10

تطوير الاختبارات للاستعمال في اللغات والثقافات المتعددة: التماس للتطوير المتزامن

نوربرت ك. تانزر
جامعة أليبانت العالمية

إن عولمة السوقِ الاقتصادية، وقابلية الحركة المتزايدة لقوة العالم العاملةِ، وظهور المجتمعاتِ المتعدّدة الثقافاتِ المعقدةِ، مثل الإتحاد الأوربي، قد جَلبَ تحديات جديدةَ إلى الاختبار التربويِ والنفسيِ الذي تجاوز المجالَ التقليدي للتقييمات الأحادية اللغة/ الأحادية الثقافة. مع الوقت، أصبحت الحاجة لأدوات صالحة في التقييمات المتعددة الثقافة/ المتعددة اللغة، أكثر أهمية (بارترام وكوين، 1998، هو وأوكلاند، 1991، أوكلاند، 1997، 2004، أوكلاند وهو، 1992).

لقد تم القبول الآن على نحو واسع، بأن تطور أدوات مثل تقييمات أو اختبارات متعددة الثقافة/ متعددة اللغة تستلزم أكثر من مجرد الترجمة، أي إعادة كتابة نص ما من لغة إلى أخرى. على الرغم من أن فريقاً من المترجمين المحترفين يستخدمون طريقة الترجمة - ترجمة مرتجعة (بريسلين، 1980)، قد ينتجون صيغ اختبارات متعددة اللغات متكافئة لغوياً. هذه النسخِ لا تَشترك في المعنى النفسي نفسه بالضرورة.

إن الترجمة الصحيحة للمادةِ "شاهدت التلفاز أكثرَ من المعتاد" سَتَبْقى مادةُ غير ملائمةُ عندما تُقدّمُ إلى سكان الساحل الذين لا يوجد لديهم كهرباءُ في بيوتِهم. (فان هامتن وفان دي فيفر، 1996).

يظهر هذا المثال بشكل واضح أن التقييمات المتعددة الثقافة/ المتعددة اللغة تَتطلّبُ معالجة متوازنة للاعتبارات النفسية، والقياسية، واللغوية والثقافة المتقاطعة، والثقافة، وذلك لكي تَضْمنَ صدق التركيب (تكافؤ التركيب). صمّمَ الاختبار لقيَاْسه ("تكافؤ أداة")، إدارة الاختبار (تكافؤ الإدارةِ")، وقد استخلصت بعض النتائج من علامات الاختبار (براكين وبارونا، 1994، بريسلين، 1980، 1986، غيسينجير، 1994، غرينفيلد، 1997، هامبلتون، 1994، فان دي فيفر ولونغ، 1997a، 1997b، فان دي فيفر وتانزر، 1997). لاحظ أن الجزء المهم لعملية التثقيف الثقافي (أو الاجتماعي) يُنقل عبر اللغة. وهكذا، فإن الأشخاص الذين لديهم لغات أمّ مختلفة لديهم أيضاً خلفيات ثقافية مختلفة. لهذا، فإن التقييمات التي تُجرى بصيغ لغوية مختلفة للممتحنين الذين لديهم لغات أمّ مختلفة هي أيضاً على الأغلب متعددة الثقافات. بناءً على هذا، التقييمات المتعددة اللغات عادة ما تكون تقييمات متعددة الثقافات. لاحظْ أيضاً بأنّ سكان المجموعات العرقية المختلفة، حتى لو كانوا يَتكلّمونَ لغة مشتركة، (أي: أغلبية/ أقلّيات)، (ما زالَتْ تُشكّلُ مجموعات ثقافية مختلفةَ. من هنا، فإن التقييمات في المجتمعات الأحادية اللغة، المتعددة الأعراق، تكون متعددة الثقافات، حتى لو أُجريت بالصيغة اللغوية نفسها لاختبار ما .(مثال، "التقييمات المتعددة الثقافات/ الأحادية اللغة").

تكافؤ البنية

بما أن الاختبارات الأحادية الثقافة/ الأحادية اللغة يمكن أن تطور بالاستناد على النظرياتِ السائدةِ للتراكيبِ التي يَنْوونَ قيَاْسها، فإن تطوير أدوات صادقة للتقييمات المتعددة الثقافة والمتعددة اللغة تَتطلّبُ قابلية التعميم للتقاطع الثقافي للتراكيب عبر كل الثقافات واللغات المقصودة. في تكوين فكرتهم عن التراكيب

النفسية، يميل مطورو الاختبارات الأحادية الثقافة إلى التعرض لخطر التحيز العرقي مما قَدْ يُهدّدُ صدق الاختبارات عبر الثقافات. على سبيل المثال، بينت دراسات سابقة متراكمة أن أدوات قياس الشخصية الغربية لا تحدد هيئة الشخصية الصينية لأحد السكان الأصليين مثل "الوجه" و"الانسجام" (سونغ، 2004، شونغ وآلو 1996، يانغ وبوند، 1990، زانغ وبوند، 1998).

تكافؤ الإدارةَ والأداة

علاوة على، وما وراء، تكافؤ التراكيب المتقاطعة ثقافياً ولغوياً، يجب على مطوري الاختبارات أن يقدموا دليلاً على تكافؤ التقاطع الثقافي واللغوي لأدواتهم (أي: غياب "تحيّز الأداة") والإجراءات المتعلقة بإدارتهم (أي: غياب "تحيّز الإدارة"). ومن أجل تسهيل هذه المهمّةِ، قامت لجنة الاختبار الدولية (ITC) بتحضير اختبارات دلائل التكييف ITC، وهي مجموعة تتألف من 22 دليلاً للممارسات الموصى بها لتطوير الاختبارات المتعددة الثقافة/ المتعددة اللغة، التي ذُكرت في الفصل الأول من هذا الكتاب (انظر أيضاً الى هامبلتون، 1994، 2002، فان دي فيفر هامبلتون، 1996). لتَأكيد سياقِ التقييمات المتعددة اللغاتِ والمتعدّدة الثقافاتِ، يقوم الدليل بتوجيه عدة معايير للتطوير (توجيهات 3-12)، الإدارة (توجيهات 13-18)، التوثيق (توجيه 19) لأدوات التقاطع الثقافي واللغوي، بالإضافة إلى تفسيرِ علامات الاختبار (توجيهات 20-22).

إن دلائل التكييف ITC تؤكّدُ، بشكل خاص، على الحاجة الى تقديم بينة على صدق التقاطع الثقافي للتركيب (توجيه رقم 2) والاختبار (توجيهات 1، 3-6، 10-11) عبر كُلّ السكان واللغات المقصودة. وقد أكدوا أيضاً على أهمية تَحديد وتَبرير مقارناتٍ صادقة بين النَتائِجِ (مثال، علامات الاختبار) التي تم تحصيلها من الصيغ اللغوية المختلفة أو مِنْ المجموعاتِ الثقافيةِ المختلفةِ (توجيهات 12، 14، 20-22). علاوة على ذلك، يُؤكّدونَ أيضاً على تَوثيق (توجيهات 4، 17، 19) كُلّ التغييرات التي

كَانتْ ضرورية أثناء عملية التكيّفِ أو التطويرِ التي تضمنت محتويات السؤال أو مادّةِ المحفّزِ (توجيهات 6، 14)، صيغ إجابة (توجيهات 5، 14)، تعليمات الاختبار (توجيهات 4، 16)، إجراءات الإدارة (توجيهات 5، 15، 16، 18)، وقواعد العلامات (توجيهات 12، 20–21). أخيراً، يُؤكّدونَ على أهمية تطبيق تقنيات إحصائيةَ وتصاميمَ مناسبة ومجموعة بيانات لاكتشاف المصادرِ المحتملة لعدم التكافؤ ولدِراسَة فعالية أيّ قياسات مقابلة تم أخذها (توجيهات 7–11، 13). أخيراً وليس آخراً، إن مؤهلات مستخدمي الاختبارات لإجراء تقييمات متعددة الثقافة/ متعددة اللغة (توجيهات 14، 18) هي أيضاً قد تم إلقاء الضوء عليها.

التطورات المتزامنة مقابل التطورات المتعاقبة لأدوات متعلقة بحالاتِ التقييمِ المتعدّدة الثقافاتِ والمتعددة اللغاتِ

بالرغم من أن السبب الجوهري وراء كُلّ توجيه تم التعليق عليه وتوضيحه بالأمثلةِ التجريبيةِ (فان دي فيفر وهامبلتون، 1996)، فإن دلائل تكييف الاختبار (ITC) معيارية أكثر منها إرشادية. فهي لا تقدم إستراتيجيات للحُصُول على اختبارات صادقة متعددة الثقافات/ متعددة اللغات. ولا تناقش حسنات وسيئات إيجاد طرق معالجة.

في طريقة معالجة تطور اختبار المحاكاة، تم تطوير أداة جديدة للاستعمال في عدد مِنْ مجموعاتٍ ثقافيةٍ محددة مسبقاً (ثقافات مرجعية) و/ أو لغات (لغات مرجعية). يستلزم ذلك عادة " طريقة معالجة لجنة"، أي إنها قدرة عمل متعددة اللغات مِنْ خلفياتٍ ثقافيةٍ مُخْتَلِفةٍ وبالخبرة المكمّلة في علْمِ النفْس "السائدِ" (يتضمن ذلك معرفة التراكيب ومقياسها)، علم النفس القياسي، تقنيات تراكيب الاختبار، بالإضافة إلى الثقافة النفسية (المحلية) وعلم النفس المتعلق بالتقاطع الثقافي واللغويات. إن فائدة هذه الطريقة هي أنه تضمن الحد الأقصى من عدم

مركزية اللغة والثقافة في تعريفِ التركيبِ والاختبار الذي صمم لقيَاْسه. إن الخواص المعينة للغة محددة (ومثال، تعابير محليّة) أو ثقافة ما (مثال، معايير اجتماعية) يُمْكِنُ أَنْ تُكتَشفَ وتُزالَ أثناء مراحل مبكّرةٍ مِنْ تطويرِ الاختبار؛ لذا، فإن طريقة المعالجة المتزامنة ذات دور فعّال في تطوير اختبارات صادقة متعددة الثقافة/ متعددة اللغة، بخصائص جيدة على حد سواء لكُلّ مِن الثقافاتِ المرجعية.

على أي حال، هذا يَتضمّنُ أكثر مِنْ مجرد إنتاج عدة صيغ لغوية لأداة ما، وتجْميع البياناتَ في المجموعاتِ الثقافيةِ المُختَارةِ بشكل عشوائي ("دراسات رحلةِ صيد").

في طريقة تطويرِ الاختبار المتعاقبةِ، بالممارسة الأكثر شيوعاً، يُطوّرُ الاختبار ويُصدّقُ بمطور، أو عِدّة مطوّري اختبار من خلفية معينة متعددة الثقافة/ متعددة اللغة، وذلك الاستخدام ضمن هذا السياق متعدد الثقافة/ متعدد اللغة ("مصدر" لغة/ ثقافة). في أغلب الأحيان، بعد أن يصبح اختباراً شائعاً بعدة سنوات، عندها فقط يتم تكييفه من قبل مطوري الاختبار الأصليين.، أو من قبل لجنة عمل جديدة للاستخدام في الثقافات الأخرى و/ أو إن علم النفس المختص بالتقاطع الثقافي يمكن أن يُؤدّي - دون قصد - إلى انحياز لتحيزات عرقية أو ثقافية تُحدّدُ التطويرَ اللاحقَ مِنْ النسخِ ذات الجودة المعادلة، في اللغات الهدف الجديدة و/ أو الثقافات.

هذه المشكلة ستتفاقم بشكل أكبر إذا اشتركت لجنةُ عمل في تكيّفِ الاختبار لا تملك نطاقاً كاملاً من الخبرة كما وصفت الطريقة الآنيةِ. على الرغم من هذا، بمثل هذا العدد من الاختبارات الممتازة الأحادية الثقافة/ الأحادية اللغة، فإن طريقة المعالجة التي تعتمد على التعاقب، سَتَستخدمُ كثيراً في المستقبلِ المنظورِ.

المخاطر والعلاج
في تطبيقات الاختبارات
المتعددة الثقافة/ المتعددة اللغة

سَتُصادفُ الباحثين المختصين في تطويرِ الاختبارات المتعددة الثقافات/ المتعددة اللغات مشكلات ومخاطر غائبة في تطبيقات الاختبارات المتعددة الثقافات/ المتعددة اللغات.

باستعمال عدد مِنْ الأمثلةِ مِنْ موجوداتِ التقاريرِ الذاتيةِ وفحوصاتِ الأهلية غير الشفهية، نُصوّرُ تشكيلة واسعة لهذه المخاطرِ، وتحدد اختبارات دلائل التكييف ITC التي قَدْ تطبق عليهم، ونَقترحُ حلولاً محتملةَ. الأمثلة الأخرى للمخاطرِ والعلاجِ المحتملِ مقدمة في غرينفيلد (1997)، فان دي فيفر ولونغ، (1997b)، و فان دي فيفر وتانزر (1997).

مخاطر سببها المواد الوحيدة:

على خلاف التركيبِ وتحيّز الأداة/ إدارةِ، "وظيفة مادة تفاضلية" (سيرسي وآلالوف،2003) أو "تحيّز مادةِ" سببت بواسطة التشويهاتِ في مستوى المادةَ، فإن مادةً متحيزة لَها معان نفسية مختلفة عبر الثقافاتِ.

بالرغم من أن تحيّز المادةِ يُمْكِنُ أَنْ يُنتَجَ من قبل بالعديد مِنْ المصادرِ، إلا أن السبب الأكثر تكراراً هو (أ) المقاييس الإنسانية مثل الترجمة الركيكة، أو الغموض في محتوى المادة الأصلية. أو من قبل (ب) "خواص ثقافية " أصيلة، مثل عدم الاطلاع/ تناسب محتوى المادةَ في بَعْض الثقافاتِ (توجيه 6). كما هو مُصَوَّرٌ في المثال التالي، إنه من الصعب في أغلب الأحيان في المواقع المتعددة اللغاتِ والمتعدّدة الثقافاتِ تَقْرير أي من هذين الاحتمالين يُمكنُ أنْ يبعد التفسيرات المختلفةَ و/ أو ردود الأفعال على الموضوعات.

محتوى المادةِ الغامضِ (المثال 1). في بيان حالة التعبير لنبرة الغضب، تعبيرِ غضبِ الميزةِ الرسميةِ (STAXI، سبيلبرغر 1988)، تم تمييز ثلاثة مِنْ أساليبِ تعبيرِ

الغضبِ. غضب موجه للخارج، وهو تعبيرُ الغضبِ نحو الناسِ أو الأشياء الأخرى. غضب داخلي، وهو إخمادُ المشاعرِ الغاضبةِ، وغضب مسيطر عليه، وهو المحاولة للسيطرة على تعبير الغضب. عموماً، تراكيب العوامل الثلاثة المُفتَرَضة لتعبيرِ الغضبِ أُكّدَت في عِدّة نماذج أمريكية (سبيلبرغر، 1988) نموذج لصينيين سنغافوريين (تانزر، سِيم، سبيلبرغر، 1996) وكذلك في التكييفِ إلى الألمانية (سكوينكميزجر، هوداب سبيلبرغر، 1999)، الإيطالية (سبيلبرغر وكومونييان، 1992) والنرويجية (هاسيث، 1996) على أية حال، في بَعْض الدراساتِ (مثال، سبيلبرغر وكومونييان، 1992، تانزر و سِيم و سبيلبرغر، 1996)، إن المادة "أنا ناقد للآخرين تماماً بشكل سري " قد انتقلت من الغضبِ الداخلي إلى الغَضب الخارجي.

تعريف المشكلة. من المُهمِ مُلاحَظَة (التوجيه 13) بِأَنَّ هذه المادة يُمْكِنُ أَنْ تُترجمَ إما كـ" إخفاء ضغائن وعدم التحدث عنها للناس الآخرين" الذي سيكون عندها تعبيراً عن غضب داخلي. أو يمكن ترجمتها كـ " التحدث بشكل سلبي بغياب شخص ما" وهذا قد يحمل تعبير غضب خارجي خفي، بدلاً من ذلك. في هيئة لدراسة التقاطع الثقافي، تم إعطاء طلبة كليات ناطقين بالألمانيةَ والإيطاليةَ بيان (STAXI بيان حالة التعبير لنبرة الغضب) بصيغة لغتهم الأصلية.

في إعادة الاختبار، سئلوا كيف سيُترجمونَ المادةَ. بالرغم من أن حوالي نِصْفِ الطلابِ الناطقين بالألمانيةِ 53 ٪، (n=347) اختاروا البديل الأولَ، ثُلث فقط مِنْ طلبةِ الكليات الناطقين بالإيطاليةِ 64 ٪، (n=241) اختاروا هذا البديلِ. هكذا، بإعطاء غموض لمحتوى المادةَ، حتى الاختلافات البسيطة بين صيغ اللغةِ المختلفةِ يُمْكِنُ أَنْ تُسبّبَ نقلات كبيرة في انتشار التفسيراتِ البديلةِ (توجيه 7).

بالرغم من أنَّ تعبير الغضب والتغلب عليه مشاعر عالميةَ، لكنها على الرغم من هذا هي محكومة بالعواملِ الثقافيةِ (مسكيتا وفريجدا، 1992). على سبيل المثال، إن المفاهيم "الحاجة إلى الانسجام "و" تَعطي وجهاً" سائدة في المجتمعاتِ الصينيةِ

(ومثال على ذلك: -، بوند، 1990؛ شونغ وآل، 1996؛ غاو، 1998) تُمْنَعُ المجابهات المفتوحة والمباشرةَ التي تُميّزُ الطرقَ الأمريكيةَ للتعبير عن الغضبِ الموجه للخارج.

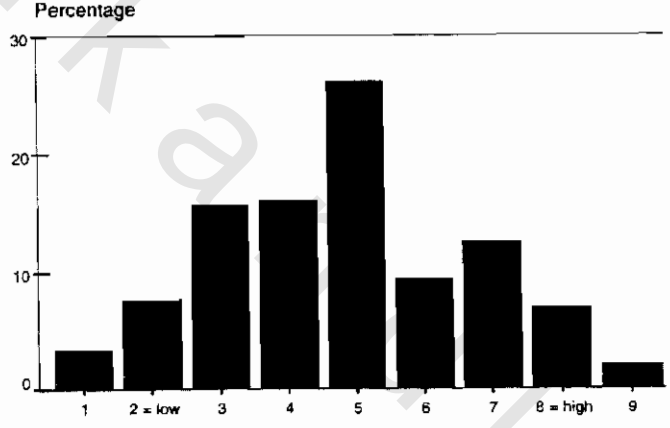
هكذا، نمط غير تجابهي غير مباشر للتعبير عن الغضبِ كما هو متضمن في التفسيرِ الثاني الذي سيتم تفضيلهُ. في هذه الحالةِ، إن النقلات المُلاحَظة تَعْكسُ اختلافات ثقافية أصيلةً في طريقة إبداء تعبير الغضبِ الموجه للخارج (توجيه 3).

المواد المثبتة الجدلية (المثال 2). إن استبانه إجهادِ العمل JSS، سبيلبرغر وريهيسر، (1994) هو تقرير ذاتي صمّم لقيَاْس تأثيرِ الإجهادِ المتعلّق بالعمل. ويتألف من 30 بياناً الذي يصفُ أحداثَ متعلقة بالعمل وصفت بالمرهقة مرهقة من قبل مستخدمين في وظائفِ مُخْتَلِفةِ (اجتهادات Job Stressors) وكُلّ ضغط عمل تم تقديره مرتين: تَقدير متعلق التكرار الذي حدث فيه خلال الستة أشهر الأخيرة (تقدير التكرار). أما الثاني فهو كمية أو درجة الإجهادِ الذي أُثير بسبب حدث واحد من الأمور المسببة للضغط (معدلات "الشدة" Severity ratings). ومن أجل تقديرات الشدة، تم توجيه الموضوعات لمُقَارَنَة كميةِ الإجهادِ المترافقة مع كل مسبب للضغط، بكميةِ الإجهادِ المثار من قبل مسبب للضغط " مهمة الواجبات المزعجة"، تم اختيارها كمرساة القيمة المسبقة التخصيص للشدَّةِ المتوسّطةِ.

تعريف المشكلة. أثناء تكييف استبانة إجهاد العمل JSS، ، تم دراسة الترجمة الألمانية للمثبت بتحويل هذا الإجهاد إلى مادة حرة التقدير (التوجيهات 8، 9، 13). على الرغم مِنْ معدل التقديرات المتوسط، تلقت هذه المادة تقديرات جدلية بثبات في عدد مِنْ الدراساتِ بغض النظر عن صيغ الرَدَّ والتعليمات المستعملة (هودابو تانزر، ماير، بيستيمر، وكورونكا، في الصحافةِ). علاوة على ذلك، تم إيجاد ثبات عال متعلق بالاختبار وإعادة الاختبار لهذا الإجهاد في دراسة لائحة ($r_{tt} = .68; N = 156$) أشارت بأنّ الفرقَ الواسعَ للتقديراتِ (انظر الشكل 10) لا يمكن أن يعزى ببساطة إلى التأثيرات العشوائية. في الدراسةِ الأخرى التي استعملت هذه المادة كمرساة مع قيمة مسبقة

التخصيص لشدة متوسطة، لوحظ أن عدداً من الموضوعات لَمْ تمتثل إلى الأمر. عندها ألغوا تَقدير الشدَّة المتوسط المسبق التخصيص، وصنفت كفئة تقدير مختلفة بدلاً من ذلك. باختصار، إن النتائج التي تشير إلى أنَّ هذه المادة لَمْ تُستدعْ كمية الإجهاد نفسها لكُلّ المستجيبين الناطقين بالألمانية، يعد طلباً ضرورياً قبل أن يمكن تأسيسه كمادة مثبتة صادقة.

الحل المقترح. اجمع "معلومات عرضية " (مثال، تفسير الممتحنين لمحتوى المادة في المثال الأول، وطرق غير قياسية من إدارة الاختبار في المثال الثاني) لاختبار صدق الاختبار المتعدد الثقافة/ المتعدد اللغة كمادة مثبتة (التوجيه 14).



الشكل رقم 10-1

توزيع معدلات الشدة لإجهادات العمل في استبانة إجهاد العمل «تحديد المهمات غير المتفق عليها» على مقياس من ٩ نقاط.

مخاطر سببها تصاميم الاختبار غير المتوافقة ثقافياً

هناك رأي واسع الانتشار بين العلماء النفسانيين السائدين، وهو أن المشكلات والمخاطر التي تمت مصادفتها أثناء تطبيقات الاختبار المتعدد اللغات، سببها بشكل رئيس استعمال المادّة الشفوية (مثال، في بيانات التقارير الذاتية أو اختبارات القدرة الشفوية مثل مقياس ذكاء فكسلر للأطفال/ مقياس تكييف ذكاء فكسلر للبالغين)

بشكل مكثف. ولذلك استخدام اختبارات "لا شفهية" مثل اختبارات جداول رافن المتعلقة بالمقاييس التقدمية (SPM) في تقييم القدرة، أو الاختبارات "التصويريةِ" في تقييمِ الشخصيةِ، سيخفض بشكل كبير مشكلة تطوير أداة صادقة للتقييمات المتعددة الثقافات/ المتعددة اللغات. تُبيّنُ الأمثلةُ التالية بأنَّ هذه لَيستْ بالضرورة القضيةَ.

التوافق الثقافي للمادّةِ التصويريةِ (المثال3). يستخدم عدد منْ الاختبارات في الشخصيةِ و الصحة العقليةِ، محفّزاتِ تتألف بشكل خاص منْ مادّةِ تصويريةِ. الأمثلة هي اختبار إبطال الصورة لروزينزويغ (مثال، روزينزويغ، 1977)، علامات ما قبل المدرسة للتقرير الذاتي (مارتيني، سترايهورن وبويغ-انتيك، 1990) و والتقدير التصويري لردات فعل الاختبار توبيانا، (1994) باعتبار أن كل المولد تشمل عناصر محددة لثقافة معينة (مثال، الأسلوب الغربي لارتداء الملابس، تصفيفات الشعر أو الملابس الثنائية الجنس). مدد الوقت أو مجموعة الأخلاقيات (مثال، وجوه بيضاء). إن تطبيقاً صادقا لتقاطع الثقافات، يَتطلّبُ إعادة رسم كاملة لكُلّ المادّة التصويرية (التعليمات 1، 3، 6). بناء على ذلك، يجب أن يتم إنشاء اختبار جديد كلياً، ويَجبُ أنْ يُصدّقَ لكُلّ ثقافة جديدة.

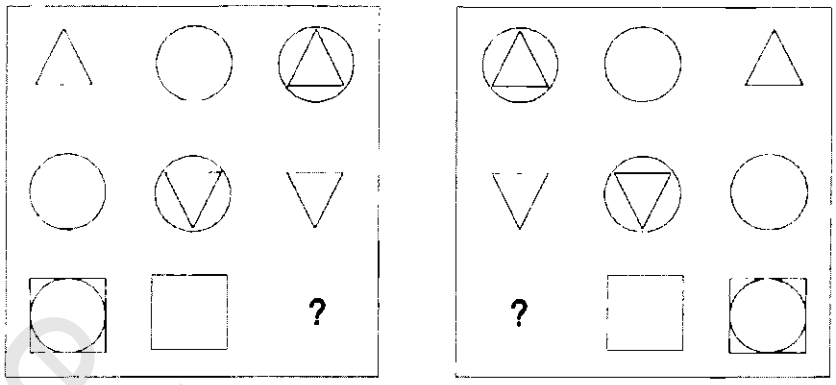
كتابة التعليمات (مثال 4). في دراسة للتقاطع الثقافي (بيسوانجر، 1975، انظر أيضاً إلى فيشر و فورمان، 1982) واختبار القوالب الفانيسي VMT (Viennese Matrices test)، فورمان وبيسوانجر، 1979 قدمت إلى طلابِ المدارس العليا العربِ والنايجيريينِ والتوجوليينِ. وقد تم مقارنة إجاباتهم مع آخرين من نموذج المعايرة الأسترالي. إن VMT هو اختبار للمنطق الاستقرائي يستعمل أسئلة جداول مماثلة لـ SPM الخاص برافن. وقد أظهر الاكتشاف الأكبر المفاجئ أن تحديد وتطبيق القاعدة من اليسار إلى اليمين كمقارنة للأعلى وللأسفل، كان أكثر صعوبة للطلاب العرب الأفارقة منْ الطلابِ النمساويينِ.

تعريف مشكلة. إن المزاوجة على نحو غير ملائم بين اتجاه كتابة في الثقافة العربية (وبمعنى آخر: من اليمين إلى اليسار في العربية، يقابلها من اليسار إلى اليمين في اللغة اللاتينية) والطريقة الغربية لتَصميم أسئلة الجداول بالعنصر المفقود في الزاوية اليمين السفلية (انظر إلى الجانب اليساري من الشكل 10-2) يعد تفسيراً معقولاً. لتوضيح النتائج الإدراكية التي يمكن أن يكون سببها تصميم الأسئلة غير المتوافقة ثقافياً، تخيل ارتباك عرب يقرؤون إعلاناً مؤلفاً من ثلاثة صور لآلة غسيل. عند ترتيب قراءتهم من اليمين إلى اليسار نجد امرأة (أ) تبدو سعيدة عند كومة من الغسيل النظيف. (ب) تضع الغسيل في الآلة وتديرها، ثم (د) بعد ذلك تفزع من أكوام الغسيل القذر.

الحل المُقترح. إن استعمال تصميم منسجم مع الثقافة لأسئلة الجداول، قد يتغلب على هذه المشكلة (التوجيهات 5، 14، 15). كما هو موضح في جهة اليمين للشكل 10-2، صيغة الاختبار "معكوسة" حيث تم ترتيب عناصر الجدول بطريقة يكون فيها العنصر المفقود عند أسفل زاوية اليسار، قد يزيل هذا النوع من التحيز للممتحنين العرب.

معنى المربكات. (المثال 5). مثال آخر لتصميم المادة غير المتوافقة ثقافياً تم تقديمه من قِبل غرينفيلد (1997). إن معنى الإرباكات في الاختبارات المتعددة الخيارات يمكن أن تُفسر بواسطة موضوعات من عدة ثقافات بشكل مختلف (التوجيهات 5، 14):

في الصيغة المتعددة الخيارات، أعطي المجيب مجموعة من الخيارات. كلها باستثناء واحد، وُظفت لتكون معلومات عديمة الجدوى. المشكلة هي أن المشاركين من عدة ثقافات سوف يفترضون أن واضع الاختبار يقدم مجموعة من المعلومات تتعلق بهدف حل المشكلة..... إن واضع الاختبار (ومطوره) يفترضان أن هدف الإجابات البديلة هي إزالة الاحتمالات غير الصحيحة، أثناء اختيار البديل الصحيح. يفترض الممتحن أن هدف الإجابات البديلة هي إنشاء حل للمشكلة. (ص: 1120-1121).



نموذجي　　معكوس

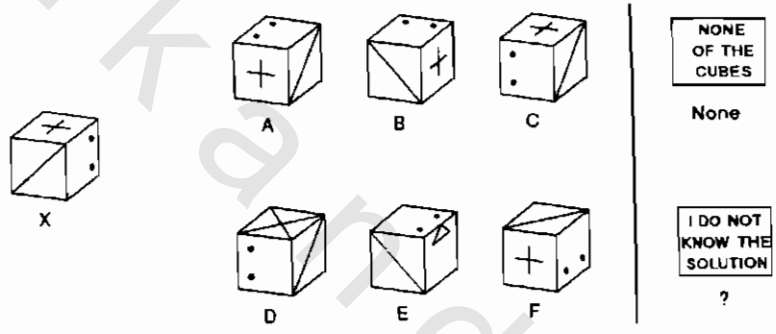
الشكل 2.10

بند اختبار معياري (بدون أسئلة الخيارات المتعددة التي تصرف الانتباه) في حاليتها النموذجية والمعكوسة

الحل المُقترح. كما أشار غرينفيلد (1979، تحمل المشكلة الحل بذاتها. لا تستعمل صيغ اختبارات متعددة في التطبيقات المتعددة الثقافات/ المتعددة اللغات، إلا إذا كان كل المتحنين مطلعين بشكل كامل على مفهوم المربكات. بدلاً من ذلك، دع المتحنين يستنتجون (مثال: استدرج) بدلاً من اختيار الحل.

تأثيرات الممارسة الضمنية للاختبار. (المثال 6). معتمدين على معارفهم وخبراتهم السابقة بالموضوعات المشابهة لأسئلة الاختبار (التوجيه 6)، سيستفيد المتحنون بشكل مختلف من تأثيرات الممارسة الضمنية للاختبار. خاصة في بداية الاختبار (توجيه 14). أُخذت الصورة الإيضاحية التالية من دراسة متعلقة بتقاطع الثقافة لاختبار مقارنة المكعب الثلاثي الأبعاد 3DC، غيللر، (1990) وهو اختبار مقارنة مكعب يقيس القدرة المكانية. نجد مثالاً لسؤال منه في الشكل 3.10 هذه الدراسة المتقاطعة الثقافة (برويز، 1996، تانزر وغيتلر وإيليس، 1995، تانزر وغيتلر وسيم، 1994) أشارت إلى أن الطلاب القادمين من البلاد التي ليس لديها تعليم رسمي في الهندسة الوصفية، مقارنة بنظرائهم النمساويين، حصلوا أكثر من العمل على أسئلة الاختبار القليلة الأولى.

الحل المقترح. استعمل عدداً كافياً من الأسئلة الإحمائية الخفية، أي أسئلة مقدمة كالأسئلة الحقيقية في بداية الاختبار ولكن ليس لاستخدامها من أجل الدرجات. في صيغة المصدر الأحادي الثقافة لـ 3DC غيللر، (1990)، تم استعمال الاختبار الأول كسؤال إحمائي مخفي، بينما استعملت أسئلة الاختبار الـ 17 المتبقية للدرجات. ولكن دراسات التقاطع الثقافي المذكورة سابقاً، قد أشارت الى أنه يلزم المزيد من الأسئلة الإحمائية الخفية، لتأمين مقارنات درجات صادقة في السياقات المتعددة الثقافات (توجيهات 20-12).



الشكل 10-3

نموذج بند استخدم في إرشادات 3DC الموجودة في الملحق

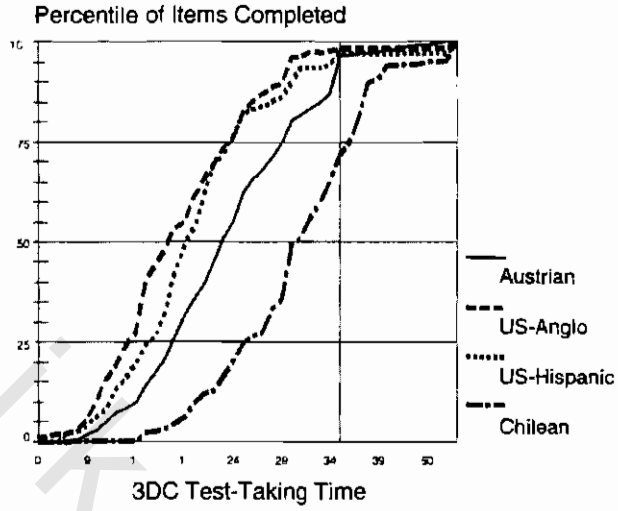
لاحظ أنه لا يمكن أن تحل هذه المشكلة بمجرد تقديم المزيد من أسئلة الممارسة مع حلولها المعطاة؛ وذلك لأن معظم الممتحنين يعالجون أسئلة الممارسة بشكل مختلف عن أسئلة اختبار حقيقية. يمكن تحديد العدد الضروري من الأسئلة الإحمائية الخفية، عبر إعطاء الممتحنين بشكل عشوائي صيغة اختبار نموذجية مقابل صيغة اختبار بترتيب أسئلة معكوس، ثم مقارنة صعوبات السؤال بين صيغتي الاختبار (توجيهات 8، 9، 22).

المخاطر المسببة بواسطة تعليمات الاختبار العرقية وإجراءات الإدارةِ

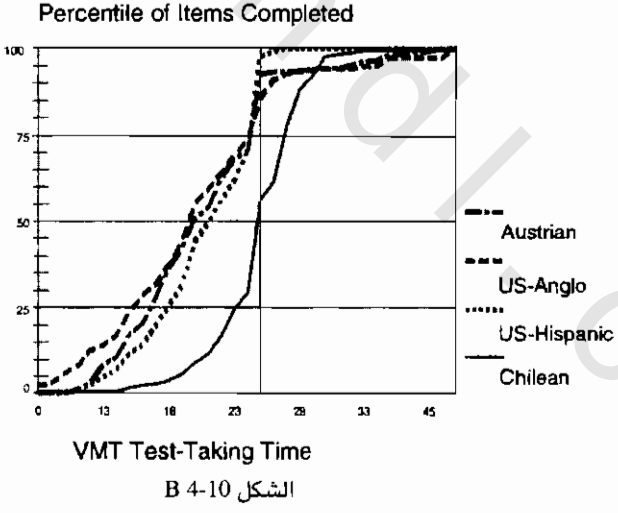
مصدر آخر للتحيز العرقي هو اختبارات الذكاء/ الأهلية غير الشفهية التي قد تكون مخفية في تعليمات الاختبار وإجراءات الإدارةِ. على سبيل المثال، إن نقاط أفضلية الممتحنين على منحنى الخط البياني لتبديل الدقة والسرعة، قد يتأثر بقيم ثقافية معينة متعلقة بالوقت والدقة.

فرض حدود الوقت في اختبارات القدرة (المثال 7).

على الرغم من أن VMTو 3DC قد صممت كاختبارات استطاعة (أي الكفاءة الذاتية)، فإن محرري الاختبارات قد أوصوا بحدود زمنية متسامحة لإدارة اختبار اقتصادي. في دراسة التقاطع الثقافي مع هذين الاختبارين (برويز، 1996، تانزز وإيليس وغيلر وبرويز، 1996، تانزز وآل)، فإن حد الوقت الموصى به والمقدر بـ25 دقيقة لـ VMT و35 دقيقة لـ 3DC، قد تم زيادته لإعطاء كل الممتحنين وقت اختبار متسعاً (توجيه 13). كان المشاركون طلاب جامعات أو كليات من النمسا (N=244)، تشيلي (N=173)، والولايات المتحدة مع أنجلو- أمريكي واحد (N=196)، ونموذج إسباني واحد (N=144) كما يمكن رؤية ذلك في الشكل 4.10، فإن نسبة كبيرة من الممتحنين التشيليين قد تجاوزا كلاً من حدي الوقت الموصى بهما.



الشكل A4-10



الشكل B 4-10

النسبة المئوية للأشخاص الذين أكملوا اختبار 3DC (الشكل A4-10) واختبار UMT (الشكل B4-10) مرسومة بالمقابل لزمن الاختبار (بالدقائق للنمساويين (N=244) والتشيليين (N=173) والإنجلو أمريكان في أمريكا (N=196) والهسبانيك في أمريكا (N=144).

بالتالي، فإن فروقات التقاطع الثقافي التي تم تحصيلها تحت تعليمات نموذجية ستكون مهمة. وبذلك، تشير بشكل غير صحيح إلى أن المجموعات تختلف في مهارات المنطق الاستقرائية (التوجيهات 14-22). باعتبار أن حدود الوقت في اختبارات القدرة يمكن أن تضر بمجموعات ثقافية معينة، فنحن نوصي بشدة تسجيل الوقت المستغرق في الاختبار وذلك للاستخدام في "المعلومات الإضافية" (توجيه 20) في تطبيقات اختبارات التقاطع الثقافي لاختبارات القدرة.

قابلية ترجمة تعليمات الاختبار (المثال 8). إن التعليمات في الاختبارات الأحادية الثقافة/ الأحادية اللغة، غالباً ما تنسجم بشكل جيد مع التطبيقات المقصودة الأحادية الثقافة/ الأحادية اللغة. وهكذا، فهي تميل إلى استغلال الخاصية اللغوية (التبحر في اللغة). وقد قام بريسلين (1986) بصيغة عدد من التوجيهات لتحسين قابلية الترجمة للمواد الشفهية التي يمكن أن تطبق أيضاً على صيغ تعليمات الاختبار (التوجيه 4).

توثيق ما بعد التعليمات (المثال 9). عادة ما تُطور تعليمات الاختبار عبر سلسلة من النسخ و" المعرفة-الكيفية" التي تم إحرازها في أثناء عملية (مثال، جزء معين من المعلومات، عناصر مهمة للتعليمات، خلفية تعليمية مطلوبة للفهم، إلخ) التطبيق في النسخة النهائية. ولكن معرفة الكيفية هذه نادراً ما توثق. وهكذا، عادة ما تكون غير متوفرة للجنة العمل المعنية في التكييفات التسلسلية للاختبار هذه، مع النسخة النهائية للتعليمات (التوجيه 17). وهذا يمكن أن يطبق كتوثيق معلومات مخزنة في حاسوب (مثال، كونكلين، 1987)، كما هو موضح في فهرس التقييمات الإنكليزية لـ DC 3 .

التكييف المتعدد الثقافات/ المتعدد اللغات لاختبار التركيز (المثال 10). إن كلاً من الأمثلة السابقة قد ركز على خطر واحد فقط. لإلقاء الضوء على عدد من المشكلات، قد يواجه المرء، حتى بالاختبارات البسيطة نسبياً، المثال التالي الذي يقدم صورة عن تكييف الاختبار المتعدد الثقافات/ اللغات.

النسخة الأصلية الألمانية. إن d2 بريكن كامب وزيلمر،1998 هو اختبار Bourdon الذي يأخذ شكل رسالة مختزلة. وهو مصمم لقياس التركيز القصير الأمد. تم تطويره في الأصل في ألمانيا (بريكن كامب، 1962) وما يزال يستخدم بشكل متكرر في البلاد الناطقة باللغة الألمانية. يتوجب على الممتحنين أن يقوموا بمسح 14 صف hW من d's و p's مع القواطع فوق و/ أو تحت الإشارات المثبتة بأسرع ما يمكن دون ارتكاب أخطاء. إن أنواع الإشارات الثلاثة هي (أ) d's مع قاطعتين فوق، (ب) d's مع قاطعتين تحت، و(د) مع قاطعة فوق وقاطعة فوق. تتألف مادة الاختبار الأصلية من ورقة اختبار مصورة مزدوجة الجهة، مع تفصيلات شخصية، أحرف الإشارات، أسطر كمثال، وجدول العلامات اليدوية على الصفحة الأمامية (الصفحة اليمنى) و4 سطر للاختبار مع مؤشرات العلامات على ظهر الصفحة (صفحة اليسرى). تتم إدارة d2 ضمن شروط السرعة (مثال، مُؤقت) بحوالي 4 دقائق لوقت الاختبار. وتعطى كل التعليمات بشكل شفهي.

تطويرات متعاقبة لنسخ اختبار لا مركزي:

بدأ فريق من علماء النفس المتعددي الثقافة/ المتعددي اللغة (تانزر، إيليس وآل، 1997) بتطوير صيغ اختبار لا مركزية ثقافياً ولغوياً. ويجب أن تكون صادقة تحت الشروط التالية المحددة مسبقاً:

(أ) يجب أن يكون لدى كل الممتحنين 8 سنوات على الأقل من التعليم الرسمي، (ب) يمكن أن لا يكون لدى الأغلبية منهم خبرة مسبقة باختبارات Bourdon المختزلة، (د) يمكن أن يجري الاختبار ضمن مجموعة صغيرة حتى المجموعات المتوسطة الحجم، (ج) لا يجب أن تعطى التعليمات باللغة الأصلية للممتحنين، (ح) يمكن أن يكون واضعو الاختبار: (1) علماء نفس أجانب مدربين، يملكون خبرة واسعة في التقييمات النفسية وفي إدارة d2، ولكن بلهجة أجنبية قوية وخبرة قليلة نسبياً بالظروف المحلية. (2) علماء نفس محليين مدربين، مع معرفة واسعة

بالظروف المحلية وخبرة في التقييمات النفسية بشكل عام، ولكن ليس بإدارة d2 بشكل خاص. أو (3) طلاب في مرحلة التخرج مع خبرة متنوعة في الاختبارات النفسية.

استناداً إلى الإجراءات الحكمية (توجيه) والبرهان التجريبي في سلسلة من دراسات التقاطع الثقافية، مع نماذج من 150–250 لطلاب كليات أو جامعات، تم تطوير الصيغ اللامركزية للغة الصينية، اللغة الكرواتية، اللغة الإنكليزية (المستخدمة في الولايات المتحدة ومن قبل النماذج الأنجلو أمريكية والإسبانية)، اللغة الفرنسية (المستخدمة من قبل النماذج النمساوية) ولهجات إسبانية متعددة. بافتراض السهولة في d2، قد يبدو أن الترجمة الصحيحة لتعليمات الاختبار ستكون كافية. ولكن، كما هو موضح لاحقاً فإن عدداً من التغيرات في تصميم مادة الاختبار، تعليمات الاختبار وإجراءات الإدارة، كانت ضرورية من أجل ضمان تكافؤ التقاطع الثقافي واللغوي.

تصميم ورقة الاختبار (**المثال** A10) تم تبسيط النموذج الطباعي للاختبار الأصلي عبر تحريك جدول الدرجات اليدوية وكل حقول البيانات الحياتية باستثناء الرمز الشخصي من تعليمات الصفحة (الصفحة اليمنى) ومؤشرات الدرجات من صفحة الاختبار (الصفحة اليسرى). من أجل التأكيد على أن كل المتحنين سوف يفهمون التعليمات تحت كل الشروط المذكورة مسبقاً، تم تقديم الجزء المركزي من التعليمات بصيغة مكتوبة على صفحة التعليمات. هذا يمكننا أيضاً من تحريك التفسيرات " ك: كما في كلمة كلب" و "خ: كما في كلمة خنزير" من التعليمات "الشفهية " المقدمة في النسخة الإنكليزية (بريكن كامب وزيلمر، 1998) وهو شيء مربك أكثر منه مساعد في الثقافات التي لا تستخدم حروف الأبجدية اللاتينية (مثال، الصين). في النهاية تم إبراز حروف الأهداف الثلاثة في مربع عُلِم مرتين على سطر النموذج.

الإدارة (المثال، B10) لضمان إدارة ثابتة تحت شروط الحقل التي تم ذكرها سابقاً، جرى تغيير الوقت المجزأ من 20 دقيقة لكل من السطور الـ 14 إلى دقيقتين لكل من مجموعتي (جزءA وجزء B) من السطور السبعة. بالإضافة إلى ذلك، تم وضع قائمة مفصلة من التوجيهات لإدارة الصيغة اللامركزية في السياقات المتعددة الثقافة (التوجيهات 8، 17).

التعليمات (المثال، C10) باعتبار أن عدداً من المختبرين استعملوا طرقاً مختلفة لتعليم الأهداف (مثال، الدرجات الصغيرة. التقاطعات، الخربشات والدوائر)، تم تأكيد وتبرير ("لتوفير الوقت") الطريقة المطلوبة " للتعليم بشحطة واحدة"، ووضحت مرتين في سطر النموذج. بطريقة مماثلة، تم معايرة الطرق المختلفة لتصحيح التعليم الخاطئ (مثال، شطب، خربشة، محي بالممحاة) إلى "الشطب" (X-out في النماذج الأمريكية)، ووضحت على السبورة أو على الزجاج المضاء.

وُجه أيضاً اهتمام خاص إلى ترجمة مفاتيح العبارات. في النسخة الإنكليزية، تمت الإشارة إلى الأنواع الثلاثة من الإشارات المحتملة (أي: كل الـ d's مع مجمل التقاطعين) مثل " أمثلة" (بريكن كامب وزيلمر، 1998) ولكن لأن تعبير: أمثلة لا يحمل معنى مجموعة مستنفذة، تم تغييرها إلى "هدف". علاوة على ذلك، فإن الصيغة الغامضة لـ:شطب" استخدمت بشكل متكرر (مثال، "لا يفترض شطب الحروف الأخرى" " الحرف خ كما في خنزير" يجب أن لا تشطب أبداً) في التعليمات الشفهية. إن الطريقة المطلوبة ("للشطب..... بإنشاء خط واحد عبر الحرف") تم تحديدها فقط مرة واحدة. باعتبار أن التعبير "شطب" يمكن أن يفسر إما كطرقة معينة للتعليم (أي، الاستبعاد " X-out") أو كمفهوم أكثر عمومية للحذف/ الإلغاء، يتضمن أنواعاً أخرى من التعليم ("حذف"، "محو")، فقد تم استبداله بعبارة أكثر دقة (التعليم بشحطة واحدة). أخيراً، إن عبارة "توقف" تابع مع B، التي أعطيت بعد أول دقيقتين قد جعلت بعض الممتحنين يتوقفون بشكل حرفي عن الحل وهكذا، تم استبدالها بـ " الوقت" تابع مع B"

حدود تكييفات الاختبار المتتابعة: بينة للتطوير المتزامن

في كل الأمثلة المقدمة، لم يتمكن مترجمون محترفون (تطبيقات الاختبار، فان دي فيفر ولونغ. 1997a و1997 b من تقديم أدوات صادقة متعددة الثقافات/ متعددة اللغات. حتى في حالة الاختبار البسيط نسبياً مثل d2 (مثال 10)، كان عدد من التعديلات الجوهرية ضرورياً " لتكييف الاختبار. قد يتطلب تكييف الاختبارات تعديل سؤال واحد (مثال 1)، وصيغ الإجابة (الأمثلة 2 و5)، وتصميم الأسئلة (مثال 4) وورقة الاختبار (مثال 10a)، وقواعد الدرجات (مثال 6) وإجراءات الإدارة (الأمثلة 7 و 10b)، وتعليمات الاختبار (الأمثلة 8، 9، 10c) وفي حالات أخرى (مثال 3)، قد يكون من الضروري تكييف الأداة إلى حد وضع أداة جديدة تماماً "تركيب اختبار" بشكل خاص.

في الواقع، على الرغم من الجهود الكبيرة لإنجاز صيغة ثقافية ولغوية لامركزية لـ d2، تبقى المشكلتان الجادتان التاليتان اللتان يمكن حلهما بإكمال إعادة تصميم الاختبار (أي تركيب الاختبار).

حروف أهداف لاتينية (المثال 11). إن استخدام الحروف اللاتينية (d's و p's) في d2 يمكن أن يضر المختبرين الذين لا يستخدمون حروف الأبجدية اللاتينية في حياتهم اليومية. من جهة أخرى، بإمكان المختبرين المدُربين على حفظ وتمييز الأحرف المعقدة مثل أحرف اللغة الصينية الشمالية، أن يقوموا بغربلة السطور بوقت واحد للمراحل الثلاثة بدلاً من مسح d's أولاً ثم مراجعة عدد التقاطعات. لاحظ أيضاً أن اختبارات إلغاء الحرف هي أمثلة نموذجية لتراكيب الاختبارات العرفية لأنه ليست هناك حاجة فعلية للاعتماد على هذا النوع من المنبهات (غيتلر وتانزر، 1990/1998، موسبراغر أولسشلاغيل، 1996).

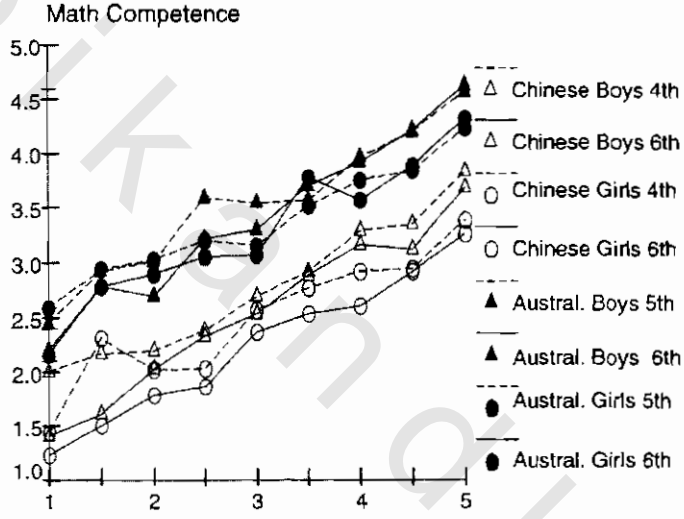
وجهة العمل (المثال 12). لم يمتثل عدد من الممتحنين من النماذج الغربية للتوجيه " اعمل على حل السطور من اليسار إلى اليمين" بل حلوا بشكل متعرج بدلاً

من ذلك. لم تلغ هذه الظاهرة أيضاً بالتوجيه الإضافي " قم بالحل كما لو كنت تقرأ كتاباً" (استخدم في النماذج الغربية). وهكذا، لم يكن من المحتمل أن هذا نتج عن سوء فهم التعليمات. بل كان السبب، على الأغلب، الممتحنين الذين يستعملون اليد اليمنى. وهؤلاء، بخلاف الذين يستعملون اليد اليسرى، تمت إعاقتهم بهذا الاتجاه من الحل؛ لأن يدهم غطت بقية الأحرف. تفاقمت هذه المشكلة في سياق متعدد الثقافة، لأن وجهة الكتابة الموصى بها قد أضرت بالممتحنين العرب (انظر أيضاً إلى المثال 4).

إن الخطر الموضح في الأمثلة السابقة له أثر فقط على صدق التقاطع الثقافي للقياس (أي الاختبار)، ولكن ليس على صدق التقاطع الثقافي للتركيب الأساسي. يوضح المثال التالي أن "تحويل الاختبار" (غرينفيلد،1997) من ثقافة أحادية إلى سياق متعدد الثقافات، يمكن أن يؤثر حتى على تكوين المفاهيم عن التركيب الأساسي (التوجيه2).

العناصر الضمنية للمفهوم الذاتي الأكاديمي (المثال 13). تكشف أبحاث المفهوم الذاتي الأكاديمي باستمرار أن القراءة وتقارير المفاهيم الذاتية الرياضية شيئان غير مترابطين (أي مجال محدد). في معظم هذه الدراسات، تم قياس المفهوم الذاتي الأكاديمي بواسطة الاستبيان الوصفي للذات (SDQ-I، مارش، 1988) الذي استخدم معايير تتألف من أسئلة إدراكية وعاطفية/ محرضة لقياس المفهوم الذاتي الأكاديمي في مجال محدد. في مقارنة للتقاطع الثقافي لطلاب نمساويين وسينغافوريين وصينيين (تانزر، 1995، 1998، تانزر وسيم، 1991 وتانزر وسيم ومارش، 1992، 1997) برزت اختلافات تتعلق بالتقاطع الثقافي في النسب المتبناه لأسئلة الكفاءة/ سهولة المهمة، ولكن ليس في أسئلة الاهتمام والاندفاع (انظر إلى الشكل 10-5). باعتبار أن النسخة الإنكليزية الأصلية لـ SDQ-I استخدمت في كلا البلدين، "تقييم متعدد اللغات/متعدد الثقافات"، فإن مشكلات الترجمة ليست بالتأكيد سبب هذه الظاهرة (التوجيهات 16.7).

إن تحليلات عوامل التداخل الثقافية اللاحقة في أستراليا (مارش، كرافن وديوس، 1999) النمسا وإيطاليا وسنغافورة والولايات المتحدة (تانزر، 1998) قد دعمت أيضاً فصل العناصر الضمنية الإدراكية (أي الكفاءة، سهولة المهمة) والعاطفية/ المحرضة لقياس المفهوم الذاتي الأكاديمي في مجال محدد. وهكذا، إن دمج أبحاث التقاطعات اللغوية والتداخلات اللغوية (أي، علم النفس الفارقي Differential Psychology) قد ساعد على اكتساب فهم أعمق للمفهوم الذاتي الأكاديمي.



الشكل رقم 5.10

انحدار جدارة الرياضيات في الاهتمام بالرياضيات ضمن المجموعات الثمانية الفرعية (العمر - الجنس - الثقافة) [الصينيون السنغافوريون مقابل الإستراليين]

يوضح هذا المثال أيضاً الحدود لطرق معالجة تطوير الاختبارات المتعاقبة: إن نتائج جديدة للتقاطعات اللغوية والتداخلات اللغوية تتعلق بتشكيل مفاهيم حول التركيب ومقاييسه، يمكن أن يتم دمجها فقط بلغة الهدف (أي المترجم إليها) ولكن ليس إلى نسخة المصدر. مثلاً، دَمج تكييف SDQ-I الألماني بشكل كامل مفهومي العنصرين للمفهوم الذاتي الأكاديمي في مجال محدد، بينما تقدم النسخة الإنكليزية الأصلية معايير "سابقة" فقط لهذه الغاية. علاوة على ذلك، فإن الجزء الكبير

للنظرية والبحث الذي جرى بواسطة الأداة الأصلية (مثال، بحث SDQ-I للمفهوم الذاتي الأكاديمي في مجال محدد، انظر الى مارش 1998، مارش وكرافن، 1997) قد لا يحافظ بشكل كامل على تركيب إعادة تشكيل المفهوم.

تصميم الاختبارات للاستخدام في الثقافات واللغات المتعددة:

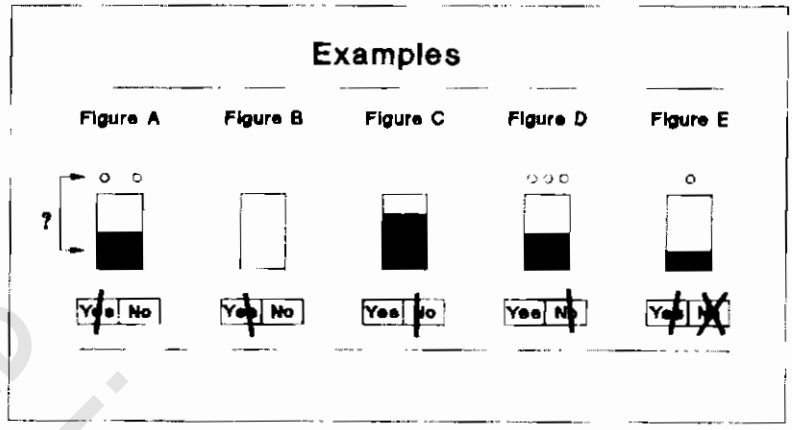
بافتراض قصور تكييف الاختبارات المتعاقبة، نحن نوصي بتطوير متزامن للاختبارات المتعددة اللغات والثقافات. بخلاف طريقة المعالجة المتعاقبة، يسمح تطوير الاختبار المتزامن بدمج النتائج من دراسات الثقافة المتقاطعة والثقافة المتداخلة؛ وذلك بإعادة صياغة إعادة تشكيل المفهوم للتركيب ومقياسه في مرحلة مبكرة نسبياً من تطور الاختبار. يمكن أيضاً لطريقة المعالجة المتزامنة، بمساعدة مراجع لغوية منتقاة بعناية، تقليص خطر تحيز التركيب بشكل كبير في سياقات التقييمات المتقاطعة ثقافياً (انظر إلى المثال 13)، وتسمح بعدم مركزية لغوية وثقافية بدرجة أكبر بكثير من تكييفات الاختبارات المتعاقبة.

التطوير المتزامن لاختبار تركيزي Concentration test (المثال14):

على نقيض الـ d2، تم تصميم سلسلة من الاختبارات المتعاقبة PTS، غيتلر وتانزر، 1995/1998 بشكل مسبق للاستخدام في الإعدادات المتعددة الثقافات واللغات. وهو يتألف من 9 اختبارات فرعية مع تعقيد معرفي متزايد، يتراوح بين الاختبارات الثانوية التي تتطلب فقط سرعة حسية وحركية، إلى الاختبارات الثانوية التي تستلزم عناصر من الذكاء السلس. مثل "الدمية الروسية". تتطلب المهمة المعرفية لكل من صيغ الاختبار الثانوية، جزءاً من الاختبار الثانوي التالي. وهكذا، تنتج بنية شبه مبسطة. يقوم هذا التصميم أيضاً بتسهيل تعليمات الاختبار، لأن الفرق بين اختبارين ثانويين، الذي يحتاج تفسيراً هو فقط.

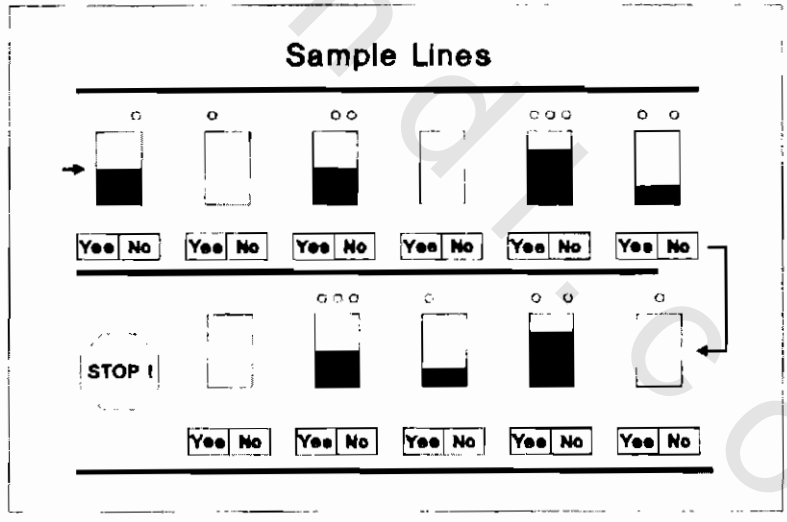
من أجل تجنب تحيز عرقي في مادة الاختبار الأصلية، تتألف أسئلة الاختبارات الثانوية التسعة من عناصر هندسية بسيطة (نقاط، دوائر، إطارات، شرائط، مناطق مظللة بالرمادي) لا بد أنها مألوفة للأشخاص المتعلمين في كل الثقافات (توجيه 6)، بطريقة مماثلة، إن خيارات الإجابات المعروضة ("0، 1، 2" "1، 2، 3"، أو " نعم/ لا") تحت كل سؤال، يمكن ترجمتها بسهولة إلى كل لغة (توجيه 5). بالإضافة إلى ذلك، فإن "معرفة الكيفية" المحصلة في تكييف d2 المتعدد الثقافة واللغة (انظر إلى الأمثلة 10-12) تم استخدامها لإعادة تصميم مواد الـ PTS (التوجيهات 13-14) كما هو موضح في الشكل رقم 6.10، تم استخدام المساعدات البصرية المتعلقة بالمهمة وتعليم تصحيح الإجابات. للموازنة بين استعمال يد معينة (أثناء الحل) وتوجيهات الكتابة في لغة المرء الأصلية (توجيه 1)، على المختبرين أن يعملوا خلال الاختبار بشكل "متعرج " كما هو موضح في الشكل 7.10. أخيراً، لضمان أنه حتى المختبرين ذوي التعليم الضئيل يمكن أن يفهموا المهمة المطلوبة (توجيهات 10.4)، يمكن أن تعطى التعليمات بذلك عبر "قصة غلاف واقعية". إن قصة الغلاف المعاد صياغته للاختبار الثانوي (انظر أيضاً إلى الشكل 6.10) هو كالتالي:

انظر إلى الإطارات [أو"النوافذ"] يمكن أن تكون فارغة أو معبأة من واحد إلى ثلاثة شرائط [أو"كتل مستطيلة"] رمادية. في الإطار الأول، يمكنك أن تحصي عدد الشرائط الرمادية. في الإطارات الأخرى، يجب أن نقدرهم من أعلى المنطقة المظللة بالرمادي. تشير الدوائر فوق كل إطار إلى عدد الشرائط التي يجب أن تكون في الإطار. إن مهمتك هي أن تقرر [نعم/ لا] لكل إطار إذا كانت الحالة كما هي الآن..



الشكل رقم 6.10

مثال لبند أُستخدم في اختبار C الثانوي السلاسل اختبار تدريجي (PTC)



الشكل رقم 7.10

مثال لبنود استخدمت في اختبار C الثانوي لسلاسل اختبار تدريجي (PTS)

خاتمة

إن تطوير الاختبارات للاستخدام في الثقافات واللغات المتعددة يستلزم الدمج الناجح لعدد من الاعتبارات النفسية، التقاطعات الثقافية واللغوية، من أجل ضمان صدق التقاطع الثقافي واللغوي للتركيب الضمني، وتصميم الاختبار لقياسه، وإدارة الاختبار والنتائج التي تم تحصيلها من درجات الاختبار. بافتراض القصور لإجراءات تكييف الاختبار المتعاقبة حتى في حالة الاختبارات السهلة نسبياً مثل d2، نوصي بمعالجة متزامنة لتطوير الاختبارات المتقاطعة ثقافياً ولغوياً؛ لأنها تسمح بالحد الأقصى من عدم المركزية اللغوية/ الثقافية. ولكن هذه الطريقة يمكن أن تضمن أدوات متقاطعة ثقافياً ولغوياً فقط إذا تحققت اللوازم التالية:

1- برهان لتكافؤ التركيب. من المتطلبات d2 الأساسية لتطوير أداة صادقة للاستعمال في اللغات والثقافات المتعددة هي قابلة التعميم للتراكيب الضمنية عبر كل مراجع الثقافات واللغات (توجيه2). بالإضافة إلى البينة التي تم جمعها من أدبيات الثقافة (الأصلية)، وعلم النفس عبر الثقافات (المقارن) وعلم الأعراق البشرية وعلم اللغويات، فإن كل من البيانات التجريبية والحكمية يجب أيضاً أن تجمع خلال عملية تطوير/ تكييف الاختبار (توجيهات 9.7).

2- تصنيف تكافؤ التقاطع الثقافي واللغوي. مطلب أساسي آخر هو تصنيف تكافؤ التقاطع الثقافي واللغوي. يجب أن يحدد هذا التصنيف تحت أية شروط يمكن لأنواع معينة من النتائج (مثال: تحميل العامل، علامات الاختبار) التي تم تحصيلها من نسخ لغوية مختلفة أو من مجموعة ثقافية (توجيهات 12-20-22). تم تقديم اقتراح شامل نسبياً لمثل هذا التصنيف من قبل فان دي فيفر ولونغ (1997a ،1997 b).

3- تصنيف التحيز المتعدد الثقافة/ المتعدد اللغة. مطلب أساسي إضافي هو تصنيف شامل لمصادر مختلفة من تحيز التركيب والقياس (يتضمن الطريقة،

الإدارة وتحيز السؤال). في تطبيقات الاختبارات المتقاطعة ثقافياً ولغوياً. يجب أن يحدد هذا التصنيف أي من أنواع التكافؤ من المرجح أن تتأثر، وبأية أشكال من التحيز. ويجب أن تؤكد بمجموعة كبيرة من التوضيحات التجريبية (التوجيهات 13-14). مع التصنيفات المقترحة من فان دي فيفر ولونغ (1997b، 1997a) أو فان دي فيفر وتانزر (1997)، يعد توجيهات تكييف الاختبار لـ ITC مع توضيحاتهم المنطقية والتجريبية خطوة تمهيدية نحو هذا التصنيف الشامل.

4- تصنيفات الحلول. إن تركيب "كتاب طهي" يحدد وصفات لكل المصادر التي ينشأ منها التحيز في التقييمات المتعددة الثقافة واللغة هو، على الأغلب، مشروع لا أمل منه. رغم ذلك، يحتاج التقدم المستقبلي في الاختبارات المتعددة الثقافات واللغات، تصنيفاً للحلول (توجيهات 1، 13) يستند على تصنيف متقاطع لأنواع مختلفة من التكافؤ ومصادر مختلفة من التحيز. إن تصنيف الحلول الذي اقترح من قبل فان دي فيفر وتانزر (1997) يعد خطوة أولى في هذا الاتجاه. ولكن من أجل تقديم رؤية أكثر عالمية، يجب أيضاً أن يتم دمج البيانات من نظرية القياس وعلم النفس عبر الثقافات (غرينفيلد).

5- البنية المنهجية لتقصي وتقييم ومعالجة التحيز. بدون منهجية ثابتة لتقصي مصادر التكافؤ، فإن التصنيفات المذكورة سابقاً للتحيز والحلول، لا يمكن أن تحول إلى تطبيقات عملية (التوجيهات 7-9، 11، 13). يجب أن تحدد البنية المنهجية المطلوبة أي الإستراتيجيات المؤثرة من أجل تقصي مصادر التحيز، وأيضاً لتقييم تأثير المقاييس المقابلة المأخوذة. "استراتيجيات صندوق الأدوات" هذه، يجب أن تتضمن إجراءات حكمية. وتقنيات قياسية (مطبقة في برامج كمبيوتر تتوفر بسهولة ومرنة مع المستخدم)، بالإضافة إلى تصاميم تجريبية (مثال، غيتلر وتانزر، 1998) يجب أن تشمل أيضاً توجيهات على مجموعة المعلومات الإضافية (مثال، تسجيل وقت العمل في سرعة الاختبار الذاتية أو

استخدام تقنيات" التفكير بصوت عالي") لأنها تعطي رؤية أعمق إلى مميزات ثقافية محتملة لعمليات تحفيزية ومعرفية (التوجيه 22).

6- الأرشيف الموثق. إن إمكانية الدخول إلى أرشيف كبير ومحفوظ بشكل جيد يقدم مستندات نصية مفصلة على تكييف وتطوير الاختبارات المتعددة الثقافة واللغة (التوجيهات 4-6، 10-11، 17-19، 22) سوف يساعد على التحديث المستمر للتصنيفات المذكورة سابقاً للتحيز وتصنيف الحلول، وعلى تقييم جدواها على أساس بينات جديدة. في حال كان هذا الأرشيف محفوظاً بشكل جيد، فسيكون بمقدور مطوري الاختبارات الاسترشاد بتصنيف الحلول لتحديث المقاييس المقابلة المؤثرة (التوجيه 13). إن الدراسة الحديثة لهامبلتون، ولي وسيرسي (2002) التي بين فيها الباحثون العديد من الأخطاء الشائعة التي وجدت في مطبوعات تكييف الاختبارات، هي خطوة أولية نحو الأرشيف المقترح.

7- المشروع التكميلي لمطوري الاختبارات. إن مطوري الاختبارات المتعددة الثقافة واللغة ليسوا فقط بحاجة لأن تكون لديهم الكفاءة في اللغات المتعددة وخبرة في علم النفس السائد (بما فيها المعرفة بالتركيب ومقياسه)، وعلم القياس وتقنيات تركيب الاختبار، ولكنهم بحاجة أيضاً إلى الكفاءة في علم اللغات والثقافة المتقاطعة والثقافة (أي الأصلية) النفسية، بالإضافة إلى العلم بالمميزات الثقافية (مثال، الموضوعات المحظورة، غودوين ولي، 1994). في كل المراجع الثقافية المعنية. من الواضح أنه لا يمكن وجود شخص يستطيع جمع كل هذه المجالات المختلفة من الخبرات؛ لذا، يجب أن يتم التركيز على كفاءة إجرائية تتعلق بإعداد وإدارة ناجحة للجنة عمل متعددة الثقافات واللغات ذات خبرة من الدراسات المكملة.

8- مؤهلات مستخدمي الاختبارات. بالإضافة إلى الخبرة في الإدارة، والدرجات، وترجمة اختبار معين، على مستخدمي الاختبارات أن يملكوا كفاءة عامة في إجراء التقييمات في السياق المتعدد الثقافة واللغة (التوجيهات 14، 15، 18).

وهذا يشمل حساً ثقافياً و كفاءة في التواصل عبر الثقافات (اسانت وغوديكونست، 1989، كوهين، 1987، لانديس و بهجت، 1996، سشنيلر، 1995، ويرلي، 1995).

واضح أن كل هذه المتطلبات تتجاوز المقدرات الفردية لمطوري ومستخدمي الاختبارات؛ لذلك، فإن تطوير وتأسيس البرامج المؤهلة لمطوري ومستخدمي الاختبارات، التي تغطي هذه المتطلبات، يجب أن يكون موضوعاً رئيساً للمنظمات العلمية والحرفية المهتمة بالاختبار النفسي والتربوي.

*****

obeikandl.com

obeikandl.com

المراجع

Asante, M. K., & Gudykunst, W. B. (Eds.). (1989). *Handbook of international and intercultural communication.* London: Sage.

Bartram, D., & Coyne, I. (1998). *The ITC/EFPA survey on testing and test use in countries world-wide. Narrative report* (Research Report). Hull, England: University of Hull, Psychology Department.

Bond, M. H. (1990). *Beyond the Chinese face. Insights from psychology.* Hong Kong: Oxford University Press.

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.

Brickenkamp, R. (1962). *Aufmerksamkeits-Belastungstest (Test d2)* [The d2 test of attention] (1st ed.). Goettingen, Germany: Hogrefe.

Brickenkamp, R., & Zillmer, E. (1998). *d2. Test of attention.* Seattle: Hogrefe & Huber.

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.

Broer, T. (1996). *Rasch-homogene Leistungstests (3DW, WMT) im Kulturvergleich Chile-Österreich. Erstellung einer spanischen Version einer Testbatterie und deren interkulturelle Validierung in Chile* [Cross-cultural comparison of the Rasch-calibrated tests 3DC and VMT between Chile-Austria and the development of a Spanish version of the test battery]. Unpublished master's thesis, University of Vienna, Austria.

Cheung, F. M. (2004). Use of western and indigenously-developed personality tests in Asia. *Applied Psychology: An International Review, 53*(2), 173–191.

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.

Cohen, R. (1987). International communication: An intercultural approach. *Cooperation and Conflict, 22,* 63–80.

Conklin, J. (1987). Hypertext: An introduction and survey. *IEEE Computer, 20,* 17– 41.

Fischer, G. H., & Formann, A. K. (1982). Some applications of logistic latent trait models with linear constraints on the parameters. *Applied Psychological Measurement, 6,* 397–416.

Formann, A. K., & Piswanger, K. (1979). *Wiener Matrizen-Test. Ein Rasch-skalierter sprachfreier Intelligenztest* [The Viennese Matrices Test. A Rasch-calibrated nonverbal intelligence test]. Weinheim, Germany: Beltz Test.

Gao, G. (1998). "Don't take my word for it."—Understanding Chinese speaking practices. *International Journal of Intercultural Relations, 22,* 163–186.

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

Gittler, G. (1990). *3DW. Dreidimensionaler Würfeltest. Ein Rasch-skalierter Test zur Messung des räumlichen Vorstellungsvermögens. Theoretische Grundlagen und Manual* [The Three-Dimensional Cube Test, 3DC. A Rasch-calibrated spatial ability test. Theoretical background and test manual]. Weinheim, Germany: Beltz Test.

Gittler, G., & Tanzer, N. K. (1990/1998). *The Progressive Test Series (PTS)*. Unpublished test, University of Vienna and University of Graz, Austria.

Gittler, G., & Tanzer, N. K. (1998, August). *Establishing cross-cultural equivalence of item complexity using the linear logistic test model (LLTM)*. Paper presented at the 24th International Congress of Applied Psychology, San Francisco.

Goodwin, R., & Lee, I. (1994). Taboo topics among Chinese and English friends. A cross-cultural comparison. *Journal of Cross-Cultural Psychology, 25*, 325–338.

Greenfield, P. M. (1997). You can't take it with you. Why ability tests 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 (Bulletin of the International Test Commission), 10*, 229–244.

Hambleton, R. K. (2002). Adapting achievement tests into multiple languages for international assessments. In A. C. Porter & A. Gamoran (Eds.), *Methodological advances in cross-national surveys of educational achievement* (pp. 58–79). Washington, DC: National Academy Press.

Hambleton, R. K., Li, S., & Sireci, S. (2003). *Pitfalls and obstacles in the test adaptation process: A meta-analysis* (Center for Educational Assessment Research Report No. 489). Amherst, MA: University of Massachusetts, School of Education.

Håseth, K. J. (1996). The Norwegian adaptation of the State–Trait Anger Expression Inventory. In C. D. Spielberger & I. Sarason (Eds.), *Stress and emotion* (Vol. 16, pp. 83–106). Washington: Taylor & Francis.

Hodapp, V., Tanzer, N. K., Maier, E. R., & Pestemer, I. A., & Korunka, C. (in press). The German adaptation of the Job Stress Survey: A multi-study validation in different occupational settings. In C. D. Spielberger & I. G. Sarason (Eds.), *Stress and emotion* (Vol. 17). Washington, DC: Taylor & Francis.

Hu, S., & Oakland, T. (1991). Global and regional perspectives on testing children and youth: An empirical study. *International Journal of Psychology, 26*, 329–244.

Landis, D., & Bhagat, R. S. (Eds.). (1996). *Handbook of intercultural training* (2nd ed.). London: Sage.

Marsh, H. W. (1988). *Self-Description-Questionnaire I. SDQ–I manual and research monograph*. San Antonio, TX: Psychological Corporation.

Marsh, H. W., & Craven, R. (1997). Academic self-concept: Beyond the dustbowl. In G. Phye (Ed.), *Handbook of classroom assessment: Learning, achievement, and adjustment* (pp. 131–198). Orlando, FL: Academic Press.

Marsh, H. W., Craven, R., & Debus, R. (1999). Separation of competency and affect components of multiple dimensions of academic self-concept: A developmental perspective. *Merril–Palmer Quarterly, 45*, 567–601.

Martini, D. R., Strayhorn, J. M., & Puig-Antich, J. (1990). A symptom self-report measure for preschool children. *Journal of the American Academy of Child and Adolescent Psychiatry, 29*, 594–600.

McFadden, J. (Eds.). (1993). *Transcultural counseling: Bilateral and international perspectives.* Alexandria, VA: American Counseling Association.

Mesquita, B., & Frijda, N. H. (1992). Cultural variations in emotions: A review. *Psychological Bulletin, 112,* 179–204.

Moosbrugger, H., & Oehlschlägel, J. (1996). *FAIR. Frankfurter Aufmerksamkeitstest* [Frankfurt attention test]. Bern, Switzerland: Huber.

Oakland, T. (1997). Test use among school psychologists: Past, current, and emerging practices. *European Journal of Psychological Assessment, 13, 2–9.*

Oakland, T. (2004). Use of educational and psychological tests internationally. *Applied Psychology: An International Review, 53*(2), 157–172.

Oakland, T., & Hu, S. (1992). The top 10 tests used with children and youth worldwide. *Bulletin of the International Test Commission, 19,* 99–120.

Paniagua, F. A. (1994). *Assessing and treating culturally diverse clients: A practical guide.* Thousand Oaks, CA: Sage.

Piswanger, K. (1975). *Interkulturelle Vergleiche mit dem Matrizentest von Formann* [Cross-cultural comparisons with Formann's Matrices Test]. Unpublished doctoral dissertation, University of Vienna, Vienna, Austria.

Ponterotto, J. G., Casas, J. M., Suzuki, L. A., & Alexander, C. M. (Eds.). (1995). *Handbook of multicultural counseling.* Thousand Oaks, CA: Sage.

Rosenzweig, S. (1977). *Manual for the children's form of the Rosenzweig Picture-Frustration (P-F) Study.* St. Louis, MO: Rana House.

Schneller, R. (1989). Intercultural and intrapersonal processes and factors of misunderstanding: Implications for multicultural training. *International Journal of Intercultural Relations, 13,* 465–484.

Schwenkmezger, P., Hodapp, V., & Spielberger, C. D. (1992). *Das State-Trait-Ärgerausdrucks-Inventar STAXI* [The German adaptation of the State-Trait Anger Expressions Inventory]. Bern, Switzerland: Huber.

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

Spielberger, C. D. (1988). *State–Trait Anger Expression Inventory research edition. Professional manual.* Odessa, FL: Psychological Assessment Resources.

Spielberger, C. D., & Comunian, A. L. (1992). *STAXI. State–Trait Anger Expression Inventory. Versione e adattamento italiano a curi di Anna Laura Comunian. Manuale* [Test manual of the Italian version of the State–Trait Anger Expression Inventory]: Firenze, Italy: Organizzazioni Speziali.

Spielberger, C. D., & Reheiser, E. C. (1994). Job stress in university, corporate, and military personnel. *International Journal of Stress Management, 1,* 19–31.

Tanzer, N. K. (1995). Cross-cultural validity of Likert-type scales: Perfect matching factor structures and still biased? *European Journal of Psychological Assessment, 11,* 194–201.

Tanzer, N. K. (1998). *Assessment of domain specificity in school-related Likert-type inventories: Conceptual issues, psychometric approaches, and cross-cultural evidence.* Unpublished habilitation monograph. Graz, Austria: University of Graz.

Tanzer, N. K., Ellis, B. B., Gittler, G., & Broer, T. (1996, August). *The use of collateral information in establishing the cross-cultural validity of tests.* Paper presented at the 26th International Congress of Psychology, Montreal, Canada.

Tanzer, N. K., Ellis, B. B., Zhang, H.-C., Sim, C. Q. E., Broer, T., & Gittler, G. (1997, July). *Cross-cultural decentering of test instructions in a letter-cancellation test: A field test of the ITC Guidelines for Test Adapations.* Paper presented at the 5th European Congress of Psychology, Dublin, Ireland.

Tanzer, N. K., Gittler, G. & Ellis, B. B. (1995). Cross-cultural validation of item complexity in a LLTM-calibrated spatial ability test. *European Journal of Psychological Assessment, 11,* 170–183.

Tanzer, N. K., Gittler, G., & Sim, C. Q. E. (1994). A cross-cultural comparison of a Rasch calibrated spatial ability test between Austrian and Singaporean adolescents. In A. Bouvy, F. J. R. van de Vijver, P. Boski, & P. Schmitz (Eds.), *Journeys into cross-cultural psychology* (pp. 96–110). Lisse, Netherlands: Swets.

Tanzer, N. K., & Sim, C. Q. E. (1991). *Self-concept and achievement attributions. A study of Singaporean primary school students* (Research Rep. No. 1991/5). Graz, Austria: University of Graz, Department of Psychology.

Tanzer, N. K., Sim, C. Q. E., & Marsh, H. W. (1992). Using personality and attitude inventories over cultures: Theoretical considerations and empirical findings. *Bulletin of the International Test Commission, 19,* 151–171.

Tanzer, N. K., Sim, C. Q. E.,& Marsh, H. W. (1997). *Where cross-cultural and differential psychology meet: Competence/task-easiness and interest/eagerness as subcomponents of academic self-concept* (Research Rep. No.1997/1). Graz, Austria: University of Graz, Department of Psychology.

Tanzer, N. K., Sim, C. Q. E., & Spielberger, C. D. (1996). Experience and expression of anger in a Chinese society. The case of Singapore. In C. D. Spielberger & I. Sarason (Eds.), *Stress and emotion* (Vol. 16, pp. 51–65). Washington, DC: Taylor & Francis.

Toubiana, Y. (1994). *Pictorial evaluation of test reactions (PETER).* Petach-Tikva, Israel: PETER.

Unterholzner, B. (1997). *Validierung einer italienischen Form des Prüfungsstressinventars (PSI) von Tanzer* [Validation of an Italian adaptation of Tanzer's Examination Stress Inventory]. Unpublished master's thesis, University of Graz, Austria.

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). Chicago: 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., & Tanzer, N. K. (1997). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology, 4,* 263–279.

Van Haaften, E. H., & van de Vijver, F. J. R. (1996). Psychological consequences of environmental degradation. *Journal of Health Psychology, 1,* 411–429.

Wehrly, B. (1995). *Pathways to multicultural counseling competence: A developmental journey.* Pacific Grove, CA: Brooks/Cole.

Yang, K.-S., & Bond, M. H. (1990). Exploring implicit personality theories with indigenous or imported constructs: The Chinese case. *Journal of Personality and Social Psychology, 58,* 1087–1095.

Zhang, J. X., & Bond, M. H. (1998). Personality and filial piety among college students in two Chinese societies. The added value of indigenous constructs. *Journal of Cross-Cultural Psychology, 29,* 402–417.