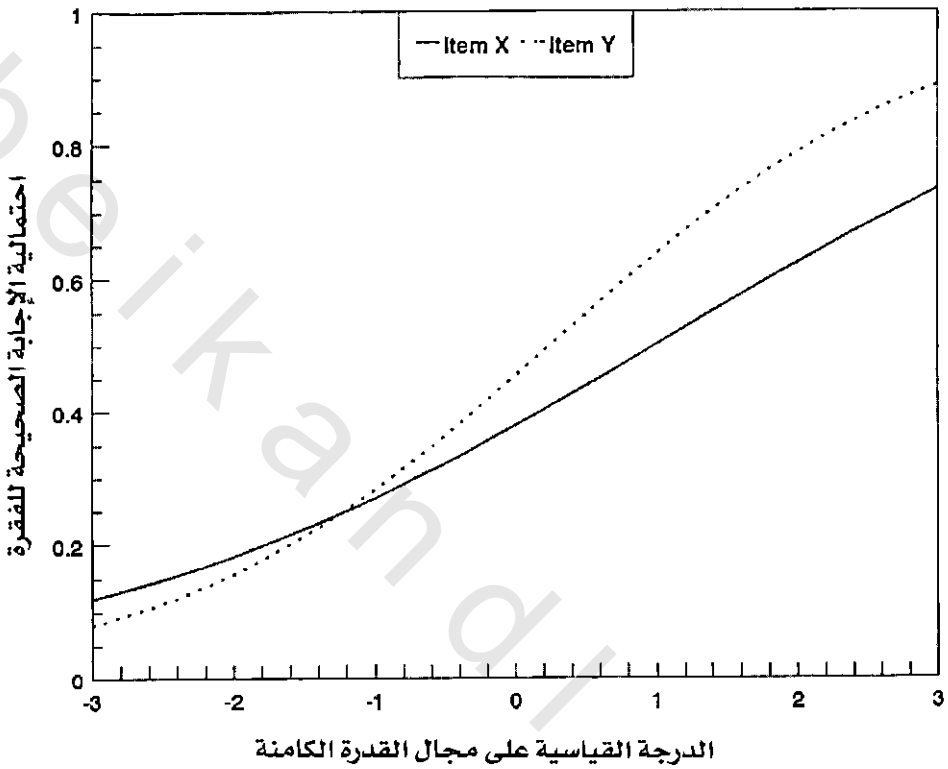
$$P(\theta) = 1/[1 + e^{-(\theta - b)}] \qquad (١٤-٢)$$

وعلى الرغم من أن نموذج راش كان أساساً للعديد من الأبحاث في مجال القياس النفسي، يبقى النموذج ثنائي المؤشر هو النموذج الأكثر شيوعاً.

وكما يوضح الشكل ٢-٣، يختلف شكل المنحنى الذي ترسمه الإجابات على فقرة اختبارية ما باختلاف قيم المؤشرات a وb. وتم رسم المنحنيين في الشكل ٢-٣ بالاعتماد على التابع ذو المؤشرين الموجود في المعادلة ٢-١٣ وكان مؤشرالصعوبة (b) هو ١,٠٠ ومؤشرالتمييز (a) هو ٠,٥، على المنحنى X، على حين أن قيم b و a على المنحنى Y كانت ٠,٧٥، و٠,٢٥، على التوالي. ومن الملاحظ أن bهي قيمة θ (النقطة على المحور الأفقي) الموافقة لـ ٠,٥ = p(θ)، وأن a هي ميل المنحنى في النقطة لـ ٥ = p(θ)، وأما في النموذج ثلاثي المؤشرات فإن b هي قيمة p(θ) الموافقة لـ ٥ (c + 1)، حيث c أو مؤشرالتخمين الزائف هي النقطة التي يتقاطع عندها منحنى الإجابات مع المحور الرأسي. ولعله من المفيد رسم عدد من منحنيات الإجابات الإحادية والثنائية والثلاثية المؤشرات باستخدام قيم المؤشرات التي يوفرها برنامج C-4 الموجود على القرص المرافق لهذا الكتاب.

الشكل ٢-٣

منحنيان لإجابات فقرات اختبار ذات مؤشرات مختلفة. (راجع النص لمزيد من المعلومات)



وقد تم التعبير عن مجال القدرة الكامنة في هذه الأشكال التوضيحية عن طريق وحدات الدرجات المعيارية (z)، إلا أن معظم التطبيقات التعليمية تحول الدرجات (z) إلى مقياس وسطه الحسابي ٣٠٠ وانحرافه المعياري ٥٠. وتساعد الرسوم البيانية (كالشكل ٢-٣) على رسم الخط البياني لنسبة الإجابات الصحيحة الموافقة لمستويات قدرة معينة كما يمكننا أن نرسم كذلك نسبة عدد الأفراد المنتمين لمستوى قدرة معين والذين إختاروا مموه معين مما يساعد في تقييم فعالية المموهات المرتبطة بفقرة اختبارية ما.

obeikandl.com

وفي الواقع العملي، تبقى مؤشرات الفقرة الاختبارية ودرجات مجال القدرة الكامنة (θs) للمتقدم للاختبار، مجهولة وتبقى المشكلة هي تحديد منحنى الإجابة الذي يتناسب بالشكل الأمثل مع الإجابات على الفقرات الاختبارية المختلفة. ويتطلب حل هذه المشكلة إجراءاً تكرارياً للاحتمالية القصوى يفترض عدداً من القيم الابتدائية لمؤشرات الفقرات الاختبارية ثم يحسب P(θ)s لمختلف قيم θ مع مقارنة الإجابات المتوقعة مع الإجابات الفعلية للفقرة، ويستمر إتباع الإجراء ذاته حتى يتم التوصل إلى الحل الأنسب. وتتطلب عملية تقديرمؤشرات الفقرات الاختبارية عدداً كبيراً من الإجابات لعدد كبير من الممتحنين يمثلون الشريحة السكانية التي ينتمي إليها كل من يُحتمل اختبارهم في المستقبل، ويتراوح هذا العدد بين ٥٠٠ للنموذج أحادي المؤشر (أو نموذج راش) و٢٠٠٠ للنموذج ثلاثي المؤشر.

ومن أهم خصائص مؤشرات الفقرات الاختبارية التي تم حسابها كون هذه المؤشرات مستقلة نوعاً ما عن مستوى القدرة المرتبط بالعينة التي تم على أساسها تحديد المؤشرات، إذ، وعلى عكس طرق الاختبار التقليدية التي تحصر درجة صعوبة الاختبار وتمييزه ضمن حدود العينة التي تم اختبارها، تعد هذه المؤشرات في نظرية الإجابة على فقرات الاختبار على المستوى النظري على الأقل مستقلة (أو لا علاقة لها) بالعينة موضع الاختبار وهذا ما يسمى بثبات (أو عدم تباين) العينة. ويساعد هذا الثبات لمؤشرات فقرات الاختبار على تحديد معدلات الدرجات المكافئة أو المعادلة في اختبارات مختلفة. وتتألف عملية المعادلة بين الاختبارات لتحويل درجات اختبار ما (أو صورة مختلفة من الاختبارذاته) إلى عدد من الوحدات مساوٍ لعدد وحدات اختبار آخر من الخطوات التالية: (١) إيجاد الرتب المئينية الموافقة لدرجات كلا الاختبارين، ومن ثم (٢) وضع جدول لدرجات مكافئة تتوافق مع الرتب المئينة في كلا الاختبارين.

وتتطلب عملية المعادلة في إطار (IRT) تحديد معادلة خطية تحول مؤشرات الصعوبة والتمييز لنموذج اختباري ما إلى نموذج اختباري آخر. وتسمى عملية إيجاد قيم ثابتة لمعادلة التحويل الخطية تضمن كون المؤشرات في كلا الاختبارين على نفس المقياس بعملية الربط. وتتطلب عملية الربط هذه أن يشترك الاختباران ببعض الفقرات الاختبارية (الفقرات الراسية Anchor items) أو أن يتم اختبار مجموعة فرعية من الأفراد بالاختبارين موضع الدراسة أو باختبار ثالث يقيس الخاصية ذاتها. كما ينطوي إجراء IRT لمعادلة الاختبارات، وهو إجراء اقتصادي، على اختبار مجموعات عشوائية من الأشخاص باستخدام مجموعات فرعية من الفقرات (اختيار الفقرات).

وتسمح عملية تحليل فقرات الاختبار (IRT) بالإضافة إلى تقدير مؤشرات الفقرات الاختبارية بتقدير درجات المتحنين بالنسبة إلى مجال القدرة الكامنة. وفي الواقع يعتبر تحديد موقع المتحن بالنسبة إلى مجال القدرة الكامنة الهدف الرئيس لنظرية (IRT). وكما كان الحال في تقدير مؤشرات الفقرة الاختبارية بناءاً على مستويات القدرة، فإن تحديد معدلات درجات القدرةهي عملية تكرارية تبدأ بتعويض قيم تجريبية لقدرة المتحن ومؤشرات الفقرة التقديرية في المعادلة اللوجستية المناسبة. ثم تقارن قيم P(θ)s التقديرية مع قيم P(θ)s الفعلية وتستمر العملية حتى الوصول إلى المعادلة الأكثر تناسباً مع القيم الناتجة. كما يمكن حساب الأخطاء المعيارية للقيم المقدرة θs (وهي مقياس لتغير قيم θs التقديرية بالنسبة إلى قيم θs الفعلية المجهولة).

وتختلف IRT عن نظريات القياس التقليدية (التي تطبق الخطأ المعياري للقياس على جميع درجات مكونات عينة ما) بكون الخطأ المعياري للقياس يختلف من معدل درجات إلى آخر. ويساوي الخطأ المعياري لقياس مستوى قدرة معينة في نظرية IRT مقلوب كمية المعلومات التي تعبر عنهاالدرجات ضمن ذلك

المستوى. وتحدد كمية المعلومات الناتجة عن فقرة اختبارية ما عن طريق دلالة معلومات هذه الفقرة بينما يعبر مجموع قيم معلومات الفقرات الاختبارية المختلفة ضمن مستوى قدرة معين عن معلومات الاختبار بكامله ضمن هذا المستوى.

ويعد ثبات قدرة الممتحنين بالنسبة إلى الفقرات الاختبارية التي تم استخدامها لتحديد هذه القدرة واحداً من الصفات الهامة لنظرية IRT والتي تستند إلى عملية تقدير θs وتعني هذه الميزة أنه يمكن استخدام أي اختبار على أي درجة من الصعوبة لتحديد موقع الممتحن ضمن مجال القدرة الكامنة. إلا أنه وللحصول على أكثر النتائج دقة، ينصح باستخدام فقرات اختبارية (أو اختبار كامل) تتوافق درجة صعوبتها مع مستوى قدرة الممتحن.

هذا ويمكن استخدام IRT لتحقيق الأهداف التالية:

- وضع وتصميم الاختبارات.
- معايرة درجات الاختبارات للتوصل إلى مرجعية يمكن تفسير الدرجات هذه على أساسها.
- قياس الاختبارات.
- تحديد فعل الفقرات التباين (DIF).
- الاختبار التكيفي.

تساعد نظرية IRT فيما يخص وضع وتصميم الاختبارات في تحديد مواقع تحليل الفقرات الاختبارية التي تعطينا قياسات أكثر دقة مما يقودنا إلى التركيز على هذه المواقع وعدم هدر الفقرات الاختبارية على مواقع أقل أهمية. وعليه يمكننا وباستخدام IRT تصميم اختبارات انتقاء شديدة الدقة واختبارات انتقائية، واختبارات ذات مرجعية محكّية واختبارات تقليدية ذات طيف واسع الامتداد

ضمن مجال القدرة. أما مقاربة IRT في مجال تحديد عمل الفقرات التباييني (DIF)، فتتم برسم المخططات البيانية لمنحنيات الإجابة لكل مجموعة سكانية على حدة (البيض والسود، الرجال والنساء... إلخ). وتعبر الأشكال المختلفة لمنحنيات الإجابة لكل من المجموعتين عن عمل فقرات الاختبار التباييني (DIF).

اثبتت الاختبارات التكيفية (أو الاختبارات التي تتناسب مع خصوصية شخص ما) أنها نوع اقتصادي ودقيق من الاختبارات المحوسبة. وتعتمد هذه الاختبارات على قدرة الممتحن التي يتم تقديرها بناءً على إجابته على الأسئلة السابقة، وبما أن أدق الاختبارات هي تلك التي تتناسب صعوبة فقراتها مع مستوى قدرة الممتحن فإن عملية إعادة تقدير قدرة الممتحن التي تتكرر خلال الاختبار تسمح باختيار فقرات الاختبار الأكثر تناسباً مع مستوى قدرة الممتحَن الفعلية. وعلى الرغم من كون الاختبارات التكيفية أقصر من الاختبارات التقليدية إلا أن الاختبارات التي تتناسب مستوياتها مع مستوى قدرة الممتحَنين يمكن أن تكون أكثر صدقاً من الاختبارات الأكثر طولاً والمصممة لاختبار القدرة ذاتها. وقد تم اعتماد الاختبار التكيفي المحوسب (CAT) في عدد من الاختبارات القياسية مثل بطارية اختبارات القابلية المهنية للقوات المسلحة (ASVAB) واختبار القدرة الدراسية (SAT) واختبارات سجل الخريجين (GRE).

وتعد الحسابات المستخدمة في تحليل فقرات الاختبار (IRT) حسابات معقدة لا يمكن أن تتم دون الإستعانة بعدد من برامج الكمبيوتر مثلLOGIST وBILOG (ميسليفي وستوكينغ، ١٩٨٩) وASCAL (فايل، ١٩٨٥) وBIGSTEPS (رايت ولينايكر، ١٩٩١). ولعل واحدة من المشكلات التي تعاني منها جميع نماذج الـIRT هي الافتراض بأن قدرةً كامنةً واحدة فقط هي المسؤولة عن الأداء الاختباري. وقد أحرزت بعض النماذج المتعددة الأبعاد نوعاً من التقدم في هذا السياق مثل نماذج (أمبرتسون، ١٩٩١ وريكيس، آكرمان وكارلسون، ١٩٨٨).

وتعتمد معظم نماذج الـ IRT على نظام توزيع درجات الفقرات (١-٠) وقد جرى العمل على نماذج أكثر تعقيداً تعتمد على نظام توزيع درجات متعددة (كقوائم الرصد) (مثال: ثيسين وشتاينبيرغ، ١٩٨٦؛ رايت وماسترز، ١٩٨٢).

الخلاصة

يتم القياس على عدة مستويات هي: المستوى الإسمي والمستوى الرُتبي والفئوي والنسبي. وتقع معظم طرق القياس النفسية ضمن المستوى الفئوي أو ما دونه. يتم تحليل مقاييس الأحداث الإدراكية والعاطفية والنفسية - الحركية بواسطة طرق إحصائية. وتتوفر إجراءات إحصائية أحادية المتغير (univariate) أو متعددة المتغيرات (multivariate) تصف وتحلل العلاقات والاختلافات بين المتغيرات النفسية. وتستخدم طرق الإرتباط كمعامل الإرتباط المولد للعزوم والتحليل العاملي والتحليل التمييزي بشكل كبير في عملية تفسير درجات الاختبارات.

وتركز نظرية الاختبار التقليدية على الصدق والثبات وقياس الاختبارات وتحليل فقرات الاختبار. ويعرّف الثبات على أنه قدرة الاختبار على قياس الشيء ذاته بالشكل ذاته في كل مرة يجرى فيها هذا الاختبار، بينما يشير الصدق إلى قدرة الاختبار على قياس الشيء الذي أُعلن أن هذا الاختبار يريد أن يقيسه، ويحدد صدق اختبار ما بواسطة الطرق التالية: صدق المحتوى والصدق المرتبط بالمحك (الصدق التلازمي والصدق التنبؤي) وصدق البناء. وتعد الإجراءات الإحصائية التي تقيِّم الصدق التنبؤي أو قدرة الاختبار على التنبؤ بالسلوك البشري إجراءات على درجة عالية من التطور.

ترى نظرية الثبات التقليدية أن المعدل الملاحظ لدرجات فرد ما في اختبار ما على أنه المجموع الفعلي لدرجات هذا الفرد مضافاً إليه خطأ معين في

القياس، وقد استندت الطرق التقليدية لتقييم الثبات (وهي طرق: إعادة الاختبار والصور المكافئة والإتساق الداخلي) إلى هذه المعادلة. تتطلب عملية حساب ثبات الإتساق الداخلي (وهي عملية بسيطة التكاليف) حساب معاملات التجزئة النصفية وكودر ريتشاردسون وكرونباخ. ويرتبط مفهوم الخطأ المعياري للقياس بالموثوقية إذ يُستخدم هذا المفهوم لحساب مجال القيمة العددية التي يمكن لنا القول ونحن على درجة محددة من الثقة إن معدل درجات الفرد الحقيقي يقع ضمنها. ويختلف ثبات الاختبار شأنه كشأن الصدق تبعاً للمجموعة قيد الاختبار والظروف المحيطة، وتؤكد نظرية التعميم على وجود عدة أنواع للثبات المرتبط بالاختبارات.

وقد قاد التركيز التاريخي على عمل الاختبارات في قياس الفروقات بين الأفراد إلى الاهتمام بقياس الاختبارات وتحديد المعايير التي يمكن تفسير نتائج الاختبارات على أساسها. ومن أهم هذه المعايير: معيارالعمر والصف الدراسي والمعايير المئينية ومعايير الدرجات المعيارية. وتعد المعايير الثلاثة الأولى مقاييساً على المستوى الرُتبي وهي المقاييس المفضلة للمدرسين وغيرهم من الذين لا يتقنون فنون القياس النفسي الأكثر تعقيداً. أما محترفو القياس النفسي فيفضلون معيار الدرجات المعيارية الذي يقيس على المستوى الفئوي. وتعتمد درجة الفائدة التي يمكن الحصول عليها من مجموعة من معايير الاختبارات على ما إذا كانت هذه المعايير قد تم الحصول عليها من عينة (ويفضل أن تكون عينة عشوائية متعددة الطبقات) تمثل المجموعات البشرية التي يستهدفها الاختبار.

تعتبر عملية تحليل عمل الفقرات الاختبارية خطوةً هامة في مجال تطوير الاختبارات، ويعتمد هذا التحليل على مقارنة الإجابات على الفقرات الاختبارية المختلفة إما مع مَحكَّات داخلية (كدرجات الاختبار الكلية) أو محكَّات خارجية (كالتحصيل الأكاديمي والأداء الوظيفي). ويحسب في حالة اختبارات الخيارات

المتعددة، بالإضافة إلى مؤشري الصعوبة والتمييز، توزيع إجابات المموهات وذلك لتحديد ما إذا كانت هذه المموهات تعمل بالشكل المطلوب. ويتم، بناءاً على عملية تحليل فقرات الاختبارات حذف أو مراجعة الفقرات غير الفعّالة، كما يتم في حالة الاختبارات القياسية تحديد ما إذا كانت هذه الاختبارات تعمل بشكل مختلف عند اختلاف المجموعة السكانية موضع الاختبار. وقد تم بحث إجراءات عمل الفقرات الاختبارية التباييني (DIF) في هذا الفصل من الكتاب.

وتهدف نظرية IRT على عكس نظريات الاختبار التقليدية التي تركز على مجموع درجات الاختبار إلى تحليل الإجابات على الفقرات الاختبارية المحددة وربطها بموقع الممتحن بالنسبة إلى مجال القدرة الكامنة. وبناءاً على النموذج الرياضي (المنطقي أو العادي) الذي يربط درجات القدرة مع احتمال الإجابة على فقرة اختبارية ما بشكل صحيح، يمكّننا الإجراء الإحصائي التكراري IRT من تقدير درجات القدرة الكامنة بالإضافة إلى المؤشرات المناسبة (كالصعوبة والتمييز والتخمين الزائف) للنموذج الاختباري. ومن أهم مميزات IRT كون نماذجه مستقلة عن الاختبار وعن الفقرة الاختبارية (على المستوى النظري على الأقل).

يمكن استخدام IRT لتحقيق أغراض أخرى مثل وضع وتصميم الاختبارات ومعادلة الاختبارات ومعايرة الفقرات الاختبارية وتقييم عمل الفقرات الاختبارية التباييني (DIF) والاختبار التكيفي.

مصطلحات للمراجعة:

راجع معاني المصطلحات المدرجة أدناه والتي تم استخدامها في هذا الفصل. الرجاء مراجعة فهرس التعريفات أو المعجم في حال وجود أي التباس حول معاني المصطلحات.

الاختبار التكيفي	التحليل العاملي
معيار العمر	الخطأ السلبي غير الحقيقي
فقرات الاختبار	الخطأ الإيجابي غير الحقيقي
النسبة القاعدية	معامل التعميم
الإرتباط القانوني	نظرية التعميم
المعامل ألفا	معيار الصف الدراسي
معامل التكافؤ	صدق الإتساق الداخلي
معامل الثبات	القياس على المستوى الفئوي
معامل الثبات والتكافؤ	تحليل فقرات الاختبار
الاختبار التكيفي المحوسب	معامل السهولة والصعوبة
الصدق التلازمي	معامل التمييز
صدق البناء	نظرية استجابة الفقرة (IRT)
صدق المحتوى	معادلات كودر- ريتشاردسون
المصادقة عن طريق التقارب	الارتباط
الاختبار ذو المرجعية المحكية	القياس

obeikandl.com

صدق المحك

عمل الفقرة الاختبارية التباييني

التحليل التمييزي

المصادقة على المميز

المظاهر

التعريف العملياتي

العملياتية

المستوى الرُتبي للقياس

معامل الصورة الاختبارية المكافئة

تحليل المسار

تقنين الاختبارات

العينة العشوائية متعددة الطبقات

جدول المواصفات

معامل إعادة الاختبار

ثبات العينة

معيار العمر العقلي

(DIF) الانحدار المتعدد

المستوى الاسمي للقياس

الدرجات المعيارية المطبّعة

الاختبار ذو المرجعية المعيارية

معدل الاختيار

طريقة التجزئة النصفية (ثبات نصف الاختبار)

الخطأ المعياري للتقدير (التخمين)

الخطأ المعياري للقياس

الدرجات المعيارية

المعيار المئيني

الصدق التنبؤي

القياس النفسي

نموذج راش

مؤشر النسب

الثبات

obeikandl.com

المراجع المقترحة:

Drasgow, F., & Hulin, C. L. (1990). Item response theory. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed.) (Vol. 1, pp. 577–636). Palo Alto, CA: Consulting Psychologists Press.

Embretson, S. (1993). Psychometric models for learning and cognitive processes. In N. Frederiksen, R. J. Mislevy, & I. I. Behar (Eds.), *Test theory for a new generation of tests* (pp. 125–150). Hillsdale, NJ: Erlbaum.

McKinley, R. L. (1989). An introduction to item response theory. *Measurement and Evaluation in Counseling and Development, 22,* 37–57.

Shavelson, R. J., Webb, N. M., & Rowley, G. L. (1989). Generalizability theory. *American Psychologist, 44,* 922–932.

Thorndike, R. L., & Thorndike, R. M. (1994). Reliability in educational and psychological measurement. In T. Husén & T. N. Postlethwaite (Eds.), *International encyclopedia of education* (2nd ed.) (Vol. 9, pp. 4981–4995). Tarrytown, NY: Elsevier.

Weiss, D. J., & Yoes, M. E. (1991). Item response theory. In R. K. Hambleton & J. N. Zaal (Eds.), *Advances in educational and psychological testing: Theory and applications* (pp. 69–95). Boston: Kluwer.

Yen, W. M. (1992). Item response theory. In M. Alkin (Ed.), *Encyclopedia of educational research* (6th ed.) (Vol. 2, pp. 657–666). New York: Macmillan.

Zeller, R. A. (1994). Validity. In T. Husén & T. N. Postlethwaite (Eds.), *International encyclopedia of education* (2nd ed.) (Vol. 11, pp. 6569–6576). Tarrytown, NY: Elsevier.

obeikandl.com