2

مسائل مفاهيمية ومنهجية في تكييف الاختبارات

فون ج. ر. فان دي فيغر ويب هـ. بورتينغا
جامعة تيلبرغ، هولندا

لنفترض أن عالماً نفسياً قرر عمل نسخة باللغة الألمانية من أحد اختبارات الذكاء الأميركية الذي يختبر المعلومات العامة الذي يتضمن البند التالي: من هو رئيس الولايات المتحدة، قد يكون بند هذا السؤال جيداً لقياس سيكولوجية عينة من الأميركيين. قد يقرر المختص الألماني استخدام ترجمة حرفية لهذا السؤال: من هو رئيس الولايات الألمانية؟ إن السؤال بصيغته الألمانية هو سؤال واضح كمثيله في اللغة الإنجليزية وأن ترجمة راجعة لهذا السؤال ستعطي النص الأصلي الإنجليزي. يضاف إلى ذلك أن الرئيس الأميركي معروف بشكل أفضل من قبل المواطنين الأميركيين، ولكن من المؤكد أنه معروف أيضاً في أوروبا. عند استخدام هذا الاختبار من المرجح أن تكون الإجابة عن هذا السؤال أقل صعوبة في الولايات المتحدة منها في هولندا. وإن هذا الفارق هو انعكاس لاختلاف المعرفة بين الشعبين، فعدد الأميركيين الذين يعرفون اسم رئيس الولايات أكثر من عدد الهولنديين، فإذا افترضنا أن هذا الاختبار هو وحيد البعد فمن المحتمل جداً أن يقيس هذا السؤال البناء العقلي الأساسي بشكل مناسب في كلا الدولتين. والنتيجة التي لا بد أن نصل إليها هي أن السؤال مفيد لأن النصين اللغويين للسؤال متماثلان تماماً ويقيس البنية العقلية الأساسية في كلتا الدولتين.

ومع ذلك فإن الاستنتاج بأن هذا السؤال صحيح ومفيد بسبب التماثل اللغوي بين النصين غير واضح، إذ إن هناك مشكلة عدم التماثل السيكولوجي بينهما: هل لهذا السؤال المعنى نفسه في كلا الدولتين. فالسؤال المماثل سيكولوجياً في هولندا قد يكون السؤال عن اسم الملكة أو رئيس الوزراء أو إحدى الشخصيات العامة الأخرى. إن تكييف النص الأصلي يفقده التماثل اللغوي إلا أنه يزيد من التماثل السيكولوجي. مع الأسف أن التكييف الذي يؤكد على التماثل السيكولوجي له مشكلاته أيضاً.

أول هذه المشكلات هي أن هذا التماثل يجب إقامته بالتجربة ولا يمكن الاكتفاء بافتراض وجوده، يضاف إلى ذلك أن طرح الأسئلة متباينة الصياغة وبلغات مختلفة سيقلل من إمكانية مقارنة النتائج بين الشعوب، مثال ذلك ما هي النتيجة التي نتوصل إليها إذا تبين أن 90٪ من الكهول في عينة الولايات المتحدة يعرفون اسم الرئيس، بينما 95٪ من الهولنديين يعرفون اسم الملكة. وإذا كانت العينة المدروسة واسعة فإن هذا الاختلاف قد يكون مهماً إحصائياً، إلا أن تفسير هذا الاختلاف قد يكون صعباً إن لم يكن مستحيلاً. فقد يعكس ذلك مثلاً اختلاف ظروف هؤلاء الأشخاص في وسائل الإعلام في بلديهما، بدل أن يكون سببه الاختلاف في بنية الاختبار أو في المعلومات العامة.

قد تكفي هذه المقدمة القصيرة للتأكيد على الموضوعات الرئيسة التي يتضمنها هذا الفصل، أولاً قد يتلاقى المنظور السيكولوجي واللغوي في أثناء الترجمة، وفي هذه الحالة تكون الترجمة دقيقة ومباشرة، لكنها قد تؤدي إلى صياغة مختلفة في اللغة المستهدفة. لكي نحصل على ترجمة جيدة للنص الأصلي يجب توافر خبرة جيدة باللغة الأصلية واللغة المستهدفة إلا أن ذلك غير كاف إذ لا بد من توافر منظور سيكولوجي حتى يمكن التوصل إلى أدوات (اختبارات) جيدة في اللغة المستهدفة (بيلينغ ولو، 2000، براكون وبارونا، 1991، برسلين، 1980، 1986،

كيسينجر، 1994؛ هامبلتون، 1994؛ جـ. هاركنس، 1998؛ ميريندا، 1994، فاليراند، 1989، فان دي فيجر، 2003).

ثانياً: هناك حاجة إلى إطار نظري نستطيع أن نحدد ضمنه بشكل دقيق ماذا نعني بتعبير "التماثل السيكولوجي".

نعرف في القسم الأول من هذا الفصل مصطلحين رئيسين يُستخدمان في تحديد التماثل وهما: الانحياز والتكافؤ. ونطبق في القسم الثاني هذين المصطلحين على الترجمة/ التكييف. ونعالج في القسم الثالث الطرق التي تعزز ملاءمة الترجمة/ التكييف. ونصف في القسم الأخير انعكاسات ذلك على استخدام الاختبارات.

يمكن اعتبار هذا الفصل كخلفية نظرية للإرشادات الخاصة في تكييف الأدوات التربوية والسيكولوجية وإحداث تكافؤ في الدرجات الموجودة في الجدول 3.1 في الفصل الأول. انطلاقاً من هذه الخلفية يصبح من السهل إدراك الأساس المنطقي للإرشادات وفهم المواقف التي تم تبنيها. وقد وردت الإشارة إلى الإرشادات كلما كان ذلك مفيداً.

الانحياز والتكافؤ

يختلف معنى كل من مصطلحي الانحياز والتكافؤ قليلاً في الأدبيات. تتضمن كلمة الانحياز وجود عوامل مزعجة (مؤذية)، يقال عن القياس إنه منحاز إذا اختلفت الدرجات المسجلة في أحد الاختبارات من لغة إلى أخرى بسبب وجود تباين غير مرغوب فيه. مثال: إذا أخذنا الترجمة السويدية - الإنجليزية في موضع الاختبار التالي: أين يحتمل أن تعيش الطيور ذات الوترات في أرجلها فإن الترجمة الراجعة من السويدية إلى الإنجليزية لعبارة (طير ذو وترات في أرجله) هي الطير ذو الأرجل السباحة هذا يعطي إشارة لمعرفة الإجابة تفوق ما يتضمنه النص الإنجليزي الأصلي (هامبلتون، 1994). مثال آخر في عمليات المسح الذي يتناول القيم في أوروبا، كانت

الدرجات للإخلاص والوفاء في إسبانيا مختلفة عن مجمل النتائج المسجلة في تلك الدولة. لدى التدقيق في الأمر تبين أن كلمة الوفاء التي استعملت في الترجمة إلى الإسبانية تضمن معنى الإخلاص الزوجي (هالمان، حزيران 1998).

يرتبط مصطلح التكافؤ ذهنياً بقياس أوجه الخلاف بين الأعراف، وما ينتج عن ذلك الانحياز. إن البند أو الأداة المنحازة ستعطي درجات غير متكافئة.

أصبح عدد التكافؤ تعبيراً عاماً يشير إلى عدم إمكان المقارنة بين الدرجات، وعملاً بهذا العرف فإننا نستعمل عدم التكافؤ كصفة مميزة لدرجات الاختبار التي تأثرت بفعل الانحياز الثقافي.

من وجهة النظر التي اتخذناها في هذا الفصل لا يعد الانحياز صفة جوهرية للاختبار وإنما هي نتيجة لتطبيق هذا الاختبار على مجموعة خاصة بغية الوصول إلى هدف محدد. وهي تشير إلى جميع أنواع العوامل المزعجة التي تعيق تفسير اختلاف الدرجات بين مجموعة وأخرى. يمكن فهم الانحياز بشكل أفضل من خلال القابلية للتعميم.

يمكن تعريف الانحياز بأنه التوافق غير المتوازن في مجال (ميدان) الملاحظة ودعائم التعميم. مثال على ذلك، لنفترض أننا طبقنا اختبار الامتداد الزمني (وهو الاختبار الذي يقيس اتساع الذاكرة القصيرة الأمد) على أطفال أميركيين وأطفال ريفيين من إفريقيا لم يحصلوا على تعليم مدرسة جيد ووجدنا أن الأطفال الأميركيين قد حصلوا على درجات أعلى. إذا فسّرت درجات الاختبار استناداً إلى عدد الأرقام التي يستطيع أطفال الفريقين الاحتفاظ بها في ذاكرتهم لمدة قصيرة، يمكن القول إن الاختبار غير منحاز؛ ذلك أن استخدام الأرقام في بعض المجالات مثل علم الحساب قد يعطي نتائج متماثلة الاختلاف. أما إذا اعتبرت علامات الامتداد الرقمي على أنها دليل على قدرات الذاكرة قصيرة الأمد (وهو التفسير الشائع لهذه الدرجات) فإن الاختبار يكون منحازاً على الأرجح. هناك إثبات في علم

النفس الشامل للثقافات، وهو أن اتساع الذاكرة قصيرة الأمد يختلف باختلاف الثقافات والأعراف (واغنر، 1981). ومع ذلك فإن اختلاف الحوافز بين الثقافات قد يكون له تأثير شامل على الدرجات في كثير من الاختبارات. إن الاختلافات التي شوهدت في اختبار الامتداد الرقمي قد لا تتكرر في الاختبارات التي تستخدم حوافز أكثر صحة من الناحية البيئية بالنسبة للأطفال الأفارقة الريفيين. لدى دراسة الانحياز سنحاول معرفة ما هي التبدلات التي تؤثر في أحداث الاختلافات المشاهدة بين الأعراق.

ثلاثة أنواع من الانحياز:

على الرغم من أن الانحياز قد ينشأ من مصادر عديدة فإن من الضروري التمييز بين عدة فئات. كثيراً ما تؤدي المصادر المختلفة إلى أنواع متماثلة من الانحياز. في اعتقادنا يوجد ثلاث فئات من الانحياز: انحياز البنية، انحياز المنهج وانحياز الموضوع (فان دي فيجر ولونغ، 1997 a، 1997 b، 2000؛ فان دي فيجر وبورتينغا، 1997؛ فان دي فيجر وتانزر 1997) (انظر الجدول 2.1).

الجدول 2.1

أنماط الانحياز

نمط الانحياز	الوصف
انحياز البنية	تداخل غير تام في البنية لدى الفئات الثقافية.
انحياز المنهج	مصطلح عام يشمل جميع عوامل الخلل الناتج عن أحد جوانب المنهج.
انحياز الأداة	أشكال الأداة التي تسبب اختلافات الدرجات في الإثنيات والتي لا علاقة لها بالبنية.
انحياز إداري	قصور التواصل بين منفذي الاختبار والخاضعين له.
انحياز الموضوع/ اختلاف عمل الموضوع	اضطراب الموضوع (مثل الترجمة السيئة).

انحياز البنية: يقصد بهذا الشكل من الانحياز التباين في البنيات الثقافية بين المجموعات المختلفة. يبين الجدول 2-2 نظرة إجمالية لأهم مصادر انحياز البنية. مثال ذلك أن العوامل التي تشكل البنية (مثل السلوك، المواقف، القواعد السلوكية) ليست متطابقة تماماً بين المجموعات المختلفة. إن أنصار نسبية المواقف التي تشاهد السيكولوجيات المحلية (اسينها، 1997) والسيكولوجيات الثقافية (كول، 1996، غرين فيلد، 1997 a، 1997 b، ميلر، 1997) يميلون للاعتقاد (فيما يتعلق بهذا الفصل) أن انحياز البنية هو القاعدة وليس الاستثناء في السيكولوجا عبر الثقافات.

الجدول 2-2

مصادر نموذجية لثلاثة أنماط من التحيز في تقويم عبر اللغات

نمط الانحياز	مصادر الانحياز
• انحياز البنية	• عدم التماثل في تعريف البنية عبر الثقافات.
	• تفاوت ملاءمة السلوك المرافق للبنية (مثال: مهارات ليس لها ذخيرة في ثقافة المجموعات).
• انحياز المنهج	• عدم إمكانية مقارنة العينات (مثال: يحدث هذا بسبب الاختلافات في التعليم، الدوافع).
	• تفاوت الألفة مع المواد المنبهة.
	• تفاوت الألفة مع إجراءات الاستجابة.
	• تفاوت طرق الاستجابة (مثال: الرغبات الاجتماعية، الدرجات القصوى، القبول).
	• تفاوت الظروف المحيطة بإجراء الاختبار، فيزيائية (مسجلات) أو اجتماعية (عدد الطلاب في الصف).
• انحياز البند	• ترجمة سيئة للبند/ بنود غامضة.
	• عوامل سيئة ذات علاقة بالبند.

ملاحظة: انظر (فان دي فيجر وتاتزر ١٩٩٧).

من الأمثلة على ذلك مفهوم الذكاء. تميل معظم اختبارات الذكاء إلى استخدام تعريف ضمني للذكاء يتألف من المحاكمة والتفكير المنطقي (كما هو الحال في اختبارات رافن) وبدرجة أقل من المعارف المكتسبة والذاكرة (كما هو الحال في مجموعات الذكاء مثل سلم فكسلر للذكاء عند الأطفال وسلم فكسلر للذكاء عند البالغين). نشاهد هذه العوامل أيضاً عندما يطلب من الأفراد بيان مميزات الشخص الذكي (ستينبرغ، كونوي، كيرتن وبيرستاين، 1981). إلا أن الدراسات في الأوساط غير الغربية بينت أن المفهوم الشائع للذكاء أوسع من ذلك ويتضمن بعض المظاهر الاجتماعية. مثال ذلك ما قالته أمهات كوكوه في كينيا من أن الطفل الذي يعرف مكانه في العائلة والسلوك هو الذي يتوقع منه مثلاً اتباع الطرق المناسبة في مخاطبة الآخرين، وأن الطفل الذكي هو الطفل المطيع الذي لا يسبب مشكلات (مندي - كاستل، 1974، سيجال، داسن، بيري وبورتينغا، 1987). وقد بينت دراسات مماثلة في زامبيا (سيربل، 1993) واليابان (أزوما وكاشيواجي، 1987) أن صفات الشخص الذكي تتجاوز ميدان المدرسة الذي يُعتمد عادة في الولايات المتحدة وأوروبا. ومثال آخر على الاختلاف في مضمون الذكاء يمكن أن يشاهد في أعمال (هو، 1996) حول طاعة الوالدين في الصين. وقد بين أنه بالمقارنة مع الغرب يميل الصينيون إلى تطبيق تعريف أوسع للذكاء. وأن الطاعة واحترام الأبوين هي عناصر تشاهد أيضاً في الدول الغربية، إلا أن المفهوم الصيني لاحترام الأبوين يشمل أيضاً العناية المادية بالأبوين عندما يتقدمان في السن ويحتاجان للمساعدة.

إن مشكلة اختيار العينات السيئ يضاعف عندما يتم العمل بوسائل مختصرة كما هي الحال في أكثر الأحيان. وقد شكا ترياندیس (1978) قبل عدة سنوات من أن قياساتنا تتناول عينات قليلة من اهتمام البشر. ويقود هذا إلى ما يدعوه أمبرسون (1983) (نقص تمثيل البنية) إذ يتم اختيار عدد قليل من الموضوعات بسب تجانسها وتعد أنها تغطي بنيات واسعة. وعلى الرغم من أن هذه القضية ليست خاصة بميدان السيكولوجيا عبر الثقافات إلا أنها تبرز في هذا المجال بشكل

خاص. وعندما تُبدي مقارنة المجموعات الثقافية اختلافات واسعة، فإن القياسات الضيقة تبدي على الأرجح انحيازاً يعود إلى محدودية تمثيل البنية.

انحياز المنهج:

انحياز المنهج هو مصطلح عام لنوع ثان من الانحياز الذي يتضمن كل المتحولات المزعجة الناتجة عن عوامل خاصة بالمنهج وقد ابتكر هذا المصطلح لأن هذه العوامل تذكر عادة في القسم الخاص بالمنهج في الدراسات التجريبية (الإمبريقية).

هناك نوعان من انحياز المنهج. الأول هو (انحياز الأداة). يتضمن ذلك كل خصائص الأداة التي لا علاقة لها بهدف الدراسة، إلا أنها مع ذلك تسبب اختلافات في درجات الاختبار. إن السبب الأكثر شيوعاً لانحياز الأداة في الاختبارات الذهنية هو تآلف الأشخاص مع المنبهات والاستجابات (أو شكل الإجابة). يمكن العثور على إيضاح لذلك في دراسة قام بها سيربل (1979).

اهتم هذا الباحث بالمهارات الإدراكية/ الحسية عند الأطفال البريطانيين والزامبيين. وقد طلب منهم استنساخ صور أشخاص باستخدام الورق وقلم الرصاص، أوضاع اليد، والأسلاك المعدنية (التي هي شائعة في زامبيا). كما هو متوقع فقد تفوق الأطفال البريطانيون في الرسوم المعمولة بالورق وقلم الرصاص، بينما حصل الزامبيون على درجات أعلى بشكل واضح في الرسوم المعمولة بالأسلاك المعدنية، وكما هو متوقع أيضاً لم تسجل اختلافات باستعمال الوسائل الأخرى. وتفسير هذه المشاهدات باختصار على أنها ناجمة عن اختلاف التآلف مع الأجوبة.

يمكن للبيانات الخاصة بالشخصية والموقف أن تُبدي انحياز الأداة. مثال ذلك ما وجده (هوي وتريانديس 1989) من أن المكسيكيين كثيراً ما يحصلون على علامات مفرطة بالمقارنة مع الأميركيين من أصل أوروبي عند استعمال سلم مؤلف من 5 نقاط. ولا يشاهد هذا الانحياز عند استخدام سلم مؤلف من 5 نقاط تشير الفئة الثانية من انحياز المنهج إلى "الانحياز الإداري" ويقصد به اختلاف الدرجات

الناجم عن التعليمات وغير ذلك من مشكلات التواصل بين المختَبرِين والمختَبَرين. تحدث هذه المشكلات بشكل خاص عندما يستخدم المختبرون أو المختبرين لغة غير لغتهم الأم. قد يكون فقد المعلومات ناجماً عن عدم القدرة على التعبير عن الأفكار بلغة ثانية (غاس وفارونيس، 1991). كما أن عدم معرفة ثقافة المختبرين قد يؤي إلى انتهاك القواعد المحلية للمجاملة.

وقد عالج المختصون بعلم النفس والباحثون في المسوحات تأثير مميزات الشخص الذي يُجري المقابلة (كالجنس، العمر، العرق) على نتائج القياس. في مراجعة للدراسات السيكولوجية الخاصة بجنس المختبر على أداء الأطفال في اختبارات الذكاء استنتج "جنسن" (1980) أن الدراسات ذات المنهجية المناسبة قليلة. (مثال: لا توجد دراسات تعالج قضية عرق المختبر (المُختبر). مع ذلك إن البينات المتوافرة لا تشير إلى أن عرق المختبر ذو أهمية كبيرة. وقد درس الباحثون في المسوحات ما يُدعى بنظرية الإذعان (الاحترام). فقد وجد كوتر، كوهين وكولتر (1982) أن الأشخاص أكثر ميلاً لإبداء مواقف إيجابية تجاه جماعة ذات ثقافة معينة عندما تجري مقابلتهم من قبل أحد أفراد تلك الجماعة (رييس، دانيلس، شوميكر، شانك وهو، 1986). مع ذلك فإن أهمية التأثيرات العائدة لصفات الباحث الذي يجري المقابلة تبدو قليلة ومتضاربة في مختلف الدراسات (سنجر وبريس، 1989).

قد يكون لانحياز الأداة تأثير شامل على درجات الاختبار، عن تأثير الاختلافات في التعليم أو تعرض المختبر لاختبارات سابقة قد يؤثر على الدرجات في بعض الموضوعات ذات العلاقة بالمدرسة. من المرجح أن هذه الاختلافات ستؤثر على معظم الموضوعات أو جميعها، وهكذا فإن انحياز المنهج قد تكون نتيجة لمقارنة قياسات نفسية، إذا عممت هذه الدرجات على الأفراد خارج نطاق المدرسة. إذاً قد ينتج عن انحياز المنهج اختلاف في الدرجات داخل المجموعات نفسها لا علاقة له بالبنية وإنما يعود إلى خداع القياسات. إن المعنى المتضمن لهذا الأمر خطير

فالباحث أو الممارس الذي يقارن الدرجات عبر الثقافات (سواء تم ذلك بشكل صريح عن طريق اختبار هذه الاختلافات إحصائياً، أو بشكل ضمني عن طريق تطبيق جدول معياري للفرد في الثقافات المختلفة) سوف يحتاج للاختيار بين تفسيرين متنافسين هما: الاختلافات المشروعة عبر الثقافات وانحياز المنهج. وغالباً ما يصعب الاختيار بين هذين التفسيرين بسبب نقص البينات التي تؤكد أو تنفي أياً منهما، وهكذا فإنه من المهم بالنسبة لمصممي الاختبارات أن يعترفوا بأهمية انحياز المنهج ويحاولوا التقليل من تأثيره إلى أبعد حد ممكن (هاميلتون، 1994؛ انظر أيضاً سيرغي، باتسولا، هامبلتون، الفصل الرابع في هذا الكتاب، وتهدف جميع التوجيهات الإدارية إلى تحقيق ذلك).

انحياز البند (اختلاف عمل البند)

يشير هذا النوع من الانحياز إلى الأخطار التي تؤثر على صحة البنود فقط، بينما يتناول انحياز البنية والمنهج المظاهر العامة للاختبار. في البدء استخدم تعبير انحياز البند (كليري وهيلتون، 1968). بعد أكثر من ثلاثة عقود من التطور المهم في القياسات النفسية لاكتشاف الموضوعات غير السوية (اكريكان، 2002؛ هولاند وووينر، 1993؛ ميلساب وافرسون، 1993؛ سيرغي وآلالوف، 2003)، تم استبدال هذا المصطلح "اختلاف عمل البند"؛ كان هناك شعور بأن انحياز البند يتضمن معنى الابتعاد عن المعيار الأوروبي/ الأميركي الذي كان وما يزال أكثر مجموعة معيارية مستخدمة في الأبحاث في الولايات المتحدة. إلا أننا نتمسك بالمصطلح الأصلي لأنه يؤكد على العلاقة الوثيقة مع الأنماط الأخرى من الانحياز ويشير إلى مظاهرها الأساسية: إنه تهديد لصحة البند ويمنع المقارنة المباشرة للدرجات.

إن أهم أسباب انحياز البند هو الترجمة السيئة واختلاف المعاني الضمنية للكلمات. مثال ذلك أنه استناداً إلى قاموس "وبستر" الأميركي فإن العدوانية "تتجلى على شكل تصميم واضح واستعداد للخصام" بينما قاموس اوكسفورد البريطاني

يعطي المعنى الأول للعدوانية بأنه "عمل أو ممارسة الهجوم دون تربص وبشكل خاص البدء بالحرب أو النزاع". يأتي المعنى السابق في القاموس الأميركي في المرتبة الثالثة. من المهم الإشارة إلى أن العدوانية في اللغات الأخرى كالألمانية والفرنسية والهولندية وغيرهم من اللغات هي أقرب إلى معناها من التعريف البريطاني وليس الأمريكي.

هناك مجموعة مذهلة من التعاريف والتقنيات الإحصائية التي تم اقتراحها لتعريف انحياز البند، إلا أن هناك في الوقت الحاضر ميلاً للتقارب في هذا المجال أكثر مما كان من قبل عشر أو عشرين عاماً. وقد أصبح أحد التعاريف ومجموعة صغيرة من التقنيات الإحصائية مقبولة على أنها الأكثر ملاءمة. ومن المهم في التعاريف المتداولة أن انحياز البند مشروط بمستوى الاستعداد أو السمة. ماذا نعني بذلك؟ لنعد إلى المثال المتعلق برئيس الولايات المتحدة، إن الدرجات العالية التي سجلها الطلاب الأميركيون في هذا البند تعكس اختلافاً حقيقياً بين المواطنين الأميركيين والهولنديين، ولا نريد أن نتجاهل هذا الاختلاف على أنه انحياز أو على أنه "تأثير معاكس" إذا استخدمنا مصطلحاً آخر يستخدم للتعبير عن الاختلافات غير الصحيحة بين الثقافات.

لتحليل مصطلح انحياز البند نقوم بتقسيم كل من العينتين إلى عدد من المجموعات بحسب الدرجات (تدعى هذه العملية التكييف الشرطي) Conditioning في المجموعة الأولى للأميركيين نضع جميع الأشخاص الذين حصلوا على درجة (1) في الاختبار، وكذلك الأمر بالنسبة لمجموعة الهولنديين التي تضم جميع الأشخاص الذين حصلوا على درجة (1). وتضم المجموعة التالية جميع الأشخاص الذين حصلوا على درجة (2)، ونقوم بالشيء نفسه بالنسبة لبقية الدرجات، يسمح لنا هذا الإجراء القيام بتحليل مفصل؛ فبدلاً من مقارنة متوسط الدرجات يصبح بإمكاننا مقارنة الأميركيين والهولنديين الذين حصلوا على المستوى نفسه من الدرجات (أي أننا نجري المقارنة المشروطة بمستوى الدرجات). يبين الاختبار الخاص برئيس الولايات المتحدة وجود انحياز البند إذا كان الأميركيون والهولنديون الذين لديهم درجة متماثلة في مجموع الاختبارات لم يحصلوا على المستوى نفسه من الدرجات في هذا

البند. وبصورة أدق نقول بوجود انحياز البند إذا كان الأشخاص الذين لديهم المستوى نفسه في اختبار البنية (وهم الأشخاص الذين لديهم درجة مماثلة في مجمل الاختبارات) لم يحصلوا على الدرجة المتوقعة نفسها في هذا البند (ويقصد بذلك متوسط الدرجة في هذا البند) (هولاند وثاير، 1988، 1997 b، 2000).

لقد وصف "ميلينبرغ" (1982، انظر أيضاً غلوستر ومازور، 1998) الفرق بين الانحياز المتساوق وغير المتساوق. يقال: إن البند منحاز بشكل متساوق إذا كان الاختلاف في مستوى أداء ثابتاً في جميع المستويات تقريباً (يقصد بذلك أنه في كل مجموعة من الدرجات يكون أداء الأميركيين المختبرين أفضل من أداء الهولنديين بالمقدار نفسه تقريباً). ويقال بوجود انحياز غير متسق إذا كان حجم الفارق يختلف بشكل مضطرد بين المجموعات (مجموعات الدرجات). مثال ذلك في المجموعات التي حصلت على درجات منخفضة كان عدد الأميركيين الذين يعرفون اسم رئيسهم أقل من عدد الهولنديين الذين يعرفون اسم الملكة، إلا أن الفارق يتضاءل تدريجياً عند المجموعات ذات الدرجات الأولى.

مستويات التكافؤ:

يتحدى الانحياز مقارنة الدرجات التي يتم الحصول عليها في المجموعات المختلفة. من الناحية التقنية يحدد الانحياز تكافؤ الدرجات. ولكي نحدد نتائج الانحياز عند مقارنة الدرجات نبين هنا أربعة أنماط من التكافؤ مرتبة بحسب صلاحيتها للمقارنة حسب الإثنيات (فان دي فيجر ولونغ 1997 a، 1997 b، 2000؛ فان دي فيج وبورتينغا، 1997، انظر الجدول 2-3). أول هذه الأنماط يدعى عدم تكافؤ البنية. يتميز هذا النمط بانعدام القابلية التامة للمقارنة، كما لو قارنا البرتقال مع التفاح". هذا النمط من التكافؤ هو نتيجة انحياز البنية. يستحيل إجراء مقارنة الدرجات عبر الثقافات استناداً إلى عمليات/ فعاليات غير تامة أو غير ملائمة.

الجدول 2-3

أنماط التكافؤ

نوع التكافؤ	الوصف
تكافؤ البنية	تقيس الأداة بنيات مختلفة في ثقافتين مختلفتين (مقارنة التفاح مع البرتقال).
تكافؤ البناء/ التكافؤ الوظيفي	تقيس الأداة ذات البناء السيكولوجي عبر مجموعات ثقافية مختلفة.
تكافؤ وحدة القياس	تشتمل الأداة على ذات وحدة القياس وأصل مختلف عبر المجموعات الثقافية المختلفة.
تكافؤ المقياس/ تكافؤ الدرجات التام	تشتمل الأداة على ذات وحدة القياس وذات الأصل عبر المجموعات الثقافية المختلفة.

يُعرف النمط الثاني من التكافؤ بأسماء مختلفة، أكثرها شيوعاً هو التكافؤ البنيوي والتكافؤ الوظيفي. يترافق هذا النمط بصنف من الإجراءات التي تُستخدم لإقامة تطابق في البنيات بين المجموعات كما هو معمول به في أحد اختبارات القياس النوعية. بشكل عام يتطلب هذا الشكل من التكافؤ أن تكون نماذج العلاقات بين المتغيرات هي نفسها في كل واحدة من المجموعات. يُستخدم هذا النمط من التكافؤ في العديد من مشروعات تكييف الاختبارات (الارشادات 2).

عندما يجري تهيئة ترجمة اختبار ما إلى لغة أخرى تظهر مسألة صدق البنية: هل يقيس النص الأصلي والنص المترجم للاختبار البنية النفسية ذاتها؟ يكون هذا التساؤل مهماً عندما يُترجم الاختبار حرفياً. في مثل هذه الحالات غالباً ما يستخدم تحليل العامل المؤكد أو تحليل العامل الاستكشافي متبوعاً بتبادل الهدف لدراسة تماثل عوامل البند عبر ثقافات سكانية (لمزيد من التفاصيل انظر: ايرن،

شافلسون وماثن، 1989؛ ليتل، 1997؛ فان دي فيج ولونغ، 1997 a، 1997 b، 2000؛ واتكينز، 1989). إن تماثل محتويات العامل في كل بند ينظر إليه كشرط أساسي للتكافؤ البنيوي، وينطبق هذا الأمر على الاختبارات المترجمة حرفياً. إذا حدثت تغييرات في اللغة المترجم إليها خلال عملية الترجمة من أجل الحصول على بنية مناسبة (أي عندما يتم إجراء أي تعديل) يجب أن تظهر العوامل السابقة نفسها، إلا أنه لا يمكن توقع حصول تطابق في كل واحد من البنود مع الآخر، على سبيل المثال هناك أكثر من أربعين ترجمة لكل من سبيلبرغر، غوسج، ولوشينيس "الاختبارات الشخصية لسمات القلق" (STAI) لم تكن الغاية المهمة من أكثر هذه الترجمات إنتاج ترجمة حرفية للنص الإنجليزي وإنما إعداد اختبار أداة قادرة على تقييم القلق بشكل مناسب في الثقافة التي يجري الترجمة إليها. إذا أخذنا هذا الهدف بعين الاعتبار يصبح العالم الخاص بتقنيات التحليل أقل ملاءمة ويصبح اختبار شبكة النواميس المهيمنة على حقل البند أكثر أهمية (كرونباخ وميهل، 1955). باختبار هذا الإجراء النموذج المتوقع للروابط العالية مع المقاييس الأخرى للقلق المتوافرة سلفاً في اللغة الهدف، كما يقوم باختبار الروابط القليلة أو المعدومة مع المقاييس الخاصة بالبنية غير ذات الصلة بالموضوع.

أما المستوى التالي من التكافؤ فيدعى "قياس وحدات التكافؤ"، يتطلب هذا المستوى أن يكون المقياس في كل مجموعة له ذات المسافة (أي أن القياسات هي نفسها على مختلف المستويات). إن مقياسين يبديان وحدات قياسية متكافئة عندما يكون لها ذات الوحدات القياسية وكلها ذات مصدر مختلف. وهذا هو أدنى مستوى من التكافؤ وفيه يمكن مقارنة مستوى الدرجات بشكل صحيح ولو كان ذلك مع بعض القيود، فالفوارق الإفرادية التي تُشاهد في المجموعة A يمكن مقارنتها بالفوارق الإفرادية في المجموعة B، مثال ذلك إذا حقق اختبار انبساط الشخصية هذا النمط من التكافؤ أمكن الإجابة عن التساؤل عما إذا كانت اختلافات الجنس (ذكر أو أنثى) في ذلك الاختبار متماثلة عبر المجموعات الثقافية المختلفة، ومع ذلك نظراً لاحتمال

وجود اختلاف في مصدر المقياس لا يمكن إجراء مقارنات بين الدرجات عبر المجموعات. وهكذا يصبح من المستحيل القول بأن إحدى المجموعات تحصل على مستوى أعلى من غيرها في هذا الاختبار (اختبار انبساط الشخصية) أو أن الشخص X من المجموعة A هو أكثر انبساطاً من الشخص Y من المجموعة B عندما تكون وحدات المقياس متكافئة.

إن الاختلافات الأساسية في مستوى الدرجات يمكن دراسته فقط عندما تبدي الدرجات تكافؤ درجات أعلى "تكافؤ المقياس/ تكافؤ الدرجات التام".

إن المقياسات التي تُبدي هذا النمط من التكافؤ تملك ذات وحدات القياس وذات المصدر في كل المجموعات. في هذا النمط من التكافؤ تستطيع الدرجات تجاوز الحدود الثقافية دون أي مشكلات كما يمكن مقارنتها بشكل سليم لدى أشخاص من ثقافات وأعراق مختلفة.

التكافؤ والانحياز:

إن التكافؤ والانحياز مرتبطان بشكل وثيق، حتى إن بعض المؤلفين يستخدمون هذين المُصطلحين للدلالة على المعنى ذاته، بعد التعرف على هذه المصطلحات أصبح من الممكن دراسة العلاقات بينها بشكل مفصل. قد يُخفض الانحياز أو لا يُخفض مستوى التكافؤ في المقارنات كما هو مبين في الجدول 2-4 وسيُحدد مستوى التكافؤ نمط المقارنة الذي يمكن اتخاذه (الإرشادات 21).

الجدول 2-4

تأثير الانحياز على مستوى التكافؤ

	مستوى التكافؤ		
	البنية	بند القياس (أ)	المقياس (ب)
- نمط الانحياز	نعم	نعم	نعم
- انحياز البنية	لا	لا	نعم
- انحياز المنهج المتسق	لا	نعم	نعم
- انحياز المنهج غير المتسق	لا	لا	نعم
- انحياز البند المتسق	لا	نعم	نعم
- انحياز البند غير المتسق			

إن مقارنة الدرجات المباشر (مثل أفراد المجموعة A أكثر انبساطاً في المتوسط من أفراد المجموعة B) يتطلب مستوى أعلى من تكافؤ البنية (مثال: هل يعد الانبساط على أنه بنية متماثلة في أفراد المجموعتين A وB) إن انحياز البنية هو أهم التحديات التي تواجه مقارنة الدرجات لأنه يستخدم شكلاً من عدم التكافؤ الذي يحول دون أية مقارنة عبر الثقافات؛ لهذا السبب قد يبدو عدم التكافؤ غير مرغوب به، ومع ذلك فإن هذا الوضع قد يكون نقطة بدء لاستشكاف الفروقات المهمة عبر الثقافات التي تتجاوز انحياز البند أو المنهج وقد تشمل المفهوم التكويني للبنية (غرينفيلد، 1997 a؛ بورتينغا وفان دير فلير، 1988).

إذا كنا مهتمين بمعرفة ما إذا كان الاختبار يقيم بعض السمات أو القدرات في المجموعات المختلفة فيكفي إقامة تكافؤ بنيوي.. في هذه الحالة قل أن يكون انحياز المنهج ذا أهمية أساسية. تأخذ التحليلات الإحصائية للتكافؤ البنيوي بعين الاعتبار العلاقات المتبادلة (الروابط) في المقام الأول ولا يؤثر انحياز المنهج على هذه الروابط؛ لذلك فإن هذه الإحصائيات لا تتأثر بانحياز البنية.

عندما يتم تحليل درجات البند تصبح الصورة أكثر تعقيداً. وقد اقترح لمعالجة موضوع التكافؤ البنيوي تحليل عاملي استكشافي يتلوه تدوير الهدف. إن الاختلافات في محتوى العامل عبر المجموعات تؤخذ على أنها دليل على انحياز البند (إذا كانت قلة من البنود منحازة) أو غياب التكافؤ البنيوي (إذا كانت عدة بنود منحازة). إذا أبدى أحد البنود انحيازاً متمثلاً فمن المحتمل أن تبقى الروابط بين البند مع غيره من البنود ثابتة. لن يُشار إلى هذا البند على أنه مشبوه في تحليل عاملي استشكافي. مع ذلك فإن التحليل العالمي الاستكشافي سيكون دقيقاً في الانحياز غير المنظم؛ لأن هذا النمط من الانحياز يُرجح أن يؤدي إلى نماذج متباينة من الروابط عبر المجموعات الثقافية بين هذا البند وغيره من البنود. تدل الكتابات حول انحياز البند أن الانحياز المتسق أكثر شيوعاً من الانحياز غير المتسق. وهكذا فقد يكون تحليل عامل الاستكشاف متبوعاً بتدوير الهدف إجراءاً كافياً لتحديد انحياز البند عندما يكون القصد تكافؤ البنية فقط. عند مناقشة الأنماط القياسية للتكافؤ (وحدات قياس وتكافؤ تام في الدرجات) يجب أن يؤخذ بعين الاعتبار التمييز بين الانحياز المتسق والانحياز غير المتسق. لا يتعارض الانحياز المتسق مع التكافؤ المقاس بالوحدات؛ لأنه مع هذا المستوى من التكافؤ لا يمكن مقارنة الدرجات مباشرة عبر الثقافات؛ إن إضافة رقم ثابت إلى جميع الدرجات ضمن مجموعة واحدة لا يسيء إلى هذا النمط من التكافؤ. ومع ذلك فإن الانحياز المتسق سوف يؤثر على فاعلية المقارنة بين الدرجات التي تبدي تكافؤ مقياس ما من جهة ثانية سوف يؤدي الانحياز غير المتسق إلى عدم تكافؤ في القياس وفي الدرجة النهائية؛ لأن هذا النمط من الانحياز يقضي على تماثل وحدة القياس عبر المجموعات.

انحياز العينة:

إن جميع أشكال الانحياز التي استعرضناها في هذا الفصل تتناول التفاوت في تقييم درجات الاختبار على البشر وعلى مجال معين من السلوك في الثقافات

المختلفة. إلا أن نمطاً واحداً من الانحياز لم يتم ذكره وهو انحياز العينة أو عدم قابلية العينات للمقارنة. هذا يتعلق بتباين العينات في قدرتها على تمثيل ثقافة السكان الذي تؤخذ منهم العينات. مثال ذلك أنه في كثير من الدراسات تم أخذ العينات من طلاب الجامعات إلا أن دخول الجامعة في بعض البلدان يعتمد في الدرجة الأولى على الأداء المدرسي فيما يكون الوضع الاقتصادي والاجتماعي للأهل في بلدان أخرى هو المقياس الأول في قبول الطلاب في الجامعة. وهكذا إن هنا اختلافات في نظام العينات فيما يتعلق بمواصفات الخلفيات مما يعلل أي تباين يلاحظ في الدرجات بالإضافة إلى الصفات الثقافية للمجموعة المعنية. إن أحد الشواهد الموثقة في الكتابات عبر الثقافات تتعلق بالنتائج المعرفية للأمية. ففي كثير من الدراسات القديمة في هذا الخصوص انقلبت المقارنة بين الأميين وغير الأميين إلى مقارنة بين الأشخاص المنتسبين إلى المدرسة وغير المنتسبين لأن الانتساب إلى المدرسة والتعلم (معرفة القراءة والكتابة) مفهومان مختلفان (يتعلم المرء القراءة والكتابة في المدرسة). وهكذا فإن الاختلاف في الانتساب للمدرسة يوفر تعليلاً بديلاً لمعرفة القراءة والكتابة عند شرح الاختلافات. وقد درس سكريبنر وكول (1981) معرفة القراءة والكتابة في ليبيريا كما درسها بينت في كندا (1991) فوجدوا أن القراءة والكتابة تكتسب عن طريق التعليم غير الرسمي (خارج المدرسة) ومن الجدير بالاهتمام أن كلاً من تلك الدراسات وجدت أن النتائج المعرفية لعدم التعليم المدرسي كانت ضئيلة.

خيارات الترجمة/ تكييف الاختبارات:

جرت المناقشة من قبل أن المعايير اللغوية والنفسية للترجمة الجيدة لا تلتقي دوماً. فالكلمة والجملة وأي نص في أي اختبار تقييم قابل للترجمة بشكل جيد إذا كان تحويل النص من اللغة الأصلية إلى اللغة الهدف يحتفظ بجميع ميزات النص الأصلي. بعبارة أخرى يُعد النص قابلاً للترجمة إذا اتفقت الاعتبارات اللغوية والنفسية على تعريف الترجمة المثلى. تُركز الاعتبارات اللغوية على تساوي دلالات

الألفاظ (هل تعطي الترجمة المعاكسة النص الأصلي؟) وقابلية الفهم، القراءة والأسلوب. أما الاعتبارات النفسية فتشير إلى غياب الأنماط المختلفة من الانحياز التي ذكرت في الفقرات السابقة وبالتالي فهي تشمل اللغة الواقعية فيما تركز الاعتبارات اللغوية على مظاهر النص، تنظر الاعتبارات النفسية إلى الاختبار في نطاقه الثقافي الواسع. تتوافر ثلاثة خيارات للترجمة استناداً إلى قابلية الاختبار للترجمة. (هناك عدة فصول في هاركنس، فان دي فيجر وموهلر، 2003 التي تعالج موضوع الترجمة وتصميم الاستبيانات متعددة اللغات).

خيار المطابقة:

تكون عملية الترجمة سهلة إذا اتضح أن الترجمة اللغوية المناسبة كانت ملائمة أيضاً في الناحية النفسية. ترجمة مثل ذلك ستكون ترجمة حرفية في الغالب ولا تتطلب تغييرات كبيرة في الكلمات. وقد أطلق فان دي فيجر ولونغ (1997 b) على هذا الخيار اسم المطابقة لأن النص في اللغة الأصلية يمكن تطبيقه ببساطة في سياق ثقافي آخر. من المحتمل إلى حد كبير أن دراسة طرق الترجمة التي يُشار إليها في الصحف عبر الثقافات، مثل مجلة "علم النفس عبر الثقافات" ستبين أن معظم المقارنات المنشورة تعتمد على الترجمة الحرفية. إن هذا الخيار هو الأكثر استعمالاً في البحوث التجريبية لسببين اثنين. الأول هو أن هذه الترجمة سهلة التطبيق مما يجعل تكاليفه معقولة، إضافة إلى ذلك فإنها تحتفظ بكل إمكانيات المقارنة المتعلقة بتكافؤ القياسات، إلا أن خيار المطابقة له عيب مهم: فهو صالح للاستعمال فقط عندما يكون الانحياز (لا سيما انحياز البنية وانحياز المنهج) أمراً مستبعداً. من المؤسف أن كثيراً من الدراسات عبر الثقافات تستخدم هذه الطريقة (السريعة والسيئة) في الترجمة الحرفية دون الأخذ بعين الاعتبار أخطار المقارنة عبر الثقافات.

خيار التكييف:

الخيار الثاني يدعى التكييف. يُقصد بالتكييف في مصطلحاتنا الترجمة الحرفية لبعض المحرضات (المنبهات) وتبديل بعضها الآخر بحيث تضاعف من ملاءمتها لثقافة اللغة المستهدفة. وقد أصبح التكييف مصطلحاً عاماً في هذا الكتاب وفي غيره من المنشورات المتعلقة بعلم النفس يشار به إلى ترجمة الاختبارات. وقد كان اختيار هذا المصطلح مقصوداً، لأنه يؤكد على قصور الترجمة الحرفية وعلى الحاجة على الأقل إلى النظر في ملاءمة هذه الترجمات من الناحية النفسية إن لم يكن تغيير المظاهر البارزة للاختبار. تظهر الحاجة للتكييف في الاختبارات التي تبدي قابلية متوسطة للترجمة: قد يكون بالإمكان ترجمة بعض معالم الاختبار كالتعليم والأمثلة والتمارين مباشرة إلى اللغة المستهدفة إلا أن الكلمات قد تكون بحاجة للتبديل، قد تكون هذه التعديلات مطلوبة للتعامل مع مختلف أشكال الانحياز كما بحثنا ذلك مسبقاً. يمكن أن نجد أمثلة على ذلك في كتابات STAI لوغلانسمان، سكافنر وسبيلبرغر 1981؛ سبيلبرغر وغيره، 1970). "تم تعديل قائمة مينيسوتا لدراسة الشخصية" لتلائم العديد من الثقافات فقد عدّلها على سبيل المثال لوشيو، ريس - لاغونس وسكوت (1994) لتلائم المكسيك وعدَّ لها تشونك (تشونك ولونغ، 1998) لتلائم الصين. ومن الأمثلة الأخرى أن ليو عدّل (اختبار القدرات المعرفية) وهو اختبار تم تطويره في الولايات المتحدة كي يُصبح صالحاً للاستخدام في تشخيص الخرف عند السكان الصينيين ذوي المستوى المنخفض من التعليم الرسمي.

لا يهدف واضعو الاختبارات المكيف إلى إجراء مقارنات عبر الثقافات، وإنما يهدفون على الأرجح إلى تغطية بنية محددة في إحدى المجموعات الثقافية. وهكذا يُعد الهدف المحدود للمقارنة عبر الثقافات أمراً مفروغاً منه (الإرشادات 12).

خيار التجميع:

الخيار الثالث هو خيار التجميع. يستخدم هذا الخيار عندما تكون الأداة (الاختبار) عسيرة الترجمة. عندما تكون الترجمة الحرفية للأداء (الاختبار) غير

ممكنة لأسباب متعلقة بانحياز البنية أو المنهج وكان تكييف الاختبار لا يغطي البنية بشكل ملائم، فقد يعمد واضع الاختبار إلى تطوير/ جمع اختبار جديد تماماً في اللغة المستهدفة.

قام سيربل (1993) بدراسة مفهوم الذكاء عند الأفراد في زامبيا ووضع اختباراً مستنداً إلى هذا المفهوم يُعد مثالاً لاختبار التجميع. والمثال الآخر على هذا الخيار هو دراسة تشرش (1987) للشخصية في الفليبين التي قادته إلى وضع توجيهات لوضع اختبار للشخصية أكثر ملاءمة من الناحية الثقافية. يمكن الإشارة في هذا المجال إلى تشونك ورفاقه (1996) الذين وضعوا قائمة لتقييم الشخصية "الصينية" وهو اختبار يحتوي على العديد من أبعاد الشخصية المحلية مثل "الوجه" و"الانسجام". في كل هذه الأمثلة حاول الباحثون وضع صورة ملائمة للتكوين النفسي مستمدة من المفاهيم المحلية التي درسوها قبل وضع الاختبار. من وجهة نظر التكافؤ لا يؤمن خيار التجميع أي مجال لمقارنة الدرجات بشكل مباشر. مع ذلك فإن هذه الدراسات وثيقة الصلة بعلم النفس عبر الثقافات إذ تبين أن المفهوم الغربي أو الاختبار لا يمكن تطبيقه على بنية بعض الثقافات الأخرى. إن هذه الدراسات فعالة في تحقيق الغاية الأساسية من دراسة سيكولوجيا الثقافات وعبر الثقافات: وهي تحديد الانحياز الغربي في النظريات والاختبارات الحالية السائدة في ميدان علم النفس.

إن انتقاء أحد خيارات الترجمة أي المطابقة أو التكييف أو التجميع أمر مهم في أي مشروع بسبب مضامينه. فإذا كانت الغاية هي تحقيق تكافؤ قياسي (وحدة قياس أو درجة تامة) فإن خيار المطابقة هو الخيار الأفضل. على الرغم من أن بعض النماذج الإحصائية تسمح بمثل هذه المقارنات باستخدام اختبارات مكيفة. يفترض خيار المطابقة عدم وجود انحياز في البنية أو المنهج (انحياز البند يمكن معالجته بعد ذلك بحذف البند). في الوقت الذي يتعرض فيه العلماء إلى ضغوط كبيرة من

أجل نشر أبحاثهم فإن خيار المطابقة قد يصبح بسهولة الخيار الأسهل تناولاً لأنه يجمع بين الجهد القليل لإنجاز الترجمة مع التعهد الضمني بالتوصل إلى تكافؤ معياري. وهذا أمر مؤسف من وجهة نظر عبر الثقافات؛ لأن مثل هذه الممارسة قد تؤدي بسهولة إلى تطبيق اختبارات غير مناسبة ثقافياً، وإلى تقييم خاطئ للتباينات السيكولوجية بين أبناء الثقافات المختلفة (بورتينغا، 1975).

إننا نتفق مع المبدأ الأساسي في مختلف الإرشادات (مثل 20 و11 -(2.4) هامبلتون، 1994، فان دي فيجر وهامبلتون، 1996؛ انظر أيضاً هامبلتون، الفصل الأول في هذا الكتاب. إن مهمة الباحثين هي تبيان ملاءمة اختباراتهم. إن هذا الاقتراح ينحرف عن الممارسة الشائعة وفيها يقع عبء الإثبات على أكتاف الذين يستخدمون الاختبار.

عوضاً عن تضخيم ملاءمة الاختبار للمقارنة عبر الثقافات يمكن أيضاً زيارة صدقه البيئي. يمكن تحقيق ذلك بوساطة خيار التجميع وبدرجة أقل بواسطة خيار التكييف. والخلاصة يبدو أن تضخيم ملاءمة الاختبار سواء للمقارنة عبر الثقافات أو صدقها البيئي في إطار ثقافي محدد قد يؤدي إلى مقاربة مختلفة في عملية الترجمة/ التكييف.

تعزيز صدق الاختبار:

يتطلب التعامل بشكل فعال مع انحيازات اللجوء إلى مقاربة تحريضية. إن الدفاع عن هذه المقاربة كما عُبّر عنها في الفقرات السابقة يبقى ناقصاً دون إلقاء نظرة إجمالية على المناهج والإجراءات لتعيين وضبط/ مراقبة وحتى حذف الانحيازات. نحاول في هذا القسم إعطاء مثل على هذه النظرة الإجمالية ومع ذلك يتطلب الأمر الإيضاح أنه من غير الممكن تقديم قائمة شاملة، إذ إن كل مشروع لتكييف اختبار جديد سوف يصادف بعض الانحيازات الخاصة بذلك المشروع. من جهة أخرى سيكون هناك العديد من النقاط المتكررة. إن هذه المناقشة للمناهج والإجراءات ترتكز على الموضوعات المهمة التي نراها مهمة.

عندما يكون هناك شك بأن انحياز البنية والسلوك المرافق للبنية غير متماثل عبر المجموعات الإثنية يصبح من المستحيل ابتكار أي أداة (اختبار) تعطي درجات قابلة للمقارنة بين السكان من ثقافات مختلفة. من المهم تعيين المدى الذي وصل إليه التشابك (الإرشادات 2). إن القول بأن الولاء للأبوين أعلى (أو أقل) في الصين مثلاً مما هو في المملكة المتحدة قول مضلل دون الإشارة إلى أن البيئة الصينية (بنية الاختبار الصيني) غير كاملة إذا كانت تعتمد على أدوات غربية تهمل المظاهر المادية للولاء للأبوين، إذا كانت النتائج المعتمدة على التعريف الصيني يصبح الاختبار مفرط الشمولية في المملكة المتحدة.

هنالك على الأقل طريقتان للتعامل مع المشكلة. الأولى هي إعادة تحديد البنية التي يُراد قياسها بطريقة تتضمن الإشارة إلى فرط الشمولية أو نقص الشمولية، فعوضاً عن المقارنة بين الولاء الأبوي بشكل عام يمكن أن نصف نتائجنا على شكل المظاهر المادية وغير المادية للولاء الأبوي. والطريقة الثانية هي تطبيق تقنيات إحصائية خاصة لمعالجة مجموعات متباينة من الحوافز مثل نظرية الإجابة على البند (فيشر ومولينار، 1995؛ هامبلتون وسوامیناثن، 1985، هاملتون، سوامیناثن وروجرز، 1991؛ فان دير ليندن وهامبلتون، 1997) أو صياغة التبادل البنيوي (بولن، 1989؛ بيرون. 1998؛ ماركوليدس وشوماخر، 1969 (الارشادات 11).

إن المبدأ الأساسي هو تفكيك عالم كبير إلى واحد أو أكثر من الحقول الثانوية المشتركة عبر الثقافات مع الاحتفاظ في الوقت نفسه في كل مجموعة ثقافية على العلاقات بين هذه الحقول الثقافية الفردية.

إن إحدى الطرق لتجنب انحياز البنية هي اللامركزية (ورنر وكامبل، 1970). يُستخدم هذا المفهوم في الوقت الحاضر بأساليب مختلفة. إن المعنى الأصلي لتعبير اللامركزية يُقصد به تطوير اختبار بعدة لغات في آن واحد. وخلافاً لما هو قائم في معظم مشاريع الترجمة/ التكييف التي يتم فيها ترجمة أداة موجودة سابقاً إلى لغة

أخرى، تؤلف مجموعة تضم واضعي الاختبارات بالإضافة إلى ممثلين للغات المستهدفة. في هذه الحالة تصبح تعريفات البنية ومظاهر البنية مثل التعليمات والبنود المنحازة نحو مجموعة ثقافية معينة قابلة للتمييز من قبل المطورين من الثقافات الأخرى.

يمكن حذف الكلمات أو الجمل التي تشير إلى عادات أو معارف ثقافية معينة وتسيء إلى قابلية ترجمة الأداة والاستعاضة عنها بما يقابلها في التعابير القابلة للاستخدام عالمياً. وبالمقابل قد يقرر الباحثون أن ترجمة متكافئة غير ممكنة وأن من المتعذر إنشاء اختبار يسمح بمقارنة كمية. إن الأمثلة على مثل هذه المقارنة نادرة على الأرجح بسبب صعوبتها؛ ولأن الباحثين الذين استخدموا أحد الاختبارات من قبل يرغبون في توسيع قاعدة معلوماتهم.

إن شيوع استخدام البريد الإلكتروني وشبكة الإنترنت حديثاً قد يشكل قوة دافعة جديدة.

استخدم مفهوم اللامركزية حديثاً جداً في ترجمة الأدوات الموجودة. ويرقى ذلك إلى حذف البنود الخاصة بثقافة معينة واستبدالها بأدوات محفزة أكثر ملاءمة. ذلك يعني في مصطلحات هذا الفصل خيار التكييف. مثال ذلك استخدم كورتيز وسميث (1979) هذه المقاربة لإعداد ترجمة إسبانية لاستبيان إنجليزي حول التثاقف.

يرتبط باللامركزية ما يدعى بالمقاربة المتلاقية Convergence approach (كامبل 1986). لنفترض أن عالماً نفسياً أميركياً وآخر من زمبابوي يهتمان بالمهارات التي يعتقد الأبوان أن من الضروري تطويرها عند الأطفال وبالسن التي يصبح الطفل بارعاً بها (تدعى النظريات الإثنية الأبوية؛ س. هاركنس وسوبر، 1992). قام كل من العاملين النفسيين بتطوير مقياس يتناسب مع البيئة الثقافية لكل منهما ومن ثم استخدمت هاتان الأداتان في كلا البلدين. في هذه الحالة تأخذ التشابهات والتباينات لتلك المعطيات أهمية خاصة. لا تعرف أي أمثلة تجريبية عن هذا الإجراء.

بعد جمع المعطيات من المجموعات المختلفة، هناك إجراءات لتحليل التكافؤ. تعالج بعض هذه الإجراءات تكافؤ البنية والمنهج في آن واحد. تبدو هذه الإجراءات جذابة بسبب قدرتها وسهولة تطبيقها. مثال ذلك يمكن سؤال الرواة المحليين أن يحكموا على صحة فهم الاختبار وملاءمته (الارشادات 2). وبشكل خاص يمكن للمجموعات ثنائية اللغة أن تقدم معلومات مفيدة حول كفاءة أحداث الاختبار (انظر، سيغي، الفصل الخامس من هذا الكتاب). وبالمثل يمكن إجراء مسوحات محلية لسكان اللغة المستهدفة تطرح فيها أسئلة حرة الاستجابة (الارشادات 6-3). يمكن الاختبار أن يستخدم أيضاً في اللغة المستهدفة بطريقة غير نظامية لطلب تفسير الأجوبة. تبين هذه الشروح ما إذا كانت الأسئلة قد فُسرت بالشكل الذي قصده مؤلف الاختبار.

إذا حصل تعديل جوهري في الاختبار أثناء عملية التكييف، يصبح من المهم بيان صدق بنية اختبار ما بعد تكييفه (الارشادات 20-10). مثال ذلك، تشونغ (189)، الذي كيف اختبار MMPI للاستخدام في الصين، درس قدرة القياس على التمييز بين الأشخاص الأسوياء والمرضى وأعد ملامح الشخصية بمجموعات مختلفة. دعمت كلاً من الطريقتين صدق البنية، إن تطبيق القوالب المتعدد السمات والمتعددة المناهج هي وسائط معقدة لمعالجة انحياز المنهج عند تكييف الاختبار يصعب الحصول عليها (كامبل وفيسك، 1959).

إن الأكثر شيوعاً من ذلك هو تطبيق تحليل العامل المؤكد وتحليل العامل الاستكشافي متبوع بتدوير الهدف (مثال، تايلوز وبوينس، 1991؛ واتكينس، 1989، ويندل وايواواكي ولينر، 1988) (مكاري وكوستا، 1997؛ بيد مونت وتشاي، 1997؛ شميدت وبه، 1992؛ فاند نبرغ وهاكستين، 1978).

إن الابتكار الحديث في تحليل العامل الاستكشافي هو تحليل العناصر المتزامن (كيرز، 1990؛ كيرز وتن بيرغ، 1989). خلافاً لما هو الحال في تحليل العامل الاستشكافي فإن جميع المعلومات تعالج مع بعضها، هناك مجموعة وحيدة من

العناصر الأساسية التي تُقيم عند سائر المجموعات. هذه العناصر الرئيسة هي بالتعريف متماثلة بين كل المجموعات؛ لذلك لا حاجة لتقويم الانسجام كما هو الحال في تحليل العامل الاستشكافي. إن نسبة التفاوت المبينة بوساطة العوامل الأساسية المشتركة تقارن مع تحليل العامل الأساسي في مجموعات منفصلة من المعطيات. يمكن إيجاد أمثلة عند زوكرمان، كوهلمان، ثورنكويست وكمبير (1991).

إن جميع الإرشادات الإدارية (18-13) واثنين من الإرشادات الوثائقية (22-19) يمكن اعتبارها مقترحات حول كيفية التقليل من انحياز المنهج. إن الطرق الفعالة لمعالجة انحياز المنهج تكمن في التدريب الشامل للذين يستخدمون الاختبار (ويتضمن ذلك التدرب على التواصل بين الثقافات فيما إذا كان المختبرون والمختَبرون من إثنيات مختلفة) وإعداد كتيبات متصلة خاصة بالاختبار تضمن تعليمات دقيقة. حتى في حالة تطبيق هذه المقترحات بدقة يبقى من المحتمل حدوث انحياز في المنهج يهدد صدق المقارنات عبر الثقافات. عندما تكون الفجوة الثقافية بين المجموعة الأصلية والمجموعة المستهدفة واسعة، فإن العينات ستختلف كثيراً في صفاتها ذات الصلة بالنتائج مثل التعليم، الحوافز، وأسلوب الإجابة، بحيث أن تقويم هذه الصفات يصبح الطريقة الوحيدة لضبط النتائج. إن الإجراءات الإحصائية مثل تحليل الخلافات المشتركة يمكن تطبيقها حينئذٍ لمعرفة إلى أي مدى يمكن لتباين الدرجات عبر الثقافات أن يكون ناجماً عن صفات خلفية العينات (الإرشادات 11).

هناك نوع آخر من الإجراءات لتحليل التكافؤ بين المعطيات وهو الاختبار وإعادة الاختبار أو الدراسة التدريبية التي تسمح بمقارنة نماذج المكتسبات عبر الثقفات. إن اختلاف نماذج المكتسبات يعطي دليلاً قوياً على سوء تكافؤ الدرجات. يمكن مشاهدة أمثلة على ذلك في نكيا، هوتو وبونت (1994). فقط طبق هؤلاء المؤلفون القوالب المعيارية لرافن ثلاث مرات على تلاميذ الصف السادس في فرنسا والكونغو. وقد وجدوا تحسناً معتدلاً في المرة الثانية ولم يجدوا أي تحسن في المرة

الثالثة في كل من المجموعتين عندما لم يكن هناك تحديد للزمن. مع ذلك عند تحديد الزمن أبدى أفراد المجموعتين تحسناً واضحاً في الدرجات في المرة الثانية كما أبدى التلاميذ من الكونغو تحسناً آخر في المرة الثالثة. إن هذه النتائج تعارض بشكل واضح قابلية مقارنة الدرجات في المرة الأولى من تطبيق الاختبار.

إن المجموعة الأخيرة من تقنيات تعزيز المصداقية هي التي تعالج انحياز البند (الإرشادات 9، 8، 7) وإن كلاً من إجراءات القياس السيكولوجي قد جرى ذكرها في مكان آخر من هذا الكتاب (انظر الفصل الأول والرابع).

الاستنتاج:

إن الشبهة بوجود انحياز ثقافي أو وجود بنية تجريبية على ذلك يعني أن الأداة (الاختبار) المقصودة لا يمكن اعتبارها بنية ذات أهمية أو نموذجاً عادلاً. إن أكثر الخطط المحافظة هي تلك التي تحاول أن تبرهن على أن الانحياز يمنع كل أشكال المقارنة. في اعتقادنا أن هذه الخطة هي الوحيدة الصحيحة عندما تتحدد البنيات حسب أسلوب ثقافة معينة. مع ذلك فإن الخطط البديلة ممكنه عندما تبدي البنيات تداخلاً عبر الثقافات وعندما يكون الانحياز غير ناجم عن البنية بحد ذاتها وإنما ينجم عن الطريقة التي يستعمل فيها اختبار معين. وقد بيّنا في هذا الفصل الفوارق المختلفة التي تساعد على معرفة الانحياز والنتائج السلبية التي تنتج عنه فيما يتعلق بتكافؤ درجات الاختبار عبر الثقافات والمقارنات التي تفيد في تجنب تلك النتائج.

إن تكييف الاختبار عبر الثقافات له جانب مفهومي وجانب مقياس متري، وقد أكدنا في هذا الفصل على الجانب الأول/ المفهومي.

إن استخدام الاختبارات عبر الثقافات وتفسير الاختلافات في نماذج الدرجات ومستوى الدرجات تواجه مآزق خطيرة، إلا أنه يمكن تجنب الكثير منها إذا كان معدو الاختبار ومستخدموه مدركين لوجودها.

*****

المراجع

Azuma, H., & Kashiwagi, K. (1987). Descriptors for an intelligent person: A Japanese study. *Japanese Psychological Research, 29,* 17–26.

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

Berry, J. W., & Bennett, J. A. (1991). Cree literacy. Cultural context and psychological consequences. *Cross-Cultural Psychology Monographs, no. 1.* Tilburg, Netherlands: Tilburg University Press.

Berry, J. W., Poortinga, Y. H., Segall, M. H., & Dasen, P. R. (1992). *Cross-cultural psychology. Research and applications.* Cambridge, England: Cambridge University Press.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Bracken, B. A., & Barona, A. (1991). State of the art procedures for translating, validating and using psychoeducational tests in cross-cultural assessment. *School Psychology International, 12,* 119–132.

Brislin, R. W. (1980). Translation and content analysis of oral and written material. In H. C. Triandis & J. W. Berry (Eds.), *Handbook of cross-cultural psychology* (Vol. 1, pp. 389–444). Boston: Allyn & Bacon.

Brislin, R. W. (1986). The wording and translation of research instruments. In W. J. Lonner & J. W. Berry (Eds.), *Field methods in cross-cultural research* (pp. 137–164). Newbury Park, CA: Sage.

Butcher, J. N. (Ed.). (1996). *International adaptations of the MMPI-2: Research and clinical applications.* Minneapolis: University of Minnesota.

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS.* Mahwah, NJ: Lawrence Erlbaum Associates.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105,* 456–466.

Campbell, D. T. (1986). Science's social system of validity-enhancing collective believe change and the problems of the social sciences. In D. W. Fiske & R. A. Shweder (Eds.), *Metatheory in social science* (pp. 108–135). Chicago: University of Chicago Press.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81–105.

Cheung, F. M. (1989). A review on the clinical applications of the Chinese MMPI. *Psychological Assessment, 3,* 230–237.

Cheung, F. M., & Leung, K. (1998). Indigenous personality measures: Chinese examples. *Journal of Cross-Cultural Psychology, 29,* 233–248.

Cheung, F. M., Leung, K., Fan, R. M., Song, W. Z., Zhang, J. X., & Chang, J. P. (1996). Development of the Chinese Personality Assessment Inventory. *Journal of Cross-Cultural Psychology, 27,* 181–199.

Church, T. A. (1987). Personality research in a non-Western setting: The Philippines. *Psychological Bulletin, 102,* 272– 292.

Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: Issues and Practice, 17,* 31–44.

Cleary, T. A., & Hilton, T. L. (1968). An investigation of item bias. *Educational and Psychological Measurement, 28,* 61–75.

Cole, M. (1996). *Cultural psychology. A once and future discipline.* Cambridge, MA: Harvard University Press.

Cortese, M., & Smyth, P. (1979). A note on the translation to Spanish of a measure of acculturation. *Hispanic Journal of Behavioral Sciences, 1,* 65–68.

Cotter, P. R., Cohen, J., & Coulter, P. (1982). Race-of-interviewer effects in telephone interviews. *Public Opinion Quarterly, 46,* 278–284.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302.

Embretson, S. E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin, 93,* 179–197.

Erickan, K. (2002). Disentangling sources of differential item functioning in multilanguage assessments. *International Journal of Testing, 2,* 199–215.

Fischer, G. H., & Molenaar, I. W. (Eds.). (1995). *Rasch models. Foundations, recent developments, and applications.* New York: Springer-Verlag.

Gass, S. M., & Varonis, E. M. (1991). Miscommunication in nonnative speaker discourse. In N. Coupland, H. Giles, & J. M. Wiemann (Eds.), *Miscommunication and problematic talk* (pp. 121–145). Newbury Park, CA: Sage.

Geisinger, K. F. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment, 6,* 304–312.

Greenfield, P. M. (1997a). Culture as process: Empirical methods for cultural psychology. In J. W. Berry, Y. H. Poortinga, & J. Pandey (Eds.), *Handbook of cross-cultural psychology* (2nd ed., Vol. 1, pp. 301–346). Boston: Allyn & Bacon.

Greenfield, P. M. (1997b). You can't take it with you: Why ability assessments don't cross cultures. *American Psychologist, 52,* 1115–1124.

Hambleton, R. K. (1994). Guidelines for adapting educational and psychological tests: A progress report. *European Journal of Psychological Assessment, 10,* 229–244.

Hambleton, R. K., & Swaminathan H. (1985). *Item response theory: Principles and applications.* Dordrecht, Netherlands: Kluwer.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage.

Harkness, J. (1998). Cross-cultural survey equivalence [Special issue]. *ZUMA Nachrichten* (no. 3). Mannheim, Germany: Zentrum für Umfragen, Methoden und Analysen.

Harkness, J. A., van de Vijver, F. J. R., & Mohler, P. Ph. (Eds.). (2003). *Cross-cultural survey methods.* New York: Wiley.

Harkness, S., & Super, C. (1992). Parental ethnotheories in action. In I. E. Sigel, A. V. McGillicuddy-DeLisi, & J. J. Goodnow (Eds.), *Parental belief systems: The psychological consequences for children* (2nd ed., pp. 373–392). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ho, D. Y. F. (1996). Filial piety and its psychological consequences. In M. H. Bond (Ed.), *Handbook of Chinese psychology* (pp. 155–165). Hong Kong: Oxford University Press.

Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel–Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129–145). Hillsdale, NJ: Lawrence Erlbaum Associates.

Holland, P. W., & Wainer, H. (Eds.). (1993). *Differential item functioning.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Hui, C. H., & Triandis, H. C. (1989). Effects of culture and response format on extreme response style. *Journal of Cross-Cultural Psychology, 20,* 296–309.

Jensen, A. R. (1980). *Bias in mental testing.* New York: Free Press.

Kiers, H. A. L. (1990). *SCA: A program for simultaneous components analysis.* Groningen, Netherlands: IEC ProGamma.

Kiers, H. A. L., & ten Berge, J. M. F. (1989). Alternating least squares algorithms for simultaneous components analysis with equal component weight matrices for all populations. *Psychometrika, 54,* 467–473.

Laux, L., Glanzmann, P., Schaffner, P., & Spielberger, C. D. (1981). *Das State–Trait Angstinventar. Theoretische Grundlagen und Handanweisung* [The State–Trait Anxiety Inventory. Theoretical background and manual]. Weinheim, Germany: Beltz Test.

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research, 32,* 53–76.

Liu, H. C., Chou, P., Lin, K. N., Wang, S. J., Fuh, J. L., Lin, H. C., Liu, C. Y., Wu, G. S., Larson, E. B., White, L. R., Graves, A. B., & Teng, E. L. (1994). Assessing cognitive abilities and dementia in a predominantly illiterate population of older individuals in Kinmen. *Psychological Medicine, 24,* 763–770.

Lucio, E., Reyes-Lagunes, I., & Scott, R. L. (1994). MMPI–2 for Mexico: Translation and adaptation. *Journal of Personality Assessment, 63,* 105–116.

Marcoulides, G. A., & Schumacker, R. E. (1996). *Advanced structural equation modeling: Issues and techniques.* Mahwah, NJ: Lawrence Erlbaum Associates.

McCrae, R. R., & Costa, P. T. (1997). Personality trait structure as a human universal. *American Psychologist, 52,* 509–516.

Mellenbergh, G. J. (1982). Contingency table models for assessing item bias. *Journal of Educational Statistics, 7,* 105–118.

Merenda, P. F. (1994). Cross-cultural testing: Borrowing from one culture and applying it to another. In L. L. Adler & U. P. Gielen (Eds.), *Cross-cultural topics in psychology* (pp. 53–58). Westport, CT: Praeger/Greenwood.

Miller, J. G. (1997). Theoretical issues in cultural psychology. In J. W. Berry, Y. H. Poortinga, & J. Pandey (Eds.), *Handbook of cross-cultural psychology* (2nd ed., Vol. 1, pp. 85–128). Boston: Allyn & Bacon.

Millsap, R. J., & Everson, H. T. (1993). Methodological review: Statistical approaches for assessing measurement bias. *Applied Psychological Measurement, 17,* 297–334.

Nkaya, H. N., Huteau, M, & Bonnet, J. (1994). Retest effect on cognitive performance on the Raven-38 Matrices in France and in the Congo. *Perceptual and Motor Skills, 78,* 503–510.

Piedmont, R. L., & Chae, J-H. (1997). Cross-cultural generalizability of the five-factor model of personality: Development and validation of the NEO PI-R for Koreans. *Journal of Cross-Cultural Psychology, 28,* 131–155.

Poortinga, Y. H. (1975). Limitations on intercultural comparison of psychological data. *Nederlands Tijdschrift voor de Psychologie, 30,* 23–39.

Poortinga, Y. H., & van der Flier, H. (1988). The meaning of item bias in ability tests. In S. H. Irvine & J. W. Berry (Eds.), *Human abilities in cultural context* (pp. 166–183). Cambridge, England: Cambridge University Press.

Reese, S. D., Danielson, W. A., Shoemaker, P. J., Chang, T., & Hsu, H.-L. (1986). Ethnicity-of-interviewer effects among Mexican-Americans and Anglos. *Public Opinion Quarterly, 50,* 563–572.

Schmidt, S. M., & Yeh, R. (1992). The structure of leader influence: A cross-national comparison. *Journal of Cross-Cultural Psychology, 23,* 251–264.

Scribner, S., & Cole, M. (1981). *The psychology of literacy.* Cambridge, MA: Harvard University Press.

Segall, M. H., Dasen, P. R., Berry, J. W., & Poortinga, Y. H. (1990). *Human behavior in global perspective. An introduction to cross-cultural psychology.* New York: Pergamon.

Serpell, R. (1979). How specific are perceptual skills? *British Journal of Psychology, 70,* 365–380.

Serpell, R. (1993). *The significance of schooling. Life-journeys in an African society.* Cambridge, England: Cambridge University Press.

Shepard, L., Camilli, G., & Averill, M. (1981). Comparison of six procedures for detecting test item bias using both internal and external ability criteria. *Journal of Educational Statistics, 6,* 317–375.

Singer, E., & Presser, S. (1989). The interviewer. In E. Singer & S. Presser (Eds.), *Survey research methods* (pp. 245–246). Chicago: University of Chicago Press.

Sinha, D. (1997). Indigenizing psychology. In J. W. Berry, Y. H. Poortinga, & J. Pandey (Eds.), *Handbook of cross-cultural psychology* (2nd ed., Vol. 1, pp. 131–169). Boston: Allyn & Bacon.

Sireci, S. G., & Allalouf, A. (2003). Appraising item equivalence across multiple languages and cultures. *Language Testing, 20*(2), 148–166.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *Manual for the State–Trait Anxiety Inventory ("Self-Evaluation Questionnaire").* Palo Alto, CA: Consulting Psychologists Press.

Sternberg, R. J., Conway, B. E., Ketron, J. L., & Bernstein, M. (1981). People's conceptions of intelligence. *Journal of Personality and Social Psychology, 41,* 37–55.

Taylor, T. R., & Boeyens, J. C. (1991). The comparability of the scores of Blacks and Whites on the South African Personality Questionnaire: An exploratory study. *South African Journal of Psychology, 21,* 1–11.

Triandis, H. C. (1978). Some universals of social behavior. *Personality and Social Psychology Bulletin, 4,* 1–16.

Vallerand, R. J. (1989). Vers une methodologie de validation trans-culturelle de questionnaires psychologiques: Implications pour la recherche en langue française [Toward a methodology for the transcultural validation of psychological questionnaires: Implications for research in the French language]. *Canadian Psychology, 30,* 662–680.

van de Vijver, F. J. R. (2003). Test adaptation/translation methods. In R. Fernández-Ballesteros (Ed.), *Encyclopedia of psychological assessment* (pp. 960–964). Thousand Oaks, CA: Sage.

van de Vijver, F. J. R. & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist, 1,* 89–99.

van de Vijver, F. J. R., & Leung, K. (1997a). Methods and data analysis of comparative research. In J. W. Berry, Y. H. Poortinga, & J. Pandey (Eds.), *Handbook of cross-cultural psychology* (2nd ed., Vol. 1, pp. 257–300). Boston: Allyn & Bacon.

van de Vijver, F. J. R., & Leung, K. (1997b). *Methods and data analysis for cross-cultural research.* Newbury Park, CA: Sage.

van de Vijver, F. J. R., & Leung, K. (2000). Methodological issues in psychological research on culture. *Journal of Cross-Cultural Psychology, 31,* 33–51.

van de Vijver, F. J. R., & Poortinga, Y. H. (1997). Towards an integrated analysis of bias in cross-cultural assessment. *European Journal of Psychological Assessment, 13,* 29–37.

van de Vijver, F. J. R., & Tanzer, N. K. (1997). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology, 47,* 263–280.

van der Linden, W. J., & Hambleton, R. K. (Eds.). (1997). *Handbook of modern item response theory.* New York: Springer-Verlag.

Vandenberg, S. G., & Hakstian, A. R. (1978). Cultural influences on cognition: A reanalysis of Vernon's data. *International Journal of Psychology, 13,* 251–279.

Wagner, D. A. (1981). Culture and memory development. In H. C. Triandis & A. Heron (Eds.), *Handbook of cross-cultural psychology* (Vol. 4, pp. 187–232). Boston: Allyn & Bacon.

Watkins, D. (1989). The role of confirmatory factor analysis in cross-cultural research. *International Journal of Psychology, 24,* 685–701.

Werner, O., & Campbell, D. T. (1970). Translating, working through interpreters, and the problem of decentering. In R. Naroll & R. Cohen (Eds.), *A handbook of cultural anthropology* (pp. 398–419). New York: American Museum of Natural History.

Windle, M., Iwawaki, S., & Lerner, R. M. (1988). Cross-cultural comparability of temperament among Japanese and American preschool children. *International Journal of Psychology, 23,* 547–567.

Zuckerman, M., Kuhlman, D. M., Thornquist, M., & Kiers, H. A. L. (1991). Five (or three) robust questionnaire scale factors of personality without culture. *Personality and Individual Differences, 12,* 929–941.