13

التكييف عبر الثقافات للاختبارات التربوية والنفسية

بيترف. ميريندا
جامعة رود آيلند

إن مجال الاختبارات والتقييم التربوي والنفسي مليء بالممارسات المغلوطة العديدة التي نجم عنها كثير من العواقب الخطيرة. لسوء الحظ، فإن علماء النفس ومن يتفق معهم من الخبراء قد فشلوا في إدراك خطورة هذه المشكلات بشكل عام والتي سببها سوء استخدام الاختبارات. من الملاحظ أنه بين الأخطاء الرئيسة (التي تأتي في المقام الأول) المرتكبة، هي تطبيق وتفسير أدوات التقييم بشكل غير منظم على الممتحنين الذين يُسبب لهم معوقات اللغة وعوامل ثقافية أخرى إبطال صدق الاختبار. إن أوضح مثال على تلك الأخطاء هو سوء استخدام الأدوات عبر الثقافات في التفسير الخاطئ للدرجات التي تم الحصول عليها من تطبيق التقاويم غير المناسبة للثقافة المتلقية.

خلال الخمسين سنة الماضية التي ظهرت خلالها هذه الاستخدامات الخاطئة والتي بدأت على نطاق واسع في أنحاء العالم كافة، قام العديد من علماء القياس النفسي وعلماء النفس الدوليين بالإضافة إلى المؤلفين، بالكتابة بشكل مكثف عن هذه الأخطاء كي يتم تجنبها وعن الطرق المناسبة لتعديل أدوات القياس ثقافياً (انظر: بيهلينغ ولو، 2000، هامبلتون و دي يونغ، 2003، فان دي فيفر و لونغ،

1997). ونذكر بعضهم أيضاً، بيري (1997)، برسلين، لونر، ثورنديك (1973)، كرونباخ و درنث (1972)، كيسنجر (1994) أولمدو (1979)، بورتينغا (1995)، وسبيلبيرغ، دفيلس، وبوكلك (1994).

إن المشكلات الأخرى، التي غالباً لا تتم مناقشتها في الموضوعات المتعلقة بنقل أدوات التقويم عبر الثقافات، هي من غير ريب الافتراضات البسيطة في الثقافات المتلقية. وهي (أ) إن الأدوات ذات الجدارة والصادقة والمناسبة في ثقافة ما، يمكن تكييفها بسهولة لثقافات أخرى. (ب) التغاضي عن حقيقة أن هذه الأدوات نفسها صحيحة حسب القياس السيكولوجي في ثقافة المنشأ كما هو مفترض (ميريندا، 1994)وهذا الأخير ينطبق على بعض الأدوات المكيفة الجديرة والمستخدمة بشكل واسع عبر الثقافات ميريندا (1995).

إن هذا الفصل يقدم بياناً مختصراً (عن السنوات الأولى من القرن العشرين) عن تكييف الاختبارات بهدف التغلب على عوائق اللغة في الولايات المتحدة والخارج. يقدم الجزء الأكبر من هذا الفصل التجارب والمشكلات التي واجهت المؤلف طوال 40 عاماً، كان يقوم خلالها بإنشاء أدوات تقويم تربوية ونفسية في الولايات المتحدة والقيام بأبحاث شاملة لتكييفهم إلى ثقافات أخرى. يوجد أيضاً موجز للإجراءات السليمة في تكييف الأدوات (انظر هامبلتون و باتسولا، 1999، فان دي فيفر و لونغ، 1997).

التكييف في الاختبارات التربوية والنفسية عبر الثقافات

إن إحدى الممارسات الخطيرة وعديمة التأثير في حقل الاختبار الثقافي والتقويم النفسي في نصف القرن الماضي، والمستمر حتى الآن، هو النقل السيئ لقياس الأدوات من ثقافة إلى أخرى أو إلى ثقافة ثانوية (ميريندا 1993). إن الممارسة السيئة تتضمن استعارة اختبار من ثقافة واستخدامه في ثقافة أخرى. لاحظ أن التأكيد هو على الاستخدام لا على تكييفه إلى ثقافة أخرى. يكون هذا

بالقيام بالترجمة الحرفية أكثر منها بالترجمة المتقدمة أو الترجمة الراجعة. وكما هو واضح في هذا الكتاب، هناك حاجة إلى الإثبات التجريبي أيضاً. بالرغم من أن ترجمة الاختبار يمكن أن يتبعه محاولة إعادة القياس، هناك ممارسة شائعة أخرى وهي ببساطة القيام بتفسير الدرجات المبنية على أساس القياسات الأصلية. لا يعطي تكييف البنود إلى الثقافة المتلقية إعادة قياس الإجراءات الإدارية والدرجات، وإثبات قياس البنية، غالباً الاعتبار المطلوب في تكييف الاختبار للاستخدام في لغة مختلفة حسب مبادئ قياس سيكولوجية معتمدة (الجمعية التربوية الأمريكية للأبحاث، الجمعية الأمريكية النفسية، المجلس القومي للقياسات التربوية 1999).

موجز عن التكييف الثقافي للاختبارات

من الممكن تتبع المحاولات الأولى لإقامة أدوات تقويم للتغلب على نقص المهارة في لغة الثقافة/ الثقافة الثانوية والتأثيرات المتحيزة على العوامل الأخرى التي تؤثر على الأداء إلى أوائل القرن العشرين. قامت إحدى المحاولات الرئيسة في الولايات المتحدة خلال الحرب العالمية الأولى مع جيش بيتا (Beta). أجرى تطوير الاختبار علماء نفس جرى تعيينهم في فيلق مهندسي التصريف الصحي، وحدة جيش خاصة تحت إمرة الميجور ر.م.يركيس، الذين قاموا بإدارة عملية تطوير الاختبار.

صُمم اختبار الجيش بيتا (Beta) لتقييم المجندين المتطوعين الذين كانوا إما أميين أو يعرفون لغة أجنبية. يتكون الاختبار من صور وأشكال تتطلب قلة أو عدم معرفة باللغة الإنكليزية. ويتضمن الاختبار أيضاً اختبارات أداء ثانوية مثل المتاهات، تصميم المجموعات وإنشاءات هندسية.استخدم اختبار بيتا بشكل كامل في الجيش بين عامي 1917 و 1918 جنباً الى جنب مع اختبار الأبجدية اللغوي للجيش.

تاريخياً، فإن اختبارات ألفا وبيتا هي اختبارات تُطبق على مجموعات ويُعمل بها في الولايات المتحدة بشكل واسع. في وقت مبكر، منذ عام 1904 طُورت لوحة استمارة لاختبارات أداء لا تتطلب معرفة باللغة الإنكليزية واستخدمت على مستوى أصغر.

لأسباب خاصة ذات دلالات تاريخية، ربما بسبب شهرة الشخص الذي أنشأ واستخدم تلك اللوحة، طُورت لوحة الاستمارة من قبل ر. س. وودورث. في عام 1904 طبق وودورث ذلك الاختبار في المعرض العالمي في سانت لويس على أطفال لا يتكلمون اللغة الإنكليزية. لاحقاً في دراسة عن الاختلافات العنصرية، استخدمت وطبعت النتائج مع لوحة الاستمارة واختبارات أداء أخرى (وودورث 1910).

قبل دخول الولايات المتحدة الحرب العالمية الأولى (1917- 1918)، أدرك هـ.أ. نوكس (1914) الخطورة والعجز في اختبار المهاجرين (الذين يعانون من علل عقلية) في جزيرة أليس باستخدام أدوات طورت باللغة الإنكليزية في الولايات المتحدة. بدأ بتطوير اختبارات لا تتطلب إجابات باللغة الإنكليزية. إن اختبار لوحة الاستمارة التي طورها نوكس كانت لوحة استمارة تقرير الإجابة "Casuist Form Board". كانت تلك اللوحة معقدة أكثر من لوحة استمارة سكوين. وتتكون بشكل رئيس من دوائر خشبية وحواجز دائرية. طور نوكس أيضاً لوحة استمارة حيث يجب على الممتحنين استخراج نسخة عن نموذج مطبوع على مجموعة من المكعبات. في ذات العقد (1910-1919)، جرت عدة تطورات مماثلة. امتحن هيلي وفيرنالد (1911) الأطفال في معهد الأحداث للاضطرابات العقلية في شيكاغو الناطقين بلغة أجنبية واحدة، أو الذين لديهم إعاقة تمنعهم من الإجابة في الاختبارات العادية الموجودة في ذلك الوقت.

أحد الاختبارات غير الشفهية التي استخدموها، كانت اختبار إكمال الصور (Ebbinghaus picture Completion).

في عام 1911 بدأ رودولف ودونالد باترسون في تطوير مقياس اختبار الأداء المعروف والمستخدم بشكل واسع. تم نشر هذا المقياس (الذي كان الاختبار الأول للاستخدام الميداني) وأصبح متوفراً للاستخدام العملي بعد عدة سنوات (بنتز وباترسون1917) يحتوي ذلك المقياس على 12 اختباراً منها 8 اختبارات شفهية. من

ضمن اختبارات الأداء في المجموعات، كان ترتيب الدوائر البلاستيكية، حل الأحجيات، اختبار إكمال الصور واختبار محاكاة.

في العقد الثاني من القرن العشرين، كان ستانلي بورتوز يطور اختباره الشهير، اختبار المتاهة، الذي تم نشره للمرة الأولى في منتصف العقد (بورتوز، 1915). في بداية العقد التالي، طور كوس (1920) اختبار تصميم المجموعات (تصميم تجريبي يُقسم فيه أفراد التجربة إلى فئات متعددة، كل فئة تمثل مجموعة متجانسة بالنسبة إلى هدف التجربة) الذي بقي حتى الآن كجزء من مجموعة قياسات أداء واختبارات ذكاء طورت في الولايات المتحدة، وجرى استعارتها لثقافات أخرى. اتبعت غودنوف الاختبارات غير الشفهية (1926) وطورٌ اختبار " ارسم رجلاً" لقياس ذكاء الأطفال بين عمر 3-13 سنة. طُلب من الأطفال أن يرسموا صورة رجل، "أفضل صورة يمكن رسمها". كانت الدرجات حسب عدد الأجزاء المهمة للرجل مثل العينين، الأصابع، الأنف، الفم. إلى ما هنالك، الموجودة في رسم الطفل. لم يكن للنوعية الفنية أي أهمية. قادت المعايير التي اعتمدت على 400 طفل إلى استنتاج العمر العقلي(MA) ومعامل الذكاء (IQ) فيما بعد، طُورت تلك المعايير كمقياس لتشخيص علم النفس المرضي للأطفال الشاذين.

التطورات خارج الولايات المتحدة

كانت المحاولة الأولى المهمة خارج الولايات المتحدة لتطوير أداة غير متحيزة ثقافياً، حيث الكتابة أو اللغة الشفهية له جزء صغير جداً أو حتى منعدم، هي اختبار «ريفن الجدول التقدمي» (Raven Progressive Matrices) (ريفن1938) اليوم، وكما استخدم في بريطانيا منذ 50 عاماً، يستمر الاختبار بكونه الأداة للتقويم الشفهي الأكثر شعبية الذي طُبق في العالم، حيث يوجد الكثير من برامج تقويم المقاييس.

في القارة الأوربية، تم تطوير اختبار شفهي فريد في إيطاليا خلال الحرب العالمية الثانية. تم تصميم اختبار "بيديني" لتوزيع الانتباه لقياس مقدرة الشخص

على تركيز انتباهه على مهمة يجب عليه القيام بها بسرعة كبيرة، ثم يحول انتباهه إلى مهمة معكوسة (ميجليوورينو1947). إن اختبار بيديني مفيد في تشخيص الضرر الدماغي أو الضرر في الجهاز العصبي. يعد ترميز ألوان التنقيط والخطوط العامل الأساسي في اختبار الأداء، والعائق الأول لأداء ناجح هو على الأغلب بدني (عمى الألوان، رؤية ضعيفة). جربت دراسة اختبار بيديني في الوقت الحاضر في الولايات المتحدة بغرض تطوير مقياس وصدق الأداة لاستخدامها في الثقافة في أمريكا الشمالية (ميريندا ودليوناردو 1992).

استخدم اختبار شفهي آخر في إيطاليا (اختبار"G" كالفي 1970) لتقويم القدرة الاستدلالية للمفاهيم لشخص ما. قدّم هذا الاختبار كونه مقياس عامل وحيد للذكاء العام الذي يمكن أن يفسّر كمعامل ذكاء تام، أو مبادئ تحليل "سبيرمان" للذكاء (g) الذي ما زال متوفراً للاستخدام. تم تصميم اختبار كالفي للاستخدام في إيطاليا، حيث لعبة الدومينو معروفة ومجموعات الألعاب شائعة في كل البيوت من البلاد. كانت الفكرة الضمنية هي إمكانية استخدامه عالمياً في كل ثقافة. بالرغم من وجود مجموعة من الإرشادات الكتابية مع الاختبار، إلا أنه يمكن أعطاؤها بالإشارة.

يتألف الاختبار من سلاسل متتالية للوجوه على قطع الدومينو. كل واحدة تعطي دلالة على تابع منطقي. على الممتحن أن يختار من عدة خيارات لقطع مختلفة الوجه الصحيح الذي يكمل تسلسل الفكرة. إن الكتيب المرافق والمنشورات التقنية صادقة للترجمة إلى لغة أي ثقافة متلقية. ولكن يبقى من الضروري للثقافة المتلقية أن تعيد إثبات خاصية القياس السيكولوجي.

إن البنود في اختبار كالفي لولبية، أي أنها تقدّم إلى الممتحنين حسب درجة الصعوبة المتنامية. يقدم بندان سهلان جداً في أول سلسلة البنود، ثم تصبح البنود أكثر صعوبة وتعقيداً وهكذا حتى آخر الاختبار.

تطور الاختبارات في الولايات المتحدة خلال الحرب العالمية الثانية

خلال الحرب العالمية الثانية، طور علماء النفس في الجيش الأمريكي اختبارين منفصلين للجنود المتطوعين. (أ) اختبار الأبجدية للجيش، وهو اختبار تجميع كتابي لتصنيف المتطوعين والمطلوبين للخدمة العسكرية المتعلمين والناطقين باللغة الإنكليزية. (ب) اختبار بيتا للأميين وللرجال الناطقين بلغة أجنبية. خلال الحرب العالمية، لم يكن هناك أي محاولة على مستوى كبير لخدمة الجنود والبحارة الذين كانوا يعانون من إعاقة بسبب عدم وجود المهارة اللغوية. كان للقوات البحرية وللقوات الجوية برامجها الخاصة للخدمات. طورت القوات البحرية اختبار مجموعة أشياء مترابطة ومصنفة للمستخدمين المجندين. تتكون المجموعة من اختبار مصنف عام (GCT)، اختبار حساب (ARITH)، اختبار كتابي (CLER)، و اختبار ميكانيكي (MECH) بينما طورت القوات الجوية مجموعة من اختبارات الأهلية واختبارات المهارات النفسية الحركية لاختيار متمرنين لثلاث وظائف لطاقم الطائرة، ربان الطائرة و المدفعي والملاح. طور الجيش اختبار التصنيف العام للجيش (AGCT) الذي يعطي قياس درجات عام يرتكز على مقياس ذي متوسط 100= (M) انحياز مقياس 20 = (SD).

يقوم هذا الاختبار بقياس القراءة، المفردات، وتقدير الحساب، الاستنتاج الحسابي والعلاقات المكانية. بالإضافة إلى استخدامه لتصنيف المجندين إلى اختصاصين مهنيين، واختيار الذين يحرزون درجات قياس أكبر من متوسط الدرجات بدرجة على الأقل (+ 120) للتدريب ليصبحوا ضباطاً. لم تكن هناك أية محاولة أثناء قيام الحرب لتطوير نسخة اختبار الجيش بيتا الذي استخدم في الحرب العالمية الأولى. لم يعط الجيش أي شرح عن عدم تطوير اختبار مواز لاختبار بيتا في الحرب العالمية الثانية. إن المعطيات التي قامت بجمعها وتحليلها هيئة أبحاث العلوم الاجتماعية وتم نشرها تحت رعايتها، تظهر بوضوح الحاجة إلى مثل

هذا الاختبار. إن اختبار التصنيف العام للجيش صنف المستخدمين إلى خمس فئات حسب الدرجات النهائية. كانت أغلب المجموعات الحاصلة على الدرجة الدنيا من السود أو الأشخاص ذوي الأصول اللاتينية (ستوفر وزملائه، 1950).

على أية حال، بعد انتهاء الحرب، وقّع قسم أبحاث المستخدمين في مكتب الجيش في واشنطن DC، على عقد اتفاقية مع مجلس البحث التربوي من كامبردج، مستشوست، لتطوير بدائل عديدة للاختبار الشفهي ليقوم مقام اختبار بيتا القديم. كان رئيس مجلس البحث التربوي، فيليب رولون، وهو أستاذ في القياسات والإحصائيات في جامعة هارفارد، كلية التربية. وقد عمل كثير من طلاب الدكتور رولون معه لإتمام شروط العقد. في عام 1952 تم تسليم الجيش اختبار الذكاء اللفظي (STI). تم تصميم تلك الأداة الشفهية لتقويم مفهوم قدرة الاستدلال للشخص. عند القيام بذلك الاختبار يتم تعليم الممتحن بالإشارة معنى إشارات خاصة عالمية، مجردة وملموسة. تم إنشاء ذلك الاختبار على مبدأ "التشبع الثقافي" بمواجهة " ثقافة محايدة". تم ربط مجموعات مركبة لأشكال هندسية مستوية مثلاً، مربعات، دوائر، معينات ومثلثات مع أشكال سوداء مظللة لحيوانات، أو أنثى إنسان في أوضاع متحركة أو ثابتة. على سبيل المثال تقفز أو في وضع الاستلقاء. لقد تم نشر الاختبار للاستخدام في الجيش في الولايات المتحدة، ولكنه لم ينشر للاستخدام المدني مثل اختبار التصنيف العام للجيش (AGCT). أراد رولون أن يسمح له الجيش باستخدام استمارة ينشرها مجلس الأبحاث التربوية بعد تقاعده من جامعة هارفارد (1967)، ولكن لسوء الحظ، توفي في ربيع 1968 وذلك قبل أن ينشر هو ورفاقه الطبعة المدنية للاختبار.

التطور بعد انتهاء الحرب العالمية الثانية

ابتدأ استخدام أدوات التقويم التربوي والنفسي (الاختبارات، المقاييس والبيانات) في حقل المعرفة والانفعالات بعد انتهاء الحرب العالمية الثانية بوقت قصير. وقد تم تطوير تلك الأدوات بشكل أساسي في الولايات المتحدة، بريطانيا

وفرنسا حيث توسعت استخدام التقاويم التربوية والنفسية بشكل سريع. وذلك في الخمسينيات والستينيات حين تسارع نمو علم النفس كمجال للعلوم في الولايات المتحدة. وبدأت الممارسة المهنية وتوسعت محفزة الاهتمام الكبير في عملية التقويم. أما في بلدان العالم الثالث، فقد بدأ علماء النفس والاختصاصيون باستعارة أدوات التقويم المطورة والمقاسة في بلدان أخرى بسبب نقص الخبراء ارتفاع التكاليف. لسوء الحظ، كان الإجراء الأساسي لنقل الاختبار من ثقافة إلى أخرى هو الترجمة اللغوية. وتستمر هذه الممارسة حتى الوقت الحالي إلى حد كبير. في بعض الأحوال يعطى قليل من الانتباه أو لا يعطى أي انتباه إلى الاختلافات الثقافية التي يجب اعتبارها عند تعديل أو تغيير البنود في الاختبار. كما لا يعطى أي اهتمام بإعادة القياس، المعايرة، وإجراء تكافؤ الدرجات والمسودة الأصلية للاختبار.

لإيضاح الفكرة السابقة، يجب الإشارة إلى أن خطأ الممارسة يحدث في البلاد الأقل تطوراً أو حتى في البلاد المتطورة. في أوربا، خاصة في بلدان البحر المتوسط. أما في بلدان شمال الألب وفي البلاد الإسكندنافية، فيبدو أنهم مطلعون على الإجراءات، ويمارسون مبادئ القياس السيكولوجي بشكل صحيح. أما في هولندا فهم يقدمون عدداً كبيراً من علماء القياس السيكولوجي المدربين جيداً والذين يمارسون على نطاق واسع في العالم. على سبيل المثال، عند مراجعة ثلاثة إصدارات لمجلة في القياس النفسي (Psychometrika) كانون الثاني 1998 حتى أيلول 1999) تبين أن 13 من أصل 25 مقالة منشورة قد كتبها علماء قياس سيكولوجي من هولندا.

ازداد في الولايات المتحدة التنوع في الثقافات والأعراق في الخمسينيات والستينيات. ولكي يستطيعوا تطبيق تشريع الإجراء الإيجابي الأول 1946، كان على ناشري الاختبارات الحصول على تقويم تربوي ونفسي لأدوات تم نشرها سابقاً في اللغة الإنكليزية وترجمت إلى لغات الأقليات. كان هذا صحيحاً بالنسبة للمجموعات من

الأصول اللاتينية، خاصة الأطفال في كوبا وبورتوريكو. على كل حال، هناك الكثير من المشكلات القياسية في تلك الطبقات باللغة الإسبانية ولغات أجنبية أخرى، لأن أكثرها تُرجم حرفياً ولم يجر تكييفه ثقافياً حسب الإجراءات القياسية الصحيحة.

لتأكيد هذا الخطأ وممارسته المستمرة في الولايات المتحدة من قبل ناشري الاختبارات، على المرء أن يدرس الفهارس الحالية التي تقوم بالدعاية لبيع مطبوعات لاختبارات اللغة الإنكليزية في لغات أجنبية. ما عدا بعض الاختبارات، فإن كل المطبوعات باللغة الأجنبية، هي ترجمة حرفية منقولة عن الأداة الأصلية.

إجراءات لتكييف الأدوات

إن الإجراءات المناسبة لتكييف الأدوات عبر الثقافات قد تم إيجازها في هذا الفصل، وجرى وصفها وشرحها في فصول أخرى من هذا الكتاب وفي مطبوعات أخرى (فان دي فيفر ولونغ 1997). هذه الإجراءات الشاملة والطويلة كانت السبب للنفقات المادية والوقت والجهد الذي أُنجز لتكييف سليم. ليس من الطبيعي أن تتطلب الإجراءات، لكي تكون ذات فاعلية، سنوات عديدة من العمل الجاد. هذه الجهود المشتركة التي قام بها اختصاصيون من البلدين المنشأ والمتلقي. الإجراءات هي كما يلي:

1- إن الخطوة الأولى هي الأخذ بالاعتبار الأدوات، التقنيات أو الوسائل المناسبة للتكييف. يجب أن يتضمن هذا الرأي تقويماً موضوعياً في صدق مواصفات القياس للمعايير في ثقافة المنشأ. يجب تجنب الخطأ، وهو افتراض صدق الأدوات لأنها ببساطة، استخدمت بشكل موسع.

2- قبل البدء بخطوة الترجمة يجب مراجعة البنود واستمارات الإجابة لاختبار طرق المجموعات الثقافية المحددة أو المجموعات الدولية.

3- في خطوة الترجمة، الترجمة الباكرة، والترجمة الراجعة، يجب الانتباه للتأكد من أن المترجمين خبراء في اللغتين ويعملون بشكل منفرد في المرحلتين. كما يجب الانتباه إلى صدق ترجمة المتغيرات في اللغة واللهجات لتأثيرها على الترجمة من حيث المفهوم والمعنى. وهذا شيء صحيح خاصة للغات اللاتينية. مثلاً اللغة الإسبانية في بورتوريكو مقابل الإسبانية في المكسيك، اللغة البرتغالية في أمريكا الجنوبية مقابل البرتغالية في البرتغال، أو اللغة الفرنسية مقابل الفرنسية في كندا.

4- يجب دراسة كل بند حسب إمكانية تكييفه إلى الثقافة المتلقية. من الثابت وجود بعض البنود التي لا يمكن نقلها مباشرة. يمكن تعديل هذه البنود وإصلاحها أو إهمالها ووضع بدائل، قبل البدء بتطوير نسخة تجريبية للاستخدام في الدراسة الاستطلاعية الأولى.

5- القيام بدراسة استطلاعية، حيث يجري استخدام النسخة التجريبية حسب الأعراف الثقافية، الممارسات، العادات، إلى ما هنالك للحصول على نماذج ذات تطابق كان في المجموعات التي تطبق عليها المقاييس.

6- عند تحليل المعطيات التي تم الحصول عليها عند تطبيق النموذج التجريبي، وهي خطوة مبكرة، يجب دراسة بنية وعينة (العامل أو العناصر) الأداة ومقارنتها مع بنية وعينة الأداة الأصلية. انظر غيرل، (2000) في هذه المرحلة، يمكن أن يقرر الباحثون ضرورة إعادة بعض الخطوات السابقة قبل الاستمرار.

7- إذا كان القرار أنه بالإمكان القيام بخطوة نحو تطوير النسخة المكيفة، فإن الخطوة التالية هي القيام بالتحليل الإحصائي الضروري والمطلوب لإثبات خصائص القياس السيكولوجي في الأداة المكيفة ثقافياً. في هذه المرحلة، وكحد أدنى، يجب حساب درجة الثبات الداخلية للبنود وصدق الدرجات والمعايير.

8- أما الخطوة الأخيرة في تطوير النموذج التجريبي، فهي القيام بدراسات للبنية ولصدق المعايير تنسجم مع الأغراض الأخرى التي تنوي الأداة استخدامها في الثقافة المتلقية؛ لأن نتائج بحث صدق الأداة الإيجابية سوف تؤمن مستخدمين محتملين وناشرين/ موزعين جداً في الثقافة المتلقية مع الثقة الحقيقية، وهي أن أداة مطورة جديدة مكيفة ثقافياً جاهزة للاستخدام الميداني.

إن هذه المجموعة من الخطوات الثمانية تتماشى مع المخطط الدولي لهيئة الاختبارات وتكييف الاختبارات الذي قُدم في الفصل الأول من هذا الكتاب (ووسع من قبل هامبلتون وباتسولا 1999).

نماذج من إجراءات استخدمت في تطوير الأدوات المكيفة اختبرها المؤلف خلال ٢٨ عاماً (- ١٩٦٧- ١٩٩٥)

جرى استخدام الأدوات المكيفة ثقافياً بشكل أولي في بحث أو أكثر من مشاريع الأبحاث عبر الثقافات وهي: (أ) تعيين المواهب في الدول النامية، 1977-1967، (ب) مطابقة الأولاد الصغار مع مشكلات التعليم، (1995-1975) و(د) الإدراك الحسي العام للقادة العالميين. 1964-1990، جرى ترجمة الأدوات لجذر من اللغة الإنكليزية إلى لغة البلد الذي يجري فيه البحث.

فيما يلي وصف مختصر لهذه الأدوات:

تحليل القوة الموجهة للنشاط (Activity Vector Analysis):(AVA) هي قائمة صفات أنشأها ولتر ف.كلارك. أصبحت AVA جاهزة للاستخدام. في 1948 استخدمت بشكل مبدئي في مجال الصناعة والتجارة في الولايات المتحدة (كلارك 1956). يحتوي الاختبار بشكله الأولي على 81 صفة للسلوك والتي تعطي درجات حسب أربع مقاييس وثلاث "أربعة عوامل" لصور صممت لقياس إدراك قدرات ذات الشخص الأساسية. الذات الاجتماعية والذات المركبة. تم إنشاؤها على

نظرية الذات الأساسية لـ لكي (لكي1945) وعلى النظرية الانفعالية لـ و.م. مارستون (1928 مارستون و.م.، كينغ ومارستون، ي. هـ 1931).

قياس المهارات (Measurement of skills)(Mos) : مجموعة من ثمانية اختبارات أنشأها ولتر ف.كلارك وزملاؤه. صمم هذا الاختبار لمساعدة العاملين في دائرة الموظفين في الاختيار السليم، والتصنيف وفي تعيين المهمات للموظفين. كانت الأدوات اختبارات قصيرة (الوقت المحدد 5 الى 7 دقائق) جرى توسيعها لمضاعفة الصدق لدقيقة من الاختبار. في هذه المجموعة من اختبارات الأداء المعرفي العملي، كانت المهارات المقاسة: (أ) المفردات، (ب) الأرقام، (ج) الأشكال (د) السرعة والدقة (هـ) التكيف (و) التفكير (ز) الذاكرة، و(ح) مهارة استعمال الأصابع (كلارك 1960).

مقياس التعريف بالطلاب في رود آيلند (Rhode Island Pupil Identifi cation Scale) (RIPIS):

صمم هذا المقياس للتعرف على الطلاب في الصفوف العادية (K-2) الذين يعانون صعوبات في التقدم المدرسي لعدة أسباب. يتألف المقياس من جزأين. يرتكز القسم الأول المكون من 22 بنداً على سلوك الطالب في الصف، والذي يمكن للمعلم مراقبته وإدراكه. ويعتمد القسم الثاني المكون من 19 بنداً على السجل المدرسي للأداء، والذي تسجله المعلمة في ملف كل تلميذ. هناك خمسة عوامل يتم الحصول عليها من القسم الأول للاختبار الأصلي: الإدراك الجسدي، تعاون حركي حسي، الانتباه، إدراك الذات وذاكرة الأحداث. تكون قيمتهم 67.5٪ من معامل التفاوت الكامل. وهناك أربعة عوامل من القسم الثاني للاختبار: الذاكرة لإعادة إخراج الرموز، التوجيه الثابت، الترتيبات المكانية المتتابعة، وذاكرة للرموز لعمليات الفهم. كانت قيمة هذا الاختبار 68 ٪ من معامل التفاوت الكامل (نوفاك، بونافنتورا وميريندا 1972/1979).

قوائم مسح المواهب في مشروع فلاناغان (Flanagan's Project Talent Inventories):

استخدمت ثلاثة من قوائم فلاناغان في الأبحاث عبر الثقافات. وكانت كالتالي (أ) قائمة الاهتمامات، (ب) قائمة نشاطات الطالب، و(ج) قائمة فراغات تملأ بالمعلومات. كانت قائمة الاهتمامات تحتوي على 205 بنود تعالج 122 مهنة ونشاط يختار الطالب منها ما يهتم به (5 نقاط من مدرج مقياس الشخصية). وتحتوي قائمة نشاطات الطالب 150 بنداً حيث يجيب الطالب على (5 نقاط من مدرج مقياس الشخصية) للجملة التالية، "الأشياء التي أقوم بها وكيف أقوم بها، هل تصفني هذه الجملة". وتقدر إجابته من جيد الى جيد جداً. أما القائمة الفارغة للمعلومات فتتألف من 394 بنداً وفيها أسئلة منتقاة بدقة عن الخلفية الاجتماعية والخطط، طموحات طلاب في المرحلة المتقدمة في المدرسة. تنقسم هذه القائمة الى 7 أقسام، يستعلم القسم الأول (K=115) عن النشاطات التي شارك فيها الطلاب حتى ذلك الوقت. يتألف القسم الثاني (K=45) من أسئلة حول العائلة والمنزل. أما القسم الثالث والرابع فلهما علاقة بطبيعة الأعمال التي قام و بها والدا الطالب أو رب العائلة. ونشاطات أخرى قام بها أفراد العائلة. يتعلق القسم الخامس بالحالة الصحية للطالب (K=42) والجزء السادس بمخططات الطالب المستقبلية (K=75) وأخيراً يتعلق الجزء السابع (K=15) بالخطط المستقبلية للانتساب إلى الجامعة (فلاناغان وزملائه، 1964، ميريندا، ميجيليورينو1997).

تم اتخاذ الحيطة في كل مرحلة للتأكد من أن الترجمة كانت صحيحة للغة المجموعة التي تم تكييف الأداة لاستخدامها عليها. على سبيل المثال (أ) في تحليل القوة الموجهة للنشاط (AVA) جرت ترجمة الصفات لمختلف المجموعات الناطقة بالإسبانية. وحسب هذا تُرجمت الى الإسبانية القشتالية (الرسمية) للاستخدام في إسبانيا. اللهجة الإسبانية المستخدمة في كوبا/ بورتوريكو للاستخدام في الولايات المتحدة الشرقية، والإسبانية شيكانو المستخدمة في الولايات المتحدة الغربية

والمكسيك. استخدمت ذات العملية في تطوير اللغة الفرنسية. جرى التكييف الى الفرنسية الباريسية للاستخدام في فرنسا، الفرنسية الكندية للاستخدام في كندا، وفرنسية معدلة للاستخدام في السنغال. طُور شكلان من اللغة البرتغالية واحد منهما للاستخدام في البرتغال والثاني للاستخدام في البرازيل. جرى ترجمة مقياس التعريف بالطلاب الى لغات عديدة: الفارسية (إيران)، الماندرين الصينية (تايوان)، البولندية (بولندا)، الدنمركية (الدنمرك)، الفرنسية الكريبولية (هاييتي)، الإيطالية (إيطالية وسيسيليا). كما جرى ترجمة استبيانات فلاناغان الى الإيطالية الرسمية (توسكانا). بالرغم من أن هذه الأدوات قد تم استخدامها في مشروعات أبحاث كبيرة جرت في سيسيليا، فإن اللغة الإيطالية الرسمية هي اللغة الوحيدة التي تستخدم في المدارس في كل إيطالية. كان من الضروري اللجوء الى اللغة السيسيلية. بالرغم من كل الحذر الشديد أثناء القيام بمرحلة الترجمة - أثناء عملية التكييف، ظهرت كثير من المشكلات. وبعضها أكثر جدية من الأخرى. كان أكبرها وأكثرها كلفة المشكلة التي حدثت في تكييف مقياس التعريف بالطلاب (RI-PIS) في ايطاليا.

أجرى دراسة النسخة الإيطالية داميكو وزملاؤه (داميكو، ميريندا وسباراسينو، 1982 كان عامل الانتباه في الاختبار يتكون من ثلاثة أقسام: "الصعوبة في الجلوس ساكناً" #9، "الصعوبة في الوقوف ساكناً" #10، " ومدى انتباه قصير" #11.

كشف تحليل العناصر الرئيسة حسب تفاوت التناوب أن معظم العناصر تتداخل مع بنية المقياس في الولايات المتحدة. على كل حال، تم الكشف أيضاً عن نتيجة غريبة في عامل الانتباه. أخفقت البنود 9-10-11 في الربط مع المجموعة. ولكن بندين آخرين " البكاء" #15، "الإخفاق في تحمل التأنيب جيداً" #16 كانا ذوي ثقل في المجموعة. في اختبار RIPIS، في الولايات المتحدة، تظهر البنود 15-16 في عامل إدراك الذات على نحو متعاقب. أدت تلك المفارقات الجدية الى تكرار

الدراسة مع نموذج جديد مكون من 1.571 طفلاً في الصف k-2 من أربع مدارس في باليرمو/ سيسيليا. تم اكتشاف خطأ في ترجمة البنود 9-10-15 من اللغة الإنكليزية الى الإيطالية. كان المقصود من البنود 9-10 هو اتخاذه تلك الوضعية، ولكن الترجمة وجهت إجابة المدرسين الى مهمة الوقوف أو الجلوس بسكون. (وافق على تلك الترجمة أربعة مترجمين ثنائيي اللغة). كان الوضوح في المعنى المقصود ناجحاً في البندين 9-10 .

تم أخذ نموذج ثالث (N = 1.311) من باليرمو من أربع مدارس في الصف k-2 أصبح مقياس التعريف بالطلاب RIPIS باللغة الإيطالية الآن نسخة مكيفة ومعدلة من النسخة الأصلية (الأمريكية) (سيرسي، كارديكا و كانجيمي، 1992) وأصبح جاهزاً للاستخدام في 1995 وهو الآن يوزع من قبل أكبر دار نشر إيطالية Organizzazioni Speciali في فلورنسا.

في أثناء عملية تطوير نسخ من لغات أجنبية لاختبار تحليل القوة الموجهة للنشاط (AVA) كان من الضروري استبدال الصفات ببنية قواعدية أخرى. على سبيل المثال، حال، مصدر، وإلى ما هنالك. لم يكن ذلك لعدم وجود ترجمة حرفية للصفة، ولكن لأن الترجمة كان لها معنى مغاير في اللغة المقصودة. مثلاً كلمة (اجتماعي) في اللغة الإنكليزية تصبح في الفرنسية (يحب العيش ضمن مجموعة). وكلمة (يُعجب) تصبح في اللغة الإيطالية (يدرك الذات). بدأت الكثير من الأبحاث حول الصفات في AVA في 1957 في الولايات المتحدة، وفي عام 1967 بدأت في بلدان أجنبية حيث جرى تكييف قوائم الصفات بهدف القيام بالأبحاث.

بالإضافة الى التأكد من صحة الترجمة والمحتوى الثقافي في البند تبعه إجراء فعال في التكييف الثقافي لأداة التقويم، وهو تغيير إجراءات الاختبار الى الأفضل إذا كانت قابلة للتعديل في الثقافة المستهدفة، وإذا تم تصديقها حسب معطيات البحث.

تمت الإشارة الى ذلك في ميريندا، مايو غواداجنولي ويو- ون (1984). كان لدى المقياس الرئيس (نوفاك، بونافنتورا، ميريندا، (1979-1972) وكل الترجمات ما عدا الصينية، معلمون في صفوف يضعون تقديراتهم في نهاية كل شهر حسب ملاحظاتهم حول سلوك كل تلميذ، بالإضافة الى تدوين أدائه المدرسي في سجله الخاص. عند التخطيط لتطوير استمارة اختبار في اللغة الصينية، أقنعني إميلي مياو أنه من الأفضل أن تقوم مدرسة كل صف دراسي بملاحظة وتدوين إجابة التلميذ للبنود في نهاية كل شهر عوضاً عن الاعتماد على استرجاع الإدراك الحسي (مثال لملاحظات المدرسة: صعوبة القفز على الحبل، صعوبة التقاط الكرة، صعوبة في تذكر أشياء شاهدها). وهذا ما حدث بالفعل وبرهن على نجاحه في تطوير جاهزية الاستمارة الصينية لاختبار RIPIS.

ظهرت بعض الأحداث غير المتوقعة وغير المتنبأ بها في أثناء عملية التكييف الثقافية لعدد من الأدوات في اللغة الانكليزية إلى لغة وثقافة بلدان مختلفة لوضعها ضمن مشاريع الأبحاث عبر الثقافات بين الأعوام 1967-1995. على الرغم من الخطوات الحذرة التي تم اتخاذها، ونوقشت سابقاً، لم يتم منع حدوث بعض المشكلات التي تطلبت التصحيح. خُصص مشروع المواهب في سيسيليا الذي بدأ في 1967، في عامي الأول لاشتراكي في مشروع التبادل الثقافي بين إيطاليا والولايات المتحدة وعملي في مخبر العلوم النفسية التطبيقية، جامعة بالريمو، لتطوير مجموعة اختبارات لتعيين المواهب الكامنة لدى شباب سيسيليا.

في مرحلة تعيين المواهب في المشروع، تم جمع المعطيات باستخدام مجموعة أدوات مترجمة منها اختبار المهارة في المفردات (MOS-1). يقدم هذا الاختبار إلى الممتحن قاموس تعريف كلمة مختصر بالإضافة الى الحرف الأول للكلمة يتبعها عدد من الفراغات حسب عدد الأحرف المطلوبة لإكمال الكلمة. أحد هذه البنود في الاختبار الأصلي يقول: "نقود، خاصة نقود جاهزة، عملة أو ما يعادلها، تُدفع مباشرة

بعد الشراء". الإجابة الصحيحة في اللغة الانكليزية هي "نقداً" في مفتاح التصحيح. في النموذج الإيطالي تم الإشارة الى حرف (ن). كلمة مكونة من أربعة أحرف تبدأ بحرف (ن)، ولكن الكلمة الإيطالية تتكون من 8 أحرف. يتوقع من طلاب المرحلة المتقدمة في إيطالية معرفة الكلمة في اللغة الإيطالية، كما يعرف الكلمة بالإنكليزية طلاب المرحلة المتقدمة في الولايات المتحدة (ميريندا، كلارك و جابكس، 1965، ميريندا، هول وباسكال. 1962، ميريندا، جابكس و كلارك، 1966).

كانت الاختبارات في مجموعة اختبار المهارة في المفردات حلزونية (تدور حول نقطة مركزية). عندما تم إقامة تحليل البند لاختبار Mos باستخدام معطيات إيطالية، كشف ذلك البند صعوبة المستوى، أي أن قيمة الاحتمال للجواب الصحيح كانت منخفضة بعكس قيمة الاحتمال العالية المعروفة في الولايات المتحدة. بدراسة لاحقة لإجابات الأفراد تم الكشف عن أن كثيراً من الطلاب الإيطاليين أجابوا بالكلمة الصحيحة كما جاءت في مفتاح التصحيح (كمبيالة) وكانت هذه مصادفة لأن الكلمة لا تتماشى مع تعريفها في الاختبار، ولكن تم تعريفها إجابة صحيحة.لهذا السبب، بإجابتين صحيحتين في البند في الاختبار الإيطالي، ارتفعت قيمة الاحتمال (p) لتتماشى مع قيمة الاحتمال في الولايات المتحدة.

قبل استخدام بطاقة الائتمان في إيطالية، كان استخدام الكمبيالة طريقة شائعة في الثقافة الإيطالية خاصة بين الطبقة العاملة. كانت الكمبيالة مماثلة للعملة الورقية وكانت تستعمل لدفع الفواتير إذا لم تتوفر العملة النقدية. ولكن، الآن في إيطاليا يجري استخدام بطاقة الائتمان؛ لذلك إذا تم استخدام Mos -1 اليوم في إيطاليا يجب تعديل مفتاح التصحيح، وتعطى درجة للإجابة الصحيحة (بطاقة ائتمان).

في المحاولة الأولى لتطوير اختبار RIPIS الى الفارسية، أثناء عملية تحضير المعطيات لتحليل عامل الأداء التي يقوم بها المدرسون في إيران، توقف الكمبيوتر

فجأة عن العمل. بعد البحث، تبين أن هناك خطأ في البند 25، " أرجع الورقة التي يوجد فيها الكثير من الكلمات الممحوة". وكانت ببساطة غير مناسبة للنموذج الذي يزود الدراسة الاستطلاعية بالمعطيات. في نقطة تفعيل جدول الارتباط، لم يكن بالإمكان حساب معامل الارتباط بين ذلك البند وبقية بنود الاختبار؛ لأنه حصل على صفر في التباين. كانت إجابة المدرسين لذلك البند "ابدأ" تعطى نقطة واحدة من درجات المقياس. بذلك، أعطى تقدير البند متوسطاً واحداً وانحرافاً معيارياً صفراً. عند تحليل السبب غير الطبيعي لتلك النتيجة مع الباحثين الإيرانيين المشاركين، تقرر أن أقلام الرصاص والممحاة لا تستخدم في المدارس التي أخذت منها النماذج. وعوضاً عن ذلك تستخدم الفرشاة والطلاء في تعليم التلاميذ الصغار الذين يبدؤون في دراسة اللغة الفارسية. على كل حال، لا تتبع كل المدارس في إيران ذات الطريقة التي تستخدم في تلك المدارس. كان ذلك مصادفة لم تكن بالحسبان. كان المقصود أن يكون ذلك النموذج متطابقاً، يمثل التلاميذ الصغار في المدارس الحكومية من مختلف الطبقات الاجتماعية/ الاقتصادية. في منطقة فقيرة أو في مدارس خاصة من مناطق طبقة غنية (ميريندا، 1990b).

تم إدراك المشكلات التي واجهها باحثو تطوير الاختبار المكيف بسرعة وجرى إصلاحها. وتوصف الآن لتوضح ما يمكن أو ما يحدث حتى في حال اتخاذ الخطوات الضرورية عند تكييف الاختبار للاستخدام في ثقافة جديدة. عندما تكون الترجمة هي الخطوة الوحيدة المتخذة في نقل الأداة من ثقافة الى أخرى، بالرغم من دقة الترجمة، يمكن حدوث بعض المشكلات دون أن يتم اكتشافها. لكي يتم تكييف الاختبار بشكل سليم يجب جمع معطيات حقيقية، وتحليلها وشرحها.

تجربة حقيقية أخرى

قبل البدء في ذكر الأحداث الجدية والمؤسفة التي يتم حدوثها بالممارسة الخاطئة عند نقل أداة القياس من ثقافة الى أخرى باستخدام الترجمة فقط، التي

غالباً ما تكون خاطئة، نقوم بوصف هذه الأحداث: في أوائل السبعينيات كنت أعمل في المجلس الاستشاري في بتراليا سوبرانا/ سيسيليا. كان الأعضاء الآخرون في المجلس سكارفيا آندرسن، ساموبل ميسيك. ميريام غولدنبرغ، مارغريت ميد وتلميذتها التي كانت تُعد رسالة الدكتوراة جوزفين دانا.

كان كاهن البلدة، كالجرو لابلاسا، قد أسس المدرسة في تلك البلدة الجميلة. بعمله الدائب، استطاع الأب لابلاسا الحصول على تمويل خارجي ليس لتأمين تعليم مجاني فقط، ولكن لتأمين إقامة مجانية للأطفال الموهوبين ولتوسيع منشآت المدرسة سنوياً. كان الأب لابلاسا يمضي أكبر قسم من العام بزيارة المدارس لإيجاد واختيار الأطفال الموهوبين للصفوف في المستقبل، ولطلب الإذن بالحديث مع الطالب الذي تعتبره إدارة المدرسة الأفضل في المدرسة. بعد ذلك يتحدث الأب لابلاسا مع الطالب، وإذا كان ذلك الطالب مهتماً بالالتحاق بالمدرسة، عندئذ يتحدث مع الأهل لإقناعهم بالسماح لولدهم في أن يغادر البلدة/ المدينة ويلتحق بمدرسة الموهوبين.

كان من عادتي أن أمضي الصيف في سيسيليا أعمل على مشروع بحثي الخاص في جامعة بالميرد. وكان الأب لابلاسا يتطلع دائماً الى تمضية بعض الوقت في التشاور معي. ذات صيف، في لقائنا الأول ذلك العام، أخذني بعيداً عن زوجتي ووضعني مع كاهن آخر في سيارة ثانية عند ذهابنا الى الغداء. كان الأب لابلاسا منزعجاً جداً وأراد أن أتناقش مع الكاهن في موضوع كان قد رواه للأب لابلاسا. كان الكاهن الذي كان قد حضر دورة في إقامة ووضع الدرجات وتفسيرها لاختبار النسخة الإيطالية لمقياس ذكاء فكسلر للأطفال، قد أجرى الاختبار لكل الطلاب المسجلين في المدرسة. ما عدا طالب "موهوب جداً" كان قد حصل على مجموع 108 في معامل الذكاء، أحرز الطلاب الآخرون درجات ضمن الحدود العادية للدرجات (70–90). من غير الضروري القول إنني أمضيت أكثر الوقت في الطريق الى المطعم أشرح للكاهنين خطأ اختبار أطفال إيطاليين باستخدام اختبار جرى تحضيره،

وقياسه ووضع معاييره في الولايات المتحدة وتمت ترجمته فقط إلى اللغة الإيطالية. ما تم ترجمته لم يكن بنود الاختبار فقط، ولكن إجراءات التصحيح وكيفية تعيين درجات المقياس ومعامل الذكاء، والوصف الكمي لمدى الذكاء حسب المعايير في الولايات المتحدة. بذلك، تكون النتائج حسب ما توقعته.

الأحداث الجدية التي من الأسف حدوثها

إن المشكلات التي جرى حدوثها خلال الوقت الطويل للبحث في تكييف أدوات الاختبار عبر الثقافات، كانت جدية لكنها لم تكن مدمرة. بالرغم من أنها كانت سبب التأخير في تقدم تكييف ناجح ومضاعفة الكلفة من حيث الوقت والمادة والجهد. ولكن جرى إيجاد الحل لها في النهاية. على كل، بعض المشكلات الأخرى التي لم يتم معالجتها فوراً أحدثت أكثر من تأخير. كانت الإحداثيات الأساسية (أ) إحداث ترجمة للأداة من اللغة الأصلية الى أخرى، و (ب) الافتراض أن الأداة الأصلية جيدة. كانت النتائج مدمرة لكل الذين اشتركوا في تلك العملية. دعوني أشرح لماذا.

في أوائل عام 1928 كان ليندرو آلميدا، أستاذاً مساعداً في قسم العلوم النفسية والعلوم التربوية في جامعة بورتو/ البرتغال، طالب دكتوراة في جامعة لوفان/ بلجيكا. كان أستاذه والمشرف على رسالته جورج موريس مؤلف كتاب " مجموعة اختبار موريس للاستدلال المنطقي الفارق". حضر ألميدا، الذي أصبح فيما بعد عالم نفسي رائد في البرتغال ومؤلف الاختبار المكيف البرتغالي الميدا، (1988)، إلى الولايات المتحدة لطلب الاستشارة حول الاختبارات النفسية والقياسات النفسية. اختار ثلاثة من الباحثين لاستشارتهم: روبرت روسنثال (هارفارد)، روبرت ستينبرغر (ييل) وبيتر ميريندا (جامعة رود آيلند). جاء أولاً الى كامبردج ولكنه علم لاحقاً أن روسنثال كان في إجازة بحث علمي ولم يكن موجوداً. عندئذ جاء إلي قبل ذهابه إلى نيو هايفن للقاء ستينبرغر. بهذا، بدأت علاقة مهنية طويلة استمرت حتى اليوم.

في عام 1984 تقاعدت مبكراً من جامعة رود آيلند. كان السبب الرئيس لذلك التقاعد هو قبولي منصباً قصير الأجل في جامعة في الخارج لتدريس القياسات السيكولوجية لأعضاء كليات علم النفس الذين كان عليهم القيام بأبحاث حقيقية للحفاظ على مناصبهم أو التقدم في رتبتهم. كان عملي الرئيس بين الأعوام 1984-1993 في جامعات ليشبونة، بورتو، مينهو في البرتغال. خلال تلك المدة (كما في كل بلدان البحر المتوسط) كما في إيطاليا واليونان، كانت الجامعات تمنح شهادة واحدة فقط، شهادة البكالوريوس. كان الكادر التدريسي في المعاهد أو أقسام علم النفس مكوناً من أستاذ وحيد فقط أو عدد قليل من الأساتذة. كانت مسؤولية تدريس الطلاب تقع بشكل أساسي على عاتق " المساعدين" الحاصلين على شهادة البكالوريوس في علم النفس. كان في البرتغال مستويات للمساعدين، "أ" أو "ب" (أو كما يسمونه الآن الأستاذ المساعد أو الأستاذ المشارك). كان البحث الموافق عليه شرط أساسي للتقدم من المرتبة "أ" إلى المرتبة "ب". وكان يجري استفتاء عن المساعدين الذين لم ينجحوا في كليات الجامعة ليصبحوا أساتذة في المدارس في أحسن الأحوال؛ لذلك كانت الأبحاث المقبولة التي يمكن الدفاع عنها أمام اللجنة الفاحصة تأخذ الأولوية في الإنجاز من المساعدين في مرتبة "أ".

كانت حلقات التدريس قصيرة المدى التي كنت أدرسها يتضمن أغلبها تدريس الإحصائيات المتعددة التباين، منهجية البحث، تحليل العوامل، طرق القياس السيكولوجي، وحلقات معالجة الكمبيوتر. كانت تلك الحلقات شبيهة بالتي كنت أدرسها في الولايات المتحدة. بالإضافة الى تدريس تلك الحلقات في البرتغال، كنت ألتقي بشكل دوري مع المساعدين لمراجعة اقتراحات بحثهم وإعطائهم النصح. وكان لي بعض التحفظات على بعض الأبحاث التي كانوا يقومون بها. كان المساعدون يقومون باختبار أداة تقويم في نطاق علم النفس، التي يهتمون بها في إقامة البحث ثم يجرون ترجمتها الى البرتغالية.

كانت إجراءات جمع النماذج في البحث والطرق الإحصائية المستخدمة في تحليل المعطيات المبنية على تلك الأدوات المترجمة (بشكل أساسي من اللغة الإنكليزية الى البرتغالية) صحيحة بشكل عام ودقيقة. على حال، في بعض الحالات، بعد عدة شهور من العمل الجاد والتكاليف الباهظة تحدث الكارثة. عندما كنت أعود إلى البرتغال في العام التالي أو بعده كنت أجد المساعدين ينتظرون عودتي بفارغ الصبر. كان السبب الأساسي لذلك أن نتائج تحليل معطيات ذلك البحث لم يكن لها معنى، لا للباحثين ولا للمشرفين على البحث. وأكثر من ذلك كانت هناك مشكلات متعلقة بإثبات عوامل بنية النسخة المترجمة للأداة، وغالباً عدم وجود معنى للبنية.

فيما يلي نناقش نموذجين من تلك التجارب التي كانت مخيبة للآمال ومحبطة لي ومؤلمة للباحث.

1- أكملت مساعدة في علم النفس في جامعة في البرتغال كانت تسعى للحصول على شهادة الدكتوراة من جامعة في فرنسا، قسماً أساسياً في بحث رسالتها للدكتوراة. كانت قد أنهت كتابة عدة فصول في رسالتها، وكانت تواجه عدة صعوبات في شرح معطيات كثيرة كانت قد عالجتها خلال وجودها في فرنسا كطالبة دكتوراة متفرغة بشكل كامل. كانت تنتظر وصولي لأبدأ مهامي التدريسية في كليتها بفارغ الصبر مع معطياتها الغامضة. لخيبة أملي ولكدرها، لم نأخذ وقتاً طويلاً لاكتشاف أساس المشكلة.

كانت أدوات التقويم التي جرى تقديمها بالطرق التمهيدية لتحليل العوامل خاطئة، لأنه لم يجر تكييفها ثقافياً بشكل جيد وتام. كان التحليل يتضمن جداول منفردة، وذلك أدى الى استخلاص جذور سلبية مستترة. لإقامة تحليل عوامل صحيح لمصفوفة ارتباط، يجب عليه أن يمتلك خاصية كونه "gramian" إن مصفوفة عامل ارتباط هي تلك التي تكون متماثلة والتي تكون جذورها الكامنة المستخلصة كلها إيجابية أو قريبة من الصفر أو صفراً. ويطلق على مصفوفة الارتباط تلك

"إيجابية شبه محددة" أو "إيجابية محددة". وكل مصفوفات الارتباط الأخرى كانت تعرف كمصفوفات ارتباط مفردة ولا يجب تحليلها بأية طريقة من تحليل العوامل. إذا تعرضت مصفوفات ارتباط مفردة الى تحليل عوامل فإنها ستعطي قيمة مستترة ولن يفي التوازن الأساسي بشروط تحليل العوامل (ميريندا، 1997، تاتسوكا، 1971) لذلك فإن نتائج تحليل العامل تكشف عن عدم وجود أية بنية للأدوات. إن حمولة العامل لأكثر البنود كانت حوالي 20s أو أقل. والكثير منها كانت قريباً من الصفر. جعلت تلك النتائج من المستحيل أن أعطيها أية نصيحة عن كيفية تعديل الأدوات وعن إعادة معايرة الدرجات.

2- كانت التجربة تخص إحدى المساعدات التي كانت سابقاً إحدى تلميذاتي، عندما ساعدتها سابقاً بإكمال مشروع بحث اختصاصي بنجاح. وذلك ساعدها في الحصول على مرتبة أعلى. كانت قد التحقت مرة ثانية بأحد صفوفي، وأنهت مشروع بحث آخر وكانت تسعى الى ترقية أخرى. في مرحلة معالجة المعطيات في بحثها، كانت قادرة على شرح الكثير من المعطيات التي سبق تحليلها. ولكنها كانت مرتبكة من نتائج استبيان لثقافة من خارج البرتغال، والذي كان أساسياً في الفرضية الرئيسة للدراسة. كانت تلك المعطيات متعلقة بالإجابات التي تم جمعها بإقامة عملية مسح على أمهات تلاميذ مدارس فعلية.

عندما عدلت الاستبيان لتكييفه حسب الثقافة البرتغالية، اتبعت إرشاداتي السابقة في كيفية ترجمة البند بشكل صحيح وكيفية تغيير بنية إجابة البند من درجات ثنائية القيمة الى فاصل مقياس تقدير بفواصل 5 أو 7 نقاط. لسوء الحظ كانت النتائج غير قابلة للتفسير. خامرني الشك في أن معاملات الارتباط المتبادلة لم تكن صحيحة للاستخدام بتحليل العوامل لعدم حيازتها على المواصفات المطلوبة. كانت، كما في التجربة الأولى، ليست معاملات ارتباط مفردة (ميريندا، 1997، تاتسوكا، 1971). عند معاينة نتائج جداول القيمة المستترة، قدمت كل الجداول قيمة

مستترة سلبية. ماذا سبب تشويه في مصفوفات معاملات الارتباط الجواب موجود في معاينة توزع مقياس إجابة البند ذي النقاط الخمسة أو السبعة. كان بعضهما غير مترابط وكان الآخر غير كامل أو كان مشوهاً بشكل كبير. لم تكن البنود الموجودة في الاستبيان قد جرى تكييفها ثقافياً. أو من الممكن أن الاستبيان الأصلي لم يكن صحيحاً حسب القياس السيكولوجي. لم يكن هناك أي خيار للباحثة إلا التضحية بسنتين من العمل المكثف في محاولة لتكييف الاستبيان والبدء من جديد.

*****

المراجع

Almeida, L. S. (1988). *Oraciocinio differencial dos jovens* [A reasoning differential test for juveniles]. Porto, Portugal: Instituto Nacional de Investigacas Cientifica.

American Educational Research Association, American Psychological Association, and National Council of Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Behling, O., & Law, K. S. (2000). *Translating questionnaires and other research instruments: Problems and solutions.* Thousand Oaks, CA: Sage.

Berry, J. W. (1997). Immigration, acculturation, and adaptation. *Applied Psychology: An International Review, 40,* 5–68.

Brislin, R. W., Lonner, W. J., & Thorndike, R. M. (1973). *Cross-cultural research methods.* New York: Wiley.

Calvi, G. (1970). *II Test G.* Firenze, Italy: Organizzazioni Speciali.

Clarke, W. V. (1956). The construction of an industrial selection personality test. *Journal of Psychology, 41,* 379–394.

Clarke, W. V. (1960). *Examiner's manual for the measurement of skill: A battery of practical placement tests.* Providence, RI: Walter V. Clarke Associates, Inc.

Cronbach, L. J., & Drenth, P. J. D. (Eds.). (1972). *Mental test and cultural adaptations.* The Hague, Netherlands: Mouton.

D'Amico, G., Merenda, P., & Sparacino, R. R. (1982). *Rhode Island Pupil Identification Scale (R.I.P.I.S.): Primo adattamento Italiano.* Palermo, Sicily: Stass.

Flanagan, J. C., Davis, F. B., Dailey, J. T., Shaycroft, M. F., Orr, D. B., Goldberg, I., & Neyman, C. A. (1964). *The American high school student* (Cooperative Research Project No. 635). Pittsburgh: Project Talent Office, University of Pittsburgh.

Geisinger, K. F. (1994). Cross-cultural normative assessment, translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Bulletin, 106,* 304–312.

Gierl, M. J. (2000). Construct equivalence on translated achievement tests. *Canadian Journal of Education, 25*(4), 280–296.

Goodenough, F. L. (1926). *Measurement of intelligence by drawings.* Yonkers, NY: World Book.

Hambleton, R. K., & de Jong, J. (Eds.). (2003). Advances in translating and adapting educational and psychological tests [special issue]. *Language Testing, 20*(2), 127–240.

Hambleton, R. K., & Patsula, L. (1999). Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving test adaptation practices. *Journal of Applied Testing Technology, 1,* 1–13.

Healy, W., & Fernald, G. M. (1911). Tests for practical mental classification. *Psychological Monographs, 13*(Whole No. 54).

Knox, H. A. (1914). A scale based on the work at Ellis Island. *Journal of the American Medical Association, 62,* 741–747.

Kohs, S. C. (1920). The block-design tests. *Journal of Experimental Psychology, 3,* 357–376.

Lecky, P. (1945). *Self-consistency: A theory of personality.* New York: Island Press.
Marston, W. M. (1928). *Emotions of normal people.* New York: Harcourt.
Marston, W. M., King, C. D., & Marston, E. H. (1931). *Integrative psychology.* New York: Harcourt.
Merenda, P. F. (1990a). Present and future issues in psychological testing in the United States. *Evaluacion Psicologia/Psychological Assessment, 6,* 3–31.
Merenda, P. F. (1990b). The Rhode Island Pupil Identification Scale (RIPIS) in cross-cultural perspective. In L. L. Adler (Ed.), *Cross-cultural research at issue.* New York: Academic Press.
Merenda, P. F. (1993). Cross-cultural current and future issues in psychological testing. *International Journal of Group Tensions, 23,* 115–132.
Merenda, P. F. (1994). Cross-cultural testing: borrowing from one culture and applying it to another. In L. L. Adler & E. P. Gielen (Eds.), *Cross-cultural topics in psychology.* Westport, CT: Praeger
Merenda, P. F. (1997). A guide to the proper use of factor analysis in the conduct and reporting of research: Pitfalls to avoid. *Measurement and Evaluation in Counseling and Development, 30,* 156–164.
Merenda, P. F., Clarke, W. V., & Jacobsen, G. (1965). Relative predictive validities of the MOS and DAT batteries for junior high school students. *Psychological Reports, 16,* 151–154.
Merenda, P. F., & DiLeonardo, C. (1992). The American adaptation of the Bedini Test of Distributive Attention. In A. L. Comunian & U. P. Gielen (Eds.), *Advancing psychology and its applications: International perspectives.* Milan, Italy: Franco Angeli.
Merenda, P. F., Hall, C. E., Clarke, W. V., & Pascale, A. (1962). Relative predictive effeciency of the DAT and a short battery of tests. *Psychological Reports, 11,* 71–81.
Merenda, P. F., Jacobsen, G., & Clarke, W. V. (1966). Cross-validities of the MOS and DAT batteries. *Psychological Reports, 19,* 341–342.
Merenda, P. F., Maio, E., Guadagnoli, E., & Yu-Wen, H. (1984). *The Chinese form of the Rhode Island Pupil Identification Scale: Standardization and validation.* Providence, RI: AVA Publications.
Merenda, P. F., & Migliorino, G. (1974). Student information blank: Comparison between Sicilian and American responses. *Annali della facolta di Economia e Commercio, Dell' Universita' "di Palermo, 28,* 385–403.
Migliorino, G. (1947). Ricerca sulla strattura psicologica del reattivo di Bedini. *Rivista di Psicologia, 43,* 154–171.
Novack, H. S., Bonaventura, E., & Merenda, P. F. (1979). *Manual to accompany Rhode Island Pupil Identification Scale: A behavior observation identification scale: A behavior observation scale for the early detection of children with learning problems.* Providence, RI, Author. (Origianl work published 1972)
Olmedo, E. L. (1979). Acculturation: A psychometric perspective. *American Psychologist, 34,* 1061–1070.
Pintner, R., & Paterson, D. G. (1917). *A scale of performance tests.* New York: Appleton.
Poortinga, Y. (1995). Use of tests across cultures. In T. Oakland & R. K. Hambleton (Eds.), *International perspectives on academic assessment* (pp. 187–206). Boston: Kluwer.
Porteus, S. D. (1915). Mental tests for the feeble-minded: A new series. *Journal of Psycho-Asthenics, 19,* 200–213.

Raven, J. C. (1938). *Progressive matrices: A perceptual test of intelligence.* London: Lewis.

Rulon, P. J. (1953). *A semantic test of intelligence. Proceedings of 1952 Invitational Conference on Testing Problems.* Princeton, NJ: Educational Testing Service.

Sperber, A. D., Develles, R. F., & Boehlecke, B. (1994). Cross-cultural translation. *Journal of Cross-Cultural Psychology, 25,* 501–524.

Sprini, G., Cardica, M., & Gangemi, A. (1992). *The Italian form of the RIPIS scale: A format modifying proposal.* Palermo, Sicily, Italy: University of Palermo.

Stouffer, S. A., Guttman, L., Suchman, E. A., Lazarsfeld, P. F., Star, S. A., & Clausen, J. A. (1950). *Measurement and prediction.* Princeton, NJ: Educational Testing Service.

Tatsuoka, M. M. (1971). *Multivariate analysis: A technique for educational and psychological research.* New York: Wiley.

van de Vijver, F., & Leung, K. (1997). *Methods and data analysis for cross-cultural research.* Thousand Oaks, CA: Sage.

Woodworth, R. S. (1910). Race differences in mental tests. *Science N. S., 31,* 171–176.