12

إنشاء وتكييف وتقييم صدق اختبارات القبول بلغات متعددة: الحالة الإسرائيلية

مايكل بيلر
هيئة الاختبار التعليمية
ناعومي غافني وبنيناهاناني
المعهد الوطني للاختبار والتقييم

يؤثر اختلاف الأهداف التي تترجم الاختبارات من أجلها على عملية الترجمة وعلى دور كل من اللغات المستخدمة. يلزم نقاش مستقل حول الاعتبارات المتعلقة بملاءمة النسخة المترجمة لكل استخدام على حدة. إن أحد الاستخدامات الشائعة للاختبارات المترجمة هو تطبيق مقياس معياري موثوق مثل اختبارات معامل الذكاء IQ أو استبيانات الشخصية. وتستخدم هذه الاختبارات (مثل مجموعة تقويم كاوفمان للأطفال) [K-A-B-C] بشكل رئيس لتطوير المعدلات الإحصائية المحلية اللازمة لأغراض البحث العلمي. تتضمن عملية الترجمة في هذه الحالة الحد الأدنى اللازم من تكييف الاختبار بورتينغا، (1995)

إن التقويمات الدولية للتعليم (مثل دراسة التوجهات الدولية في الرياضيات والعلوم [TIMMs] هي أمثلة على الأبحاث عبر الدول التي لا تعتبرها أية لغة أو محتوى مسيطراً بل تحدد كل الدول المشاركة محتويات التقويم معاً مؤمنة بذلك قواسم مشتركة عظمى.

obeikandl.com

تترجم النسخة المتفق عليها إلى كل اللغات المشاركة (انظر غريسي، 2003 وهو مثال ممتاز) تكون المقارنة الرئيسة من هذه الحالة المقارنة بين الدول مع أن بعض الحالات (كما في كندا) المقارنة في بلد واحد بين مجموعات مختلفة.

هدف آخر لترجمة الاختبارات هو وضع سياسة انتقاء عادلة وصحيحة للمرشحين في مختلف المجموعات اللغوية للمتقدمين للقبول في معاهد التعليم العالي في بلد ولغة تعليم معينين. هذا الأمر شائع في البلدان التي تستقبل أعداداً كبيرة من المهاجرين (مثل: الولايات المتحدة وكندا وإسرائيل) فقد تؤثر عملية الترجمة في هذه الحالة وبشكل كبير على صدق قرارات فردية ذات نتائج خطيرة. إن استخدام درجات اختبارات القبول المجراة لكل المجموعات بلغة المصدر تؤدي إلى الخلط بين المنشأ المقاس ومستوى إتقان لغة المصدر؛ لذلك هناك حاجة لإيجاد الوسائل للتقليل من الخلط بين هذين المتغيرين وذلك على سبيل المثال بترجمة اختبارات القبول إلى لغات المتقدمين المختلفة وقياس إتقانهم للغة المحلية بشكل منفصل.

يجب معالجة أهداف الترجمة مع أخذ المجموعة السكانية المستهدفة بعين الاعتبار، حيث يختلف وضع هذه المجموعات من حالة لأخرى: في بعض الحالات يملك البلد الواحد أكثر من لغة رسمية واحدة (مثال: سويسرا، كندا، إسرائيل) وبالتالي فإن الاختبارات تترجم بشكل روتيني. وحتى في مثل هذه الحالات هناك اختلافات بين البلدان، فبعضها يؤمن للمجموعات اللغوية المختلفة نظاماً تعليمياً كاملاً بلغتها الخاصة (مثل سويسرا) فيما تؤمن بلدان أخرى نظاماً تعليمياً نصف منفصل (كندا، إسرائيل). ففي إسرائيل، على سبيل المثال، تشكل اللغة العربية لغة رسمية ثانية ويوجد نظام تعليمي منفصل للسكان الناطقين يمتد من الحضانة حتى نهاية التعليم الثانوي (باستثناء بعض الموضوعات التي تدرس باللغة العبرية). أما اللغة المعتمدة في التعليم العالي، فهي العبرية والعبرية فقط. هناك نظام تعليمي واحد لكلتا المجموعتين السكانيتين.

يشكل المهاجرون في البلدان المختلفة مجموعات سكانية إضافية مستهدفة في ترجمة الاختبارات. ويجب أن لا يعامل المهاجرون من بلد معين وكأنهم مجموعة متجانسة. فهم يتفاوتون في معرفتهم بلغتهم الأم وباللغة المحلية الجديدة وفي معرفتهم بالثقافة المحلية والنظام التعليمي (يتعلق ذلك بعمرهم عند الهجرة وطول المدة التي أمضوها في البلد الجديد، ومستوى انغماسهم في ثقافة البلد الجديد). لا تضمن ترجمة اختبار إلى لغة معينة (للأسباب المبنية أعلاه) مقارنة صحيحة حتى بين أفراد مجموعة لغوية واحدة تؤثر في الدرجات المختلفة في إتقان كل مجموعة سكانية للغة المصدر واللغة الهدف في طريقة ترجمة الاختبارات. فقد يكون من المفضل في بعض الحالات ترجمة تعبيرات معينة من الاختبار عوضاً عن ترجمته كاملاً. فعلى سبيل المثال، قد يفضل المهاجرون القدامى إلى إسرائيل (الذين درسوا في إسرائيل لعدة سنوات) أن يقدموا اختبارات القبول الجامعية باللغة العبرية مع ترجمة تعبيرات معينة إلى الروسية عوضاً عن أخذ الاختبار كاملاً باللغة الروسية.

يركز هذا الفصل على بعض المشكلات الكبرى المتعلقة بترجمة الاختبارات من جهة استخدام الاختبار وتحليل الدرجات. ويعالج هذا الفصل بشكل خاص المدى الذي يجب أن يصل إليه ترجمة أو تكييف أو تغيير النسخة الأصلية من الاختبار. كما يعالج تحديد معايير تقديم نوعية الترجمة ومقاربة معايرة الدرجات في النسخ اللغوية المختلفة.

تناقش أساليب معالجة هذه الموضوعات ويتم شرحها باستخدام النسخ اللغوية المختلفة من اختبار الدخول القياسي النفسي PET اختبار الدخول القياسي النفسي PET، هو اختبار أهلية مدرسية ويوضع ويجرى من قبل المعهد الوطني للاختبار والتقويم NITE ويستخدم بالإضافة إلى شهادة قبول لاتخاذ قرارات القبول في كل الجامعات الإسرائيلية ومعاهد التعليم العالي الأخرى. أما شهادة القبول فهي توضع بناء على تقييم المدرسة واختبارات أداء خارجية تجرى على نطاق وطني. يقيس

PET قدرات معرفية ومدرسية مختلفة في محاولة لتقدير النجاح المستقبلي في الدراسات الأكاديمية. ويتألف PET من ثلاثة اختبارات فرعية مختلفة متعددة الخيارات، المحاكمة اللغوية (V)، المحاكمة الكمية (Q)، والإنجليزية كلغة أجنبية (E) لا يوجد تصحيح للتخمين في وضع درجات الاختبار، بل يشجع الممتحنون على التخمين عندما لا يعرفون الجواب الصحيح. (انظر بيلر، 1994 للحصول على وصف تفصيلي لـ (PET لقد واجه الذين يقومون بالقياس النفسي والمخططون عند وضع سياسة القبول في جامعات إسرائيل مشكلة إيجاد الأساليب المثلى للتنبؤ بالنجاح الأكاديمي للمتقدمين غير الناطقين بالعبرية (بالإضافة إلى الناطقين بالعبرية) في معاهد التعليم العالي (حيث لغة التعليم هي العبرية). كان الهدف بعبارة أخرى تصنيف الممتحنين في المجموعات اللغوية المختلفة على مقياس واحد يقيس معياراً مشتركاً في السياق الثقافي نفسه. تقرر بناء على ذلك إجراء PET باللغة التي يتقنها المتقدم بشكل أفضل لأنه كان يُعتقد أن هذا الأمر يؤمن للممتحن فرصة الأداء الأمثل. ويجري ترجمة PET حالياً إلى اللغات التي يتكلمها أغلب الممتحنين: العربية، الروسية، الفرنسية، الإسبانية والإنكليزية. إن عملية الترجمة جهد مستمر، حيث تترجم سنوياً أربعة نماذج إلى العربية. نموذجان إلى كل من الروسية والإنكليزية ونموذج واحد إلى كل من الفرنسية والإسبانية (من أصل عشر نماذج إلى ثمانية عشر نموذجاً بالعبرية).

ينشر المعهد الوطني للاختبار والتقويم كتيّب معلومات يحتوي اختبارات مجراة سابقاً مع شرحها، وذلك لتعريف الممتحنين بالـ PET والتأكد من فهم كل الأشخاص لمتطلبات كل مهمة في الاختبار بشكل فاعل. يترجم هذا الكتيّب إلى اللغات الخمسة المذكورة سابقاً. وهذا الإجراء مهم لأن المجموعات اللغوية المختلفة تتفاوت في مدى خبرتها السابقة مع الاختبارات متعددة الخيارات. من بين الـ 66731 شخصاً الذين تقدموا للاختبار (PET) عام 1998 على سبيل المثال، اختار 27 ٪ آخذ الاختبار من واحد من هذه اللغات (العربية15 ٪، الروسية 10 ٪، 2٪ من لغات أجنبية أخرى)

تطلب بعض المعاهد من المتحنين الذين أخذوا PET بلغة أجنبية أن يأخذوا اختباراً إضافياً لإتقان اللغة العبرية(HP) وتوضع درجاته بشكل منفصل. إن النسخ غير العبرية من PET هي أساساً ترجمات للنسخ العبرية المجراة للناطقين بالعبرية، ولذلك فهي ذات بناء متماثل. يكون اختبار اللغة الإنكليزية الفرعي واحداً لكل النسخ اللغوية. ويترجم اختبار المحاكمة الكمية في العبرية، وأساس ذلك أنه يمكن مقارنة مواد الرياضة المترجمة بشكل مباشر بالمصدر. أما اختبار المحاكمة اللغوية فهو مترجم جزئياً فقط حيث تنتقى معظم البنود من مجموع البنود العبرية ويتم إنشاء بنود أخرى بشكل خاص للنسخ اللغوية المختلفة (مثل بنود المفردات والتعبير). نناقش في الفقرات التالية مسائل إجرائية وجوهرية في ترجمة PET، ويعطى اهتمام خاص لترجمة اختبار المحاكمة اللغوية.

عملية الترجمة

انتقاء نموذج اختبار للترجمة: بعض الاعتبارات

تترجم الاختبارات في اللغات الخمس (العربية، الروسية، الفرنسية، الإسبانية، الإنجليزية) من نماذج اختبار عبرية مجراة سابقاً. وهذا يعني أن المواد المنتقاة للترجمة ذات نوعية عالية بمعايير القياس النفسي. يجب مراعاة الاعتبارات التالية عند انتقاء النسخ العبرية للترجمة:

1- نوعية المعايرة: لتلافي حدوث مشكلات معايرة، نحاول تحديد نوعية المعايرة لنموذج مجرى سابقاً لمتحنين ناطقين باللغة العبرية مشابهين نسبياً من حيث توزيع القدرات للممتحنين باللغات الأخرى (المجموعة الهدف).

2- الجدارة: لقد وُجد جدارة الدرجة الكاملة في النسخ المترجمة كان قريباً جداً من جدارة النسخة العبرية الأصلية (انظر الجدول 12-1 أما في الماضي، فكانت جدارة اختبار المحاكمة اللغوية للمجموعة الناطقة بالعربية أقل من جدارة الاختبارات الفرعية المترجمة الأخرى (انظر الجدول 12-1 للقيم الحالية). كان السبب الرئيس هو أن اختبار المحاكمة اللغوية كان صعباً جداً للناطقين

بالعربية. ولزيادة جدارة اختبار المحاكمة اللغوية في هذه المجموعة تم إنشاء اختبار أسهل بانتقاء النصف الأسهل من المواد من نموذج عبري (ليستعمل لاحقاً في المعايرة) وإكمالها ببنود سهلة من مصرف الأسئلة.

3- الحفاظ على تواتر التعبيرات التقنية في نصوص القراءة والفهم: تم تحاشي نصوص القراءة والفهم التي تزخر بالتعبيرات التقنية (تعبيرات علمية، لغة قانونية، مصطلحات نفسية) عند انتقاء نصوص للترجمة. وذلك لأن هذه التعبيرات قد تكون مفهومة في لغة معينة دون الأخرى. بالإضافة إلى ذلك، فإن تواتر استخدام هذه التعبيرات يختلف باختلاف السياق الثقافي واللغوي، فلا يترجم نص ما يزخر بالكلمات الأجنبية إلى العربية لأن الممتحنين الناطقين بالعربية لا يصادفون مثل هذه الكلمات في المدرسة الابتدائية أو الثانوية.

4- السياق الثقافي: يجب أن يكون السياق الثقافي مألوفاً لكل الممتحنين. يجب عدم ترجمة نص قراءة فهم يحوي دلالات ثقافية محلية.

5- مراجعات الحساسية: تخضع مواد الاختبارات لمراجعات حساسية لتفادي اختبار مواد قد تكون مثيرة أو مؤذية في النسخ المترجمة. فعلى سبيل المثال لا نستعمل أي مادة تحتوي كلمة انتفاضة بالعربية بسبب الحساسية السياسية للكلمة. ويتم تفادي النصوص التي تناقش السياسة أو الدين أو الجنس، إلى ما هنالك.

جدول رقم 12-1

جدارة الدرجة الكاملة في النسخ المترجمة لاختبارات PET الثانوية

Language	*V*	*Q*	*E*	*PET*
Eebrew (65)	0.90	0.90	0.94	0.96
Arabic (23)	0.84	0.86	0.82	0.93
Russian (18)	0.86	0.88	0.92	0.94
French (9)	0.82	0.87	0.90	0.93
Spanish (9)	0.82	0.87	0.93	0.94
English (15)	0.90	0.90	0.96	0.96

ملاحظة: تظهر أعداد نسخ الاختبار بين أقواس.

مراحل عملية الترجمة:

تحقق عملية الترجمة شروط تكييف الاختبارات التي وضعتها الهيئة الدولية للاختبارات (هامبلتون، 1994، انظر أيضاً الفصل الأول من هذا الكتاب)، هناك أربع مراحل في عملية الترجمة:

1- الترجمة الأولية: يقوم مترجم مؤهل وخبير يتقن اللغتين ويعرف الثقافتين، خاصة الهدف، بترجمة النسخة العبرية الأصلية إلى اللغة الهدف. يناقش المترجم المشكلات التي تظهر في أثناء عملية الترجمة مع عالم القياس النفسي الذي يقود عملية الترجمة. تجرى حالياً ترجمتان مستقلتان إلى اللغة الروسية، وذلك بعد توصية من هامبلتون (في اتصال شخصي معه عام 1997). وتشير التجربة حتى الآن مع هذا الإجراء المكلف أنه يحسن نوعية المراجعة التي تجري في المرحلة التالية. ويجري الآن تقويم الجدوى الاقتصادية لهذا الإجراء الإضافي.

2- المراجعات المستقلة: توضع النسخ المترجمة لمراجعة نقدية من قبل عدة مراجعين ثنائي اللغة، يملك بعضهم خلفية قوية في الرياضيات والمنطق، فيما يبرع الآخرون في المحاكمة اللغوية. يقرأ مراجعون أمريكيون وبريطانيون النسخة الإنكليزية، فيما يقرأ مراجعون من عدة بلدان ناطقة بالإسبانية النسخة الإسبانية. يُطلب من المراجعين نقد النسخة المترجمة أولاً دون النظر إلى النسخة العبرية. وبعدها فقط يقارنون النسخة المترجمة بالنسخة العبرية الأصلية. ثم يُطلب منهم الانتباه بشكل خاص إلى دقة الترجمة، ووضوح الجمل، ومستوى صعوبة الكلمات، وسلاسة النص. يُقوّم كل مراجع كل بنود الاختبار ويتحرى أي تغيرات حدثت في المنطق الداخلي للبند، ويتأكد من أن كل بند ما زال يملك حلاً واحداً فقط لا غير وأن الإجابات المضللة مناسبة من حيث جاذبيتها. يناقش المترجم وعالم القياس النفسي تعليقات واقتراحات ويتخذان التعديلات المناسبة.

3- الترجمة الراجعة: يقوم خبير ثنائي اللغة، لم يسبق له رؤية النسخة العبرية، بترجمة النسخة المترجمة وبشكل شفهي إلى اللغة العبرية. وتُجرى هذه المرحلة شفهياً، وذلك يسمح بالنقاش بين عالم القياس النفسي والمترجم الراجع. تقارن الترجمة الراجعة مباشرة مع النسخة العبرية وتعدل المواد المترجمة عند الضرورة.

4- المراجعة النهائية قبل الإجراء الأولي للاختبار: تعطى النسخة المعدلة من الترجمة إلى ناطق أصلي باللغة الهدف. ويطلب منه حل الأسئلة دون النظر إلى الأصل العبري والتأكد من وجود حل واحد صحيح فقط لكل سؤال. يُقوّم عالم القياس النفسي الإجابات ويبحث عن الإجابات الخاطئة التي قد تنتج عن أخطاء في الترجمة.

المشكلات الخاصة بترجمة اختبار المحاكمة اللغوية:

إن فقرات المحاكمة اللغوية هي الأكثر إشكالية في الترجمة؛ لأن الكلمات والمفاهيم في إحدى اللغات قد لا تحافظ على المعاني والدلالات نفسها والاعتياد أو مستوى الصعوبة عند ترجمتها إلى لغة أخرى. تشكل العبارات والتعابير مشكلة شائعة حيث لا يمكن غالباً ترجمتها على الإطلاق. كما تختلف الكلمات في غناها اللفظي في مفرداتها. فعلى سبيل المثل، تزخر العبرية بالكلمات المتعلقة بالزراعة. ففي الإنكليزية يقال يقطف العنب أو يقطف الزيتون، وهكذا. أما في العبرية فهناك فعل مختلف لقطف العنب وقطف الزيتون وما شابه ذلك. كذلك يوجد في العبرية كلمات مختلفة لغسل الأرض أو غسل الصحون أو غسل الثياب. في حين يستعمل الناطقون بالإنكليزية فعل "غسل" لكل هذه الأعمال. تملك كل من العبرية والإنكليزية كلمة واحدة لـ "جمل"، في حين يوجد في العربية عدد كبير للتعبير عن الأنواع المختلفة من الجمال حسب مواصفاتها.

تشكل الترجمة إلى العربية مشكلة كبيرة. ففي حين أن العربية الفصحى هي نفسها لكل العرب في كل البلاد العربية، تختلف العربية المحكية بشكل كبير عن

العربية المكتوبة. كما تختلف من بلد إلى آخر وحتى من منطقة إلى أخرى من البلد نفسه. فكلمة "coat" تعني "معطف" بالعربية الفصحى، و"كبوت" باللغة العامية. أما "hat" فتعني "قبعة" في العربية الفصحى و"طاقية" بالعربية العامية. تتوخى الترجمة العربية للاختبار تجنب استخدام الكلمات العامية. وتستخدم عوضاً عن ذلك الكلمات الفصحى بالرغم من أنها أكثر صعوبة. فقد يواجه الممتحنون الذين يقرؤون الأدب هذه الكلمات، في حين قد لا يعرفها آخرون.

تناقش الفقرات التالية بعض مشكلات الترجمة التي تتراوح بين بعض أنواع البنود التي لا يمكن ترجمتها على الإطلاق، إلى تلك التي يمكن ترجمتها مباشرة أو بعد تعديل طفيف.

بنود استبدال الحروف. تستند هذه البنود إلى خاصية الصرف من اللغات السامية والتي لا تشاركها فيها اللغات الهندو -أوربية. وهي أن معظم المفردات في اللغة العبرية، أي أن كل الأفعال ومعظم الأسماء والصفات، تتألف من جذر + نهاية صرفية. تتألف بنود استبدال الحروف من أربع جمل تعدّل في كل منها كلمة واحدة بتغيير حروف جذرها إلى قالب ثابت هو عادة الحروف (PTI) يحل هذا القالب الثابت، في ثلاث من الجمل، محل الجذر نفسه، في حين يحل في الجملة الرابعة محل جذر مختلف وعلى الممتحنين تعيين هذه الجملة. يمكن استخدام هذا النوع من البنود في اللغة العربية لأنها لغة سامية. لكن البنود لا يمكن ترجمتها من العبرية، بل تكتب في العربية ويتم اختبار البنود الجديدة من هذا النوع قبل استخدامها في وضع الدرجات. لا يمكن استخدام هذا النوع من البنود في النسخ اللغوية الأخرى.

الكلمات والتعابير. لا يمكن ترجمة الكلمات والتعبيرات من العبرية. بل تكتب مباشرة باللغة الهدف. وقد وُجد أنه من الضروري والمجدي اقتصادياً في نسخة الاختبار العربية اختبار هذه البنود قبل استخدامها لأغراض التصحيح.

التشبيهات. هذا النوع من البنود هو الأصعب في الترجمة. حيث يتعلق بمعاني ودلالات كلمات إفرادية والعلاقة بين أزواج من الكلمات. يندر إيجاد كلمات تملك المعنى الدقيق نفسه والدلالات نفسها مستوى الصعوبة نفسه في لغة أخرى. يجب عند ترجمة التشبيهات الحفاظ بما يمكن من الدقة على العلاقة بين كلمتي كل زوج من الكلمات مع الإبقاء على مستوى صعوبة المفردات. يصمم التشبيه الأصلي عادة لاختبار إتقان مفردات اللغة العبرية، بالإضافة إلى القدرة التحليلية. وغالباً ما تكون البنية المترجمة أكثر سهولة.

إتمام الجمل. تستلزم هذه البنود هذه العلاقات المنطقية واللفظية في جملة مركبة. وتصعب ترجمة بنود إتمام الجمل. فمن أجل تأليف جملة طبيعية وانسيابية في اللغة الهدف، يلزم غالباً تغيير بنية الجملة مما يؤثر على موقع الكلمات الناقصة في الجملة المترجمة. ويجب على المترجم أن يتأكد من أن الإجابات المضللة تنتج جملاً صحيحة نحوياً وإعرابياً بحيث يعتمد اختيار الجواب الصحيح على المنطق الداخلي وليس على تلميحات بنيوية ونحوية. كما يجب الحفاظ على مستوى اللغة (عادية، رسمية، أدبية) وتعقيدات الكلمات المفقودة وعدد الفراغات.

تظهر المشكلات مثلاً في العربية حيث لكل اسم جنس نحوي قد لا يطابق جنس مقابله العبري. كما أنه يوجد في العربية شكلان من الجمع، الجمع الأكثر من بندين (الجمع) والجمع لبندين (المثنى) حيث يصرف الفعل بناءً على ذلك. كما تبدأ الجمل في العربية بفعل حين تبدأ في العبرية باسم. تستدعي كل هذه المشكلات عدة تعديلات في بنية البند وتعقد مهمة ترجمة بنود إتمام الجمل.

عندما تتغير بنية الجملة في اللغة الهدف، قد تحتوي الجملة ثلاثة فراغات عوضاً عن أربعة في النسخة العبرية الأصلية. ولا يوجد سبب مسبق لعدم استخدام مثل هذه البنود، أو إذا وُجد أن هذا البند في تحليل المعايرة أسهل بكثير من مقابله في العبرية، فسوف يلغى من المعايرات اللاحقة.

المنطق. يجب ترجمة بنود المنطق بعناية ودقة. يجب على المترجم أن يحاول الحفاظ على كل العناصر المنطقية في البند العبري مع الحفاظ على البنية الموجودة نفسها في البند الأصلي. يجب الانتباه إلى السياق فيما إذا كان حقيقياً أو خيالياً وتعديل الأسماء ووحدات القياس (كيلومترات أو أميال) بحيث تكون المصطلحات المستخدمة مألوفة لدى المتحنين.

في محاولة للحفاظ على البنية الأصلية لفقرات المنطق الأصلية والنسخة المترجمة (مثل الحفاظ على النفي أو النفي المضاعف أو أحرف الربط مثل فقط.... إلخ (يلزم أحياناً تغيير البنية بحيث يكون التركيب الإعرابي في اللغة الهدف صحيحاً. أحد الأمثلة على ذلك أن التركيب الإعرابي للجملة العبرية "all p`s are not q" غامض في الإنكليزية. وهكذا فإن مقولة بالعبرية مثل "All birds of prey are not green" والتي تعني "أنه لا يوجد طير جارح واحد أخضر اللون"، لا يمكن ترجمتها مباشرة إلى الإنكليزية بل يجب تغييرها كالتالي "No birds of Prey are green"وتظهر صعوبة مماثلة عند ترجمة الجملة نفسها إلى الفرنسية. ففي الفرنسية، لا يمكن أن يأتي نفي بعد "كل" وبالتالي تصبح الترجمة "Aucun oiseau predate n`est vert"

القراءة والفهم. يتم التأكيد في ترجمة النص على النقاط التالية: دقة الترجمة، الحفاظ على سلاسة وغنى وأسلوب اللغة باستعمال مفاهيم مألوفة في اللغة الهدف واستعمال المصطلحات التي تظهر في النص بشكل متناسق.

إحدى الانتقادات التي توجه إلى الاختبار العابر للثقافات، هو أن النص المترجم لا يوصل المعنى نفسه ولا يحافظ على مستوى الصعوبة نفسه في النص الأصلي. لهذا، أنشئ في المعهد الوطني للاختبار والتقويم فريق خاص بهدف إيجاد نصوص خاصة للناطقين بالعربية، والذين يشكلون أكبر مجموعة من المتقدمين للاختبار غير الناطقين بالعبرية. تُكتب النصوص بالعربية ويتم تكييفها لمتطلبات اختبارات المعهد

الوطني للاختبار والتقويم. أحد الأسئلة التي يجب الإجابة عنها في المستقبل، -هو فيما إذا كانت الدرجات الناتجة عن اختبار يستخدم نصوص فهم مكتوبة أصلاً بالعربية يمكن مقارنتها مع الدرجات الناتجة عن نسخ الاختبار التي تستخدم نصوصاً" مكتوبة أصلاً باللغة العبرية.

تصحيح النسخ اللغوية

يتم تصحيح كل اختبار فرعي بشكل منفصل باستعمال صيغة قاعدة درجات. ويتم معايرتها على مقياس كان له في المجموعة المعيارية الأصلية (المتقدمين للاختبار الناطقين بالعبرية) 1984 متوسط قيمته 100 وانحراف معياري قيمته 20 أما الدرجة الكلية PET فهي المجموع المثقّل لدرجات الاختبار الفرعية ، (2V, 2Q + E) وله متوسط قيمته 500 وانحراف معيار قيمته 100 (لوصف أكثر تفصيلاً انظر بيلر، 1994).

تطبق الضوابط نفسها في تصحيح الاختبارات الفرعية للمحاكمة الكمية والإنكليزية في كل النسخ اللغوية (وذلك بافتراض أن الترجمة لا تغير معنى الأسئلة الكمية). يتم القيام بعملية معايرة شبيهة بالتي وصفها آنغوف ومودو (1973)، وتُستخدم في تصحيح اختبار المحاكمة اللغوية. تم تأسيس قاعدة مشتركة بين النسخة العبرية وكل من النسخ اللغوية الأخرى وذلك بانتقاء بنود ذات مؤشرات قياسية نفسية متشابهة وترتيب صعوبة متشابهة (باستخدام التقنيات البيانية دلتا) لمجموعتين من الممتحنين. عند وضع القاعدة المشتركة تطبق أساليب التعديل الخطي.

يظهر الجدول 12-2 المتوسطات والانحرافات المعيارية للنسخ اللغوية المختلفة لـ PET واختباراته الفرعية للعام الدراسي 1997/ 1998 المجموعات الناطقة بالفرنسية والإسبانية والإنكليزية صغيرة جداً، بحيث لا يمكن تعميمها. أما المجموعات الناطقة بالعربية والروسية فهي كبيرة ومستقرة عبر السنين بشكل كاف لتكون ممثلة لأكبر مجموعتي أقليات تتقدم للدخول إلى معاهد التعليم العالي في إسرائيل.

وهكذا، فإن أغلب التحليل والنقاش التاليين مبني على النسخ العبرية والعربية والروسية. كما ذكر سابقاً، يملك السكان الناطقون بالعربية في إسرائيل نظاماً تعليمياً منفصلاً لغة التعليم فيه هي العربية. يظهر التفاوت بين المجموعات الناطقة بالعبرية وتلك الناطقة بالعربية في الصفوف الأولى، كما أظهرت عدة تقويمات وطنية للنظام التعليمي. كما يظهر مستوى الأداء الأعلى بعض الشيء في الرياضيات لدى الناطقين بالعربية نسبة إلى التحصيل اللغوي في سن مبكرة.

من اللافت للنظر أن الفرق الأكبر في الأداء بين المجموعتين الناطقتين بالعربية والروسية والمجموعة العبرية، هو في اختبار اللغة الإنكليزي الذي لا يخضع للترجمة.

الجدول رقم 12-2

المتوسطات والانحرافات المعيارية لصور لغوية متعددة لاختيار PET والاختبارات الفرعية المصاحبة له في العام 1997/1998.

		Total Score		*Verbal Reasoning*		*Quantitative Reasoning*		*English*	
Language	*N*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Hebrew	48,897	554	101	108	20	111	19	111	23
Arabic	9,949	431	85	86	16	92	19	82	16
Russian	6,366	512	101	92	18	106	18	95	22
French	511	521	84	99	16	104	19	112	17
Spanish	363	480	82	90	14	96	17	108	22
English	645	552	106	100	21	107	21	131	23

جودة النسخ المترجمة

بالإضافة إلى عملية الترجمة الدقيقة التي وُصفت سابقاً، تستعمل المعايير الكمية التالية لتقويم جودة النسخ المترجمة: الأثر التفاوتي للتخمين، تحليل البنود، الأداء التفاوتي للبنود ذات الجدارة العالية، تكافؤ البنية، المصداقية والانحياز في

التنبؤ بالدرجات. تتأثر هذه المعايير بالترجمة وبعوامل أخرى متنوعة متعلقة بثقافة المجموعة وتوزيعها الثقافي. تُعطى اللغة الروسية اهتماماً خاصاً لأنها لغة أكبر مجموعة مهاجرين في إسرائيل (15٪ من سكان إسرائيل).

الأثر التفاوتي للتخمين: بحثت دراسات أجراها غافني وميلاميد (1994) الظاهرة التالية: بالرغم من توجيه الممتحنين إلى التخمين إذا لم يعرفوا الجواب الصحيح، فإن 75 ٪ - 93٪ منهم فقط (حسب الاختبار الفرعي) أجابوا على كل بند في PET لقد افتُرض أن المجموعات اللغوية المختلفة قد تظهر أنماطاً مختلفة من السلوك من ناحية التخمين. كان من المتوقع، على سبيل المثال، أن يكون الناطقون بالإنكليزية أكثر اعتياداً على اختبارات الخيارات المتعددة، أن يتبعوا تعليمات الاختبار بدقة. أما الناطقون باللغة الروسية من جهة أخرى، فقد يكونون أقل ميلاً للتخمين. وذلك لقلة اعتيادهم على هذا النوع من الاختبارات. كما افتُرض وجود تأثير لمعرفة الجمهور العام للاختبارات ذات الخيارات المتعددة.

أوحت النتائج أن الناس المختلفي الخلفيات الثقافية مختلفون في ميلهم إلى التخمين. وُجد تأثير للمجموعة وتأثير الاعتياد. في عام 1984 كان الممتحنون الناطقون بالروسية والعربية والفرنسية يميلون إلى إغفال بنود أكثر من الناطقين بالعبرية والإنكليزية والإسبانية. وفي عام 1987 (أي بعد أربع سنوات من إجراء PET كان الممتحنون بالروسية يميلون إلى إغفال بنود أكثر من كل المجموعات الأخرى. لكن نسبة البنود المغفلة انخفضت بشكل ملحوظ في كل المجموعات، حيث أصبح الاختبار أكثر اعتياداً، وتحضير الاختبار أكثر شيوعاً. أوصت دراسة في عام 1994 بالتركيز على تعليمات الاختبار، خاصة بين أفراد المجموعات الأكثر ميلاً لتجنب التخمين.

تحليل البنود والأداء التفاوتي للبنود:

تاختبار نوعية كل بند مترجم (من حيث مستوى الصعوبة ودرجة التحيز). كما ياختبار بالإضافة إلى ذلك، الأداء التفاوتي لكل بند مترجم وذلك بمقارنة الممتحنين

الناطقين بالعبرية مع غير الناطقين بها (يشير تعبير الأداء التفاوتي للبند إلى الملاحظات البسيطة في أن البند يُظهر خصائص إحصائية مختلفة للمجموعات المختلفة، وذلك بعد استبعاد فروق القدرات بين المجموعات). إذا كانت الخصائص الإحصائية لبعض البنود المترجمة سيئة، تتم مراجعة هذه البنود والبحث عن أسباب فشلها. بعد ذلك يتم اتخاذ قرار حول إدخاله في القاعدة المشتركة المستخدمة للمعايرة.

قام كل من غافني ويهو شافات (1993) باختبار الأداء التفاوتي للبنود في اختبار المحاكمة اللغوية لثلاثة نماذج روسية من PET باستخدام تقنية الخط البياني دلتا المقترحة من قبل آنغوف (1927). ولوحظ أن أكبر أداء تفاوتي للبنود كان في التشبيهات. وأصغر أداء تفاوتي للبنود كان في بنود المنطق وإتمام الجمل. هذه النتائج مشابهة لتلك التي وجدها آنغوف وكوك (1988) للممتحنين الناطقين بالإنكليزية والإسبانية في اختبار التقويم المدرسي (SAT) وذلك باستثناء بند المنطق غير الموجود في اختبار التقويم المدرسي اللغوي. وقد أظهرت بنود القراءة والفهم أداءً تفاوتياً للبنود أكبر نسبياً منه في دراسة آنغوف وكوك.

وقد فسر كل من آلالوف وهامبلتون وسيرسي (1999) وسيرسي وآلالوف (2003) نتائج بحثهم عن علاقة الأداء التفاوتي للبنود مع نوع البند (باستعمال طريقة مانتل – هانسل) ووضعوا نظريات حول أسباب ومصادر الأداء التفاوتي للبنود. قام هؤلاء بتحليل ثلاثة نماذج من النسخ العبرية والروسية لـPET تعكس النتائج مدى المشكلات المتعلقة بترجمة الأنواع المختلفة للبنود اللغوية كما وضحنا في الفقرة السابقة. فقد وجدوا أن 42 من أصل 125 بنداً أظهرت أداءً تفاوتياً عبر اللغات. كانت التشبيهات الأكثر إشكالية، حيث أظهرت 65 ٪ منها أداء تفاوتاً. وقد أظهر الممتحنون الناطقون بالروسية نتائج أفضل في معظم هذه البنود من الناطقين بالعبرية. كما أن نسبة كبيرة من بنود إكمال الجمل (45٪) أظهرت أداءً تفاوتياً. ولكن في هذه الحالة كان أداء المجموعتين اللغويتين متقارباً. وتم الطلب من

المترجمين أن يفكروا في أسباب الأداء التفاوتي لكل بند، فكانت الأسباب الرئيسة المقترحة هي تغييرات صعوبة الكلمات، تغيرات في الصيغة، تغيرات في العلاقة الثقافية وتغييرات في المحتوى.

الجدارة:

تقدر الجدارة الداخلية لكل اختبار فرعي وللعلامة الكاملة بشكل روتيني لكل نسخة لغوية. يظهر الجدول 12-2 متوسط معاملات الاتساق الداخلي (KR-20) للاختبارات الفرعية الثلاثة وللعلامة الكاملة في النسخ اللغوية المختلفة من PET قد تفسر جدارة النسخ الأجنبية الأولى نسبياً، بمشكلات متعلقة بالترجمة. ولكن الجدارة الداخلية لا تُحدد فقط بجودة بنود الاختبار وجودة الترجمة، بل أيضاً بالتباين الحقيقي في مجموعة الممتحنين. تظهر الخبرة المكتسبة في المعهد الوطني للاختبار والتقويم أنه في كثير من المهارات تم الخلط بين نوعية الترجمة وفروقات الأداء الحقيقية. فعندما تتفاوت مجموعتان في القدرة، فإن ذلك بحد ذاته قد يخلق فروقاً في الجدارة وقابلية المقارنة و الأداء التفاوتي للبنود. فعندما تكون البنود صعبة لمجموعة معينة، تضعف جدارة الاختبار في هذه المجموعة. فعلى سبيل المثال: كان متوسط الجدارة لاختبار المحاكمة اللغوية في النماذج العربية الخمسة الأولى التي وضعت بين عامي (1984 و1989) 0.68 (بيلر وغافني 1995). ولرفع هذه الجدارة، تم إنشاء اختبار محاكمة لغوية بشكل خاص للنسخة العربية حيث استعمل فيه بنوداً أكثر سهولة فارتفع متوسط الجدارة الناتج عن 23 بنداً إلى 0.84 وعلى الرغم من أن جدارة هذا الاختبار الجديد ارتفعت، إلا أنها تسبب خطأً معدلاً أكبر من الاختبار السابق.

وعلى الرغم من أهمية هذه النتائج، إلا أن الجدارة هي شرط لازم وليس كاف لتطوير الاختبار. إن مصداقية النسخ المترجمة المختلفة هي التي تعطي المبرر لاستخدام الدرجات الناتجة عنها.

تكافؤ البنية:

للتأكد من أن النسخ المترجمة لـ PET تقيس البنية نفسها التي تقيسها النسخة العبرية، استخدم آلالوف وباستاري وهامبلتون وسيرسي (1997) تحليل العامل الاستكشافي والقياسي متعدد الأبعاد، وتحليل العامل المؤكد لتقويم التكافؤ البنيوي لاختبار المحاكمة اللغوية في نسختين عبرية وروسية. وقد حللوا بالتحديد أربعة من مجالات المحتوى الخمسة: التشبيهات، النطق، القراءة والفهم، وإتمام الجمل. تضمن التحليل 41 بنداً. وقد وُجد في التحليلات التي أجريت على النسختين أن بنية PET كانت متشابهة عبر النسختين اللغويتين بالنسبة لهذه البنود.

الصدق:

تُختبر عملية الانتقاء بشكل روتيني وذلك باختبار الصدق التنبؤي لـ PET ومقارنته بمعيار متوسط الدرجات النسبي (GPA) في نهاية السنة الدراسية الجامعية لأولى، وعند إتمام الدرجة العلمية الأولى. يركز هذا الفصل على المقارنة بين نتائج المتحنين الناطقين بالعبرية والروسية. أما نتائج دراسة الصدق وانحياز الاختبار، بالنسبة للنسخة العربية منPET، فيمكن مراجعتها في بيلر، غافني وهاناني (1999).

الصدق التنبؤي للنسخة الروسية مقارنة بالنسخة العبرية:

تتأثر الصدق التنبؤي للنسخة الروسية بدرجة أقل بعوامل مثل الفوارق الكبيرة في المقدرات بين المجموعتين الناطقتين بالعبرية والعربية. لهذا تناقش.الأبحاث المتعلقة بالترجمة الروسية بشكل أوسع من الأبحاث المتعلقة بالنسخة العربية.

قامت دراسة أجراها مؤخراً غافني وبرونر (1998) بحساب الصدق التنبؤي لعلامة PET ومقارنتها مع نظيرتها في المجموعة الناطقة بالعبرية. كانت المتنبئات في هذه الدراسة هي PET واختباراته الفرعية الثلاثة (V, Q, and E) ودرجة القبول (Adm) كانت درجة القبول في هذه الدراسة مرتكزة على ثقلين متساويين في PET ودرجة تحصيل معطاة إما في المدرسة الثانوية (باغروت) أو في برنامج تحضره للممتحنين الذين لم يدرسوا في مدرسة ثانوية إسرائيلية. وقد احتوت هذه

الدراسة متنبأً إضافياً هو علامة اختبار إتقان اللغة العبرية (HP) الذي يُجرى لكل الممتحنين غير الناطقين بالعبرية. ثم تم حساب معاملات الصدق للمتنبئات المختلفة للطلاب الذين بدؤوا دراستهم الجامعية بين العامين 1992- 1996 لمعيار في متوسط الدرجات النسبي للسنة الأولى (FGPA) ومتوسط الدرجات النسبي للسنة الثالثة (TGPA) أجريت التحليلات لكل قسم جامعي ضمن كل جماعة شرط أن تضم خمسة طلاب على الأقل من كل من المجموعتين اللغويتين. تم إصدار النتائج لـ463 قسماً حقق الشروط السابقة لـ (FGPA و83 قسماً حقق الشروط لـ TGPA).

يظهر الجدول 12-3 عدد الطلاب والمتوسط والانحراف المعياري لمختلف المتنبئات والمعايير.

الجدول رقم 12-3

المعدلات والانحرافات المعيارية (داخل الأقواس) لدرجات كل من المتنبئ والقاعدة (العبرية والروسية) لكل من TEPA, FGPA

Language	*PET*	*V*	*Q*	*E*	*Adm*	*HP*	*FGPA*	*TGPA*
FGPA								
Hebrew	600	118	117	118	101	–	80	–
N = 55,434	(60)	(13)	(13)	(16)	(6.8)		(8.8)	
Russian	561	111	116	100	99	93	73	–
N = 7,313	(52)	(12)	(12)	(16)	(6.2)	(16)	(11.9)	
TGPA[a]								
Hebrew	590	116	116	116	101	–	82	84
N = 6,612	(57)	(12)	(13)	(15)	(6.2)		(6.7)	(6.9)
Russian	540	108	112	96	98	87	76	81
N = 1,011	(54)	(13)	(12)	(16)	(6.1)	(14.4)	(8.4)	(8.3)

كان الفارق الأكبر بين المجموعتين في اختبار الإنكليزية (E)،(P) حيث أظهر الممتحنون الناطقون بالعبرية أداء أفضل. لم يوجد أي فارق في اختبار المحاكمة

الكمية (Q) أما الفارق في اختبار المحاكمة اللغوية (V)، فكان في الوسط. كما وُجد فارق طفيف لصالح الممتحنين الناطقين بالعبرية في علامة القبول (Adm) مما يوحي بنمط معكوس في الفوارق في درجة التحصيل (غير المعايرة) بالمقارنة مع PET كان الفارق في معيار (FGPA) مماثلاً للفارق في PET، ومن اللافت للانتباه أن الفوارق تقلصت في (TGPA) مقارنة ب (FGPA).

يظهر الجدول 12-4 معاملات الصدق للمجموعتين اللغويتين عبر معدلها في كل الأقسام، وتصحح الارتباطات الملاحظة (بين قوسين) لمحدودية المدى. كان معدل معاملات الصدق لكل من درجة القبولPET في مجموعة (FGPA) متماثلاً في كل من المجموعتين العبرية والروسية عبر كل مجالات الدراسة.

الجدول رقم 12-4

صدق التنبؤ (صححت الارتباطات لمحدودية المدى) لكل من درجة PET؛ درجة القبول، واختبار الكفاءة العبري (HP) والمعدل العام (GPA) في نهاية السنة الجامعية الأولى (FGPA) والسنة النهائية (TGPA) لكل من متحدثي الروسية والعبرية من الممتحنين (يظهر الترابط الخام داخل الأقواس).

Language	*PET*	*V*	*Q*	*E*	*Adm*	*HP*
FGPA						
Hebrew	.39	.32	.36	.24	.48	
	(.26)	(.21)	(.26)	(.12)	(.37)	
Russian	.35	.26	.30	.29	*	*
	(.27)	(.16)	(.21)	(.24)	(.38)	(.23)
TGPA						
Hebrew	.44	.36	.37	.29	.54	
	(.20)	(.16)	(.17)	(.12)	(.28)	
Russian	.45	.35	38	.35	*	*
	(.26)	(.17)	(.19)	(.22)	(.33)	(.20)

معاملات الصدق لكل من درجة القبول PET في مجموعة (FGPA) متماثلاً في كل من المجموعتين العبرية والروسية عبر كل مجالات الدراسة.

لكن نمط معاملات الصدق في الاختبارات الفرعية لـ PET كان مختلفاً في المجموعتين اللغويتين حيث إن Q كان الأعلى مصداقية في المجموعة الناطقة بالعبرية. في حين كان E الأكثر صدقا في المجموعة الناطقة بالروسية مما قد يشير إلى أن V لا يقيس تماماً البنية نفسها في اللغتين. وهذا يعود إما إلى ترجمة تكييف الاختبارات أو بسبب محتوى الاختبار بعينه الذي تم اختياره أصلاً للمجموعة الناطقة بالعبرية. يبدو أن عوامل عدة تحدد مصداقية اختبار في مجموعة معينة، وجودة الترجمة واحد منها فقط. يمكن أن تعزى المصداقية العالية نسبياً لاختبار الإنكليزية (E)، (P) في المجموعة الناطقة بالروسية إلى متغيرات معدلة لم يتم تحريها في هذه الدراسة. فمثلاً يمكن أن يكون الطلاب الذين هاجروا إلى إسرائيل من مدن كبرى ذات نظام تعليمي متطور وحصلوا على فرص أفضل لتعلم اللغة الإنكليزية من أولئك الآتين من بعض المدن البعيدة التي تفتقر إلى نظام تعليمي حديث ومتطور. كما أنه من الممكن أن يكون بعض الممتحنين الناطقين بالروسية قد هاجروا إلى إسرائيل منذ سنين عديدة ودرسوا في النظام التعليمي الإسرائيلي لعدة سنوات قبل تقديم PET وينعكس هذا في درجة اللغة الإنكليزية ودرجتهم المعيارية.

انحياز الاختبار:

قام المعهد الوطني للاختبار والتقويم بإجراء بحث لتقصي وجود انحياز في الاختبار للممتحنين الناطقين بالروسية (غافني، برونر، 1998) ويشير التعبير "انحياز" إلى أخطاء منهجية في الصدق والتنبئي أو صدق المفهوم المتعلق بمجموعة معينة من الممتحنين. تم استخدام الطرق المستمدة من التعريفات الواردة في دارلينغتون (1971) والنقاش الذي قام به لين (1984) لتحري الانحيازات في المتنبئات المختلفة. ويجب تفسير النتائج المتعلقة بمتنبأ واحد بحذر، وذلك لتأثير إخراج متنبأ ما من معادلة راجعة فيها فروق موجودة أصلاً بين المجموعات لينو ورتس (1971).

الانحياز في اختبار المتحنين الناطقين بالروسية:

تتألف العينة الأولى من 55.434 ممتحناً ناطقاً بالعبرية، و7.313 ممتحناً ناطقاً بالروسية. بدؤوا دراستهم في إحدى السنوات 1992- 1996 وكانت درجاتهم في FGPA وPET متوفرة. تم إشراك ست جامعات إسرائيلية وما مجموعه 463 قسماً في الدراسة. كانت درجة القبول (Adm) متوفرة لعينة قريبة تتألف من 26.875 ناطقاً بالعبرية و3.478 ناطقاً بالروسية. أُجري التحليل لهذا المتنبأ في 259 قسماً فقط (غافني، برونر 1998). يظهر الجدول 12-5 عدد حالات الانحياز المهمة التي تم الكشف عنها في كل الأقسام واستخدام درجة القبول وAdm، PET كمتنبئين.

جدول رقم 12-5

عدد حالات الانحياز المهمة التي تم الكشف عنها في كل الأقسام واستخدام درجة القبول وPET،Adm كمتنبئين

Predictor	*Bias Against Russian Speakers*	*Bias Favoring Russian Speakers*
FGPA		
PET	3.0	2.0
V	0.5	1.5
Q	1.0	7.0
E	1.5	0.0
Adm	1.0	9.0
TGPA		
PET	8.0	5.5
V	2.7	5.5
Q	2.7	5.5
E	11.0	0.0
Adm	2.7	0.0

لم يظهر انحياز واضح عند استخدام PET كمتنبئ منفرد: ففي 3٪ من الـ463 حالة كان هناك مؤشر واضح على الانحياز ضد المتحنين الناطقين بالروسية، وكان الانحياز في 2 ٪ لصالحهم. وقد وُجدت نتائج مماثلة لـ "V" و "Q" (الاختبارات

obeikandl.com

المترجمة) مع ميلها إلى مبالغة النتائج المتوقعة لـ FGPA للمجموعة الناطقة بالروسية. وقد وجد ميل معاكس في اختبار الإنكليزية E أما في درجات القبول، فكان هناك في حوالي 10 ٪ من الـ 259 قسماً مؤشرات واضحة على الانحياز، وغالباً لصالح الممتحنين الناطقين بالروسية.

لا تعطى النتائج المذكورة آنفاً في إجابات واضحة على انحياز التنبؤ الممكن حدوثه نتيجة الترجمة. فقد أظهر اختبار الإنكليزية E، وهو غير مترجم، انحيازاً ضد الناطقين بالروسية، بينما أظهر اختبار المحاكمة الكمية Q، وهو غير متأثر بالترجمة نسبياً، انحيازاً لصالح هذه المجموعة. أما اختبار المحاكمة اللغوية الأكثر تأثراً بالترجمة فقد أظهر نتائج لا تشير إلى أي مشكلات أكيدة في الترجمة. يمكن الاستنتاج أنه إذا نجحت عملية الترجمة في الحفاظ على مستوى الصعوبة نفسه في النسختين اللغويتين، وإذا كان معنى ما يتم قياسه متشابهاً قدر الإمكان، فلا يجب توقع أي انحياز نتيجة للترجمة بحد ذاتها.

إن أحد الانتقادات الرئيسة لدراسات الانحياز التقليدية هي أنها تبالغ في تنبؤات الدرجات المعيارية لمجموعات الأقليات، إن النقلة إلى كلية يشكل فيها الغالبية السكانية القسم الأكبر من الطلاب يفرض متطلبات أكثر على طلاب الأقليات منه على طلاب الغالبية السكانية. ولهذا يتوقع المرء أن تختفي المبالغة التنبئية في السنة الجامعية الثالثة. تم اختبار عينة فرعية مؤلف من (6.612) ممتحناً ناطقاً بالعبرية و (1.011) ممتحناً ناطقاً بالروسية تتوفر درجاتهم فيPET وFGPA لاختبار هذا الافتراض. وقد ضم هذا الاختبار 83 قسماً. كانت درجة القبول (Adm) متوفرة لعينة فرعية مؤلفة من (2.687) ممتحناً ناطقاً بالعبرية و(338) ممتحناً ناطقاً بالروسية. توفر التحليل لهذا المتنبأ في 37 قسماً فقط. يظهر الجدول 12-5 عدد الأقسام التي تم فيها اكتشاف انحياز واضح ضد ولصالح الممتحنين الناطقين بالروسية. وبشكل عام، عند استخدام TGPA كمعيار، ظهر نمط مماثل لـ FGPA في انخفاض بسيط من ميل المتنبئات إلى انحياز لصالح المجموعة الناطقة بالروسية كما هو متوقع.

الاستنتاجات

يجب مقاربة قضية ترجمة الاختبارات مع التفكير بالأبعاد المختلفة هدف الترجمة، المجموعة السكانية المستهدفة ونوع ومحتوى الاختبار. يمكن ترجمة اختبار ما لأغراض الأبحاث حيث تكون الأهمية الأولى للفروق بين المجموعات. كما يمكن استخدام الترجمة لاتخاذ قرارات فردية مهمة مثل القبول في الجامعات. إن متطلبات جودة الترجمة أعلى في الاستخدام اللاحق من السابق.

تم وصف عملية ترجمة PET وهو اختبار يستعمل في القبول في التعليم العالي في إسرائيل من العبرية إلى العربية والروسية والفرنسية والإسبانية والإنكليزية بشكل مفصل مع توضيح المشكلات المتأصلة في الترجمة (بشكل خاص في الفقرات اللغوية). تمت مراجعة جودة الترجمة بتطبيق طرائق نوعية وكيفية متنوعة، وذلك لإيضاح الجوانب المختلفة في عملية الترجمة للنسخ اللغوية المختلفة. تم اتخاذ خطوات متعددة لتأمين جودة الترجمة:

(أ) استثمار جهد كبير في المراجعة النوعية للترجمة مع استخدام مترجمين مستقلين في بعض الحالات.

(ب) اختبار أنماط الإجابات والميل التفاوتي للممتحنين إلى التخمين في بنود الخيارات المتعددة.

(ج) اختبار تحليل البنود والأداء التفاوتي للبنود.

(د) اختبار الجدارة وعلاقتها بمستوى قدرات المجموعة.

(هـ) تحري الصدق التنبئي لمعيارين (معدل الدرجات النسبي للسنتين الدراسيتين الجامعية الأولى والثالثة).

(و) تحليل انحياز الاختبار التنبئي للمجموعات الفرعية المختلفة.

إن المجموعة الواسعة من التحليلات المقدمة في هذا الفصل والمتعلقة بـPET، والنسخة المترجمة تعطي الكثير من المعلومات حول تعقيد مسألة تساوي وتكافؤ

اختبارات القبول المترجمة. يؤخذ هذا الرأي القائل إنه يجب تبني وجهة نظر أوسع لعدالة الاختبار وتكافؤ النسخ. عند تقويم جودة الاختبارات المترجمة في سياق قرارات فردية مهمة، يجب القيام ببعض الصدق التنبئي وانحياز الاختبار، بالإضافة إلى التحليلات الأخرى الأكثر شيوعاً مثل اختبار الأداء التفاوتي للبنود. إن مسألتي صدق الاختبار وانحيازه ضروريتان لوضع الأساس لاستخدام نتائج الاختبار، ليس أقل أهمية من الاختبار نفسه في سياق اختبارات القبول.

تظهر النتائج المقدمة في هذا الفصل أنه عند تطبيق عملية ترجمة جيدة لـPET، يمكن إنتاج مجموعة من الاختبارات المترجمة المتكافئة البينية والجدارة الصحيحة والعادلة نسبياً. ولكن حتى عندما تُتخذ كل هذه الخطوات للتأكد من جودة الترجمة وتساوي الدرجات، يظل الحصول على نسختين واحدة أصلية وأخرى مترجمة أمراً مستحيلاً. ويظل الخيار الآخر الذي يقضي باختبار الناطقين بغير العبرية حلاً أقل عدالة. كما يجب اعتبار عوامل أخرى مثل التأثيرات الاقتصادية والفوائد المتوقعة (أي تقدير الكلفة المرافقة لتحسين العملية الحالية).

شكر

نرغب بشكر مايا بار- هيدل، غيرشون بن- شاكار، روث فورتس، شافا كاسلوجيري ليفينسون لملاحظاتهم العميقة، وشمول برونر لمساعدته في تحليل المعطيات.

*****

المراجع

Allalouf, A., Bastari, B., Hambleton, R. K., & Sireci, S. G. (1997). *Comparing the dimensionality of a test administered in two languages* (Laboratory of Psychometric and Evaluative Research Rep. No. 319). Amherst, University of Massachusetts, School of Education.

Allalouf, A., Hambleton, R. K., & Sireci, S. G. (1999). Identifying the causes of DIF in translated verbal items. *Journal of Educational Measurement, 36*(3), 185–198.

Angoff, W. H. (1972, August). *A technique for the investigation of cultural differences.* Paper presented at the meeting of the American Psychological Association, Honolulu. (ERIC Document Reproduction Service No. ED 069686)

Angoff, W. H., & Cook, L. L. (1988). *Equating the scores of the Prueba de Aptitud Academica and the Scholastic Aptitude Test* (College Board Rep. No. 88-2). New York: College Entrance Examination Board.

Angoff , W. H., & Modu, C. C. (1973). *Equating the scales of the Prueba de Aptitud Academica and the Scholastic Aptitude Test* (Research Rep. No. 3). New York: College Entrance Examination Board.

Beller, M. (1994). Psychometric and social issues in admissions to Israeli universities. *Educational Measurement: Issues and Practice, 13*(2), 12–20.

Beller, M., & Gafni, N. (1995). Translated scholastic aptitude tests. In G. Ben-Shakhar & A. Lieblich (Eds.), *Studies in psychology* (pp. 202–219). Jerusalem: The Hebrew University, The Magnes Press.

Beller, M., Gafni, N., & Hanani, P. (1999, June). *Constructing, adapting, and validating admissions tests in multiple languages.* An invited paper presented at the International Conference on Adapting Tests for Use in Multiple Languages and Cultures, Georgetown University, Washington, DC.

Darlington, R. B. (1971). Another look at "cultural fairness." *Journal of Educational Measurement, 8,* 71–82.

Gafni, N., & Bronner, S. (1998, April). *An examination of criterion-related bias for Hebrew- and Russian-speaking examinees in Israel.* Paper presented at the meeting of the American Educational Research Association, San Diego.

Gafni, N., & Canaan-Yehoshafat, Z. (1993). *An examination of deferential item functioning for Hebrew and Russian-speaking examinees in Israel.* Paper presented at the 24th Annual Conference of the Israeli Psychological Association, Ramat-Gan.

Gafni, N., & Melamed, E. (1994). Differential tendencies to guess as a function of gender and lingual-cultural reference group. *Studies in Educational Evaluation, 20,* 309–319.

Grisay, A. (2003). Translation procedures in OECD/PISA 2000 international assessment. *Language Testing, 20*(2), 225–240.

Hambleton, R. K. (1994). Guidelines for adapting educational and psychological tests: A progress report. *European Journal of Psychological Assessment, 10,* 224–229.

Linn, R. L. (1984). Selection bias: Multiple meanings. *Journal of Educational Measurement, 21,* 33–47.

Linn, R. L., & Werts, C. E. (1971). Considerations for studies of test bias. *Journal of Educational Measurement, 8*(1), 1–4.

Poortinga, Y. H. (1995). Cultural bias in assessment: Historical and thematic issues. *European Journal of Psychological Assessment, 11*(3), 140–146.

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

obeikandl.com