

Las pruebas psicométricas en la investigación empírica sobre los procesos cognitivos del traductor

ALICIA BOLAÑOS MEDINA
Universidad de Las Palmas
Grupo de investigación PETRA
abolanos@dfm.ulpgc.es

Resumen

El creciente interés académico por la traductología cognitiva ha venido acompañado de un incremento de los diseños de investigación empírica sobre los procesos mentales del traductor (Fernández Sánchez y Muñoz Martín 2008). Entre otros aspectos metodológicos, en este tipo de experimentos es necesario atender a las diferencias de carácter psicológico de los sujetos de la muestra. Según sea el diseño de investigación, dichas diferencias individuales pueden actuar como variables independientes organísmicas, manipuladas por el investigador para determinar su efecto, o como variables extrañas, no relacionadas directamente con el propósito del estudio. En ambos casos es necesaria su cuantificación.

En este trabajo se describen los diferentes tipos de instrumentos de evaluación psicológica al alcance del investigador. A continuación, se presentan los tests más utilizados para medir diferentes constructos psicológicos que pueden ser de interés para la traductología. Por último, tras revisar los factores que se deben tener en cuenta a la hora de elegir la prueba psicométrica más adecuada, se exponen algunas recomendaciones para la administración de los tests estandarizados, en consonancia con las normas profesionales del principal organismo regulador internacional de este campo.

Palabras clave: traductología cognitiva, pruebas psicométricas, metodología de investigación, procesos cognitivos del traductor

Abstract

The recent interest in cognitive translatology has entailed a growing number of empirical study designs in translation process research (Fernández Sánchez & Muñoz Martín 2008). Among other methodological issues, subject profiling becomes a key factor if we are to control an experiment. Individual differences can act as independent variables, manipulated by researchers in order to gauge their effect, or extraneous variables, not directly related to the purpose of the study, which need to be controlled. In both cases, it is advisable to quantify them.

Different types of instruments for psychological evaluation are described and most used tests measuring psychological constructs of interest for translatology are presented, together with some factors to bear in mind when selecting the appropriate test for a given study design. Finally, according to the Standards for educational and psychological testing, some recommendations for adequate test administration are given.

Keywords: cognitive translatology, psychometric tests, research methodology, subject-profiling, translation process research

1. INTRODUCCIÓN

En los últimos años, el creciente interés académico por la traductología cognitiva ha venido acompañado de un incremento de los diseños de investigación empírica sobre los procesos mentales del traductor (Fernández Sánchez y Muñoz Martín 2008), pero también se ha generado cierta preocupación por la metodología utilizada. Jääskeläinen (2000: 73) propone el enriquecimiento de los estudios empíricos sobre traducción a partir de la experiencia acumulada por la investigación psicológica y, en la misma línea que Hansen (2003), afirma que es necesario obtener más información sobre el perfil de los sujetos experimentales de la que hasta ahora se ha venido recopilando en nuestro campo. Las conclusiones de una reciente investigación preliminar de Muñoz Martín (en prensa) parecen apuntar en la misma dirección.

Entre otros aspectos metodológicos, en este tipo de experimentos es necesario atender a las diferencias de carácter psicológico de los sujetos de la muestra. Según sea el diseño de investigación, dichas diferencias individuales pueden actuar como variables independientes organísmicas, manipuladas por el investigador por medio de selección para determinar su efecto, o como variables extrañas, no relacionadas directamente con el propósito del estudio, y que habrá que controlar. En ambos casos es aconsejable su cuantificación.

Ante la importancia de la elección de los instrumentos adecuados de recogida de datos sobre el perfil de los sujetos, así como de su correcta administración, en este trabajo se describen los diferentes tipos de pruebas psicométricas al alcance del investigador y se presentan los tests más utilizados para medir diferentes constructos psicológicos, responsables de la variabilidad intersujetos, que pueden ser de interés para la traductología. Por último, tras revisar los factores que se deben tener en cuenta a la hora de seleccionar el instrumento más adecuado para cada diseño de investigación, se exponen algunas recomendaciones para la administración de los tests estandarizados, en consonancia con las normas profesionales del principal organismo regulador internacional de este campo.

2. TIPOS DE INSTRUMENTOS DE EVALUACIÓN PSICOLÓGICA AL ALCANCE DEL INVESTIGADOR

Los instrumentos de evaluación psicológica son procedimientos de recogida de información que permiten la observación y descripción de determinadas conductas y se pueden clasificar en autoinformes, observación, técnicas subjetivas, técnicas proyectivas y técnicas objetivas (Moreno Rosset 2005: 93).

Entre los tipos de autoinforme se encuentran: los cuestionarios, las escalas, los inventarios, las técnicas de autoregistro y la entrevista. Los primeros se caracterizan por recoger los ítems en forma de interrogación. Ejemplos cercanos de su utilización en el ámbito de la investigación empírica sobre traducción, son el cuestionario sociolingüístico del grupo PETRA (Marín García 2008: 63) y los cuestionarios de problemas y sobre conocimientos de traducción del grupo de investigación PACTE (Presas 2006).

Las escalas son conjuntos de enunciados que sirven para medir el grado de acuerdo o desacuerdo de un individuo hacia su contenido y son propias de la valoración, entre otros, de los aspectos motivacionales. Los inventarios recogen una lista de rasgos, preferencias, actitudes, intereses o capacidades con el fin de evaluar las características o habilidades de un individuo y se suelen utilizar a menudo para valorar aspectos de la personalidad.

La técnica de autorregistro, una de las más utilizadas en el ámbito clínico, presenta la gran ventaja de, a diferencia de la entrevista, permitir la recogida de las variables de interés en el mismo momento en el que ocurren. En la misma línea, los protocolos de verbalización del pensamiento han sido ampliamente utilizados en el ámbito de la investigación sobre los procesos cognitivos del traductor. El método ha ido evolucionando y, en la actualidad, coexisten la verbalización en tiempo real, la retrospectiva y su uso conjunto con el diálogo. Las sesiones se graban para su posterior análisis. Entre los aspectos más criticados por sus detractores se encuentran el hecho de que no tienen en cuenta los procesos automáticos implicados en la traducción y la distorsión que produce en la actividad traductora y en los resultados la propia verbalización (Halverson 2009).

La entrevista implica una conversación y relación interpersonal con unos fines dados, en la que cada interviniente adopta un rol determinado. Existen diferentes tipos; según sea su finalidad, se pueden distinguir la encaminada a decidir sobre la adscripción o no de un sujeto a una investigación concreta; la llamada «diagnóstica», que permite contrastar la información recogida con la obtenida mediante otros instrumentos o la consultiva, para dar respuesta a una pregunta concreta. También podemos diferenciar la estructurada, estandarizada y con guión preestablecido, con la ventaja de que se pueden resolver las dudas que surjan a cada individuo sobre la marcha; en la semiestructurada, aún existiendo un

guión, se pueden incluir preguntas espontáneas durante su desarrollo. Por último, en la entrevista libre, las preguntas abiertas permiten al entrevistado decidir en qué aspectos prefiere abundar. También es necesario tener en cuenta las características de los sujetos, en especial su edad y formación académica, así como el momento del proceso de investigación en el que tienen lugar, si es al comienzo, durante la fase de recogida de datos o al final, como es el caso de la entrevista retrospectiva de PACTE (Presas 2006).

La metodología observacional permite la recogida sistemática de datos de la realidad sin ejercer ningún tipo de control sobre las variables. En la observación participante, se obtienen los datos del contacto directo con el sujeto, mientras que en la no participante apenas se produce interacción. En la modalidad de participación y observación, ambas figuras se acercan dada la relación previa entre ellos (por ejemplo, profesor-estudiante). Hay que tener precaución para evitar la aparición de los tres sesgos principales: la reactividad, o alteración de la conducta de los sujetos al saberse observados, la influencia de las expectativas del observador en lo que cree percibir, y los sesgos de tipo técnico que puedan afectar al equipo usado o a la organización del proceso.

En los últimos años, gracias a la aparición de nuevos instrumentos tecnológicos, es posible observar y registrar la conducta de los sujetos de una forma mucho más precisa (Dam-Jensen & Heine 2009: 26). Más allá de las grabaciones en vídeo, la aplicación Translog, desarrollada por Jakobsen, registra en vivo el proceso de escritura de la traducción, al recoger las pulsaciones en tiempo real y, al contrario de los métodos de verbalización del pensamiento, no interfiere en el proceso. Se estudia el ritmo y la velocidad de creación del texto traducido, y entre los parámetros investigados hasta el momento se encuentran la segmentación y revisión textuales, la conexión entre las pausas y las acciones textuales inmediatamente anteriores o posteriores, la distribución de las subtareas y el ritmo de alternancia entre las pulsaciones y las pausas (Halverson 2009: 214). Los programas de captura de pantalla, que cuentan con una elevado índice de validez ecológica, permiten registrar todas las actividades que se realizan en un ordenador en forma de vídeo digital, aunque no ofrecen información sobre qué segmentos e imágenes captan la atención de los usuarios. Éste último problema se puede solucionar con el uso conjunto de herramientas de seguimiento del globo ocular, también cada vez más utilizadas, y que permiten relacionar el esfuerzo cognitivo y el movimiento ocular al calibrar los puntos en que se fija la mirada, su recorrido y el tamaño de la pupila de los sujetos (Göpferich *et al.* 2008).

Por métodos subjetivos entendemos los destinados a recopilar información sobre la estructura psicológica, los contenidos y los procesos de las visiones subjetivas o significados personales acerca de uno mismo y el mundo. Los más extendidos son la técnica de rejilla, el diferencial semántico y los métodos her-

menéuticos y narrativos, pero, dada su poca relación con los contenidos tradicionalmente investigados en nuestra disciplina, y las limitaciones de espacio de este artículo, no nos detendremos en ellos. Algo parecido sucede con las técnicas proyectivas, tradicionalmente usadas para el estudio de la personalidad a partir de las «proyecciones» (percepciones, sentimientos, etc.) de los sujetos ante series de estímulos estandarizados (tests de Rorschach y de apercepción temática), en las que tampoco profundizaremos.

A la hora de medir las diferencias individuales, los tests psicológicos estandarizados suelen ser la herramienta de elección, ya que, por ser objetivos, sistemáticos y tipificados, permiten comparar la conducta de dos o más personas con ciertas garantías de fiabilidad. Se definen como un instrumento o procedimiento en el que se recoge una muestra de la conducta de un sujeto, en un ámbito específico, para su evaluación mediante una puntuación asignada siguiendo un proceso estandarizado. Las instrucciones dadas a los sujetos, las condiciones en que se realiza la prueba y el proceso de asignación de puntuaciones de cada test sigue un patrón específico.

Existen varias categorías de tests psicométricos, según sea la naturaleza del constructo que pretenden medir. Los tests de inteligencia pretenden calibrar la capacidad básica y potencial de un individuo para comprender el mundo que le rodea, asimilar su funcionamiento y aplicar dicho conocimiento para mejorar su calidad de vida. Los tests de personalidad se utilizan con fines de investigación o diagnóstico. Los de rendimiento y de aptitud, se suelen utilizar en entornos educativos y de recursos humanos, e intentan medir el grado de conocimientos de un sujeto sobre un tema concreto o hasta qué punto domina un área. Los neuropsicológicos evalúan los posibles déficits del funcionamiento cognitivo que pueden provenir de algún tipo de daño cerebral. Los ocupacionales intentan emparejar los intereses de un individuo con los de diferentes profesiones para hallar cuál es la que más se adapta a su perfil. Los clínicos miden factores tales como el nivel de ansiedad o depresión de un individuo. En el apartado siguiente trataremos algunas pruebas pertenecientes a las dos primeras categorías.

3. CONSTRUCTOS Y PRUEBAS PSICOMÉTRICAS DE INTERÉS PARA LA INVESTIGACIÓN SOBRE LOS PROCESOS COGNITIVOS DEL TRADUCTOR

El test de inteligencia más utilizado en investigación traductológica en nuestro país es el WAIS III, del inglés *Weschler Adult Intelligence Scales*, que permite obtener, además de tres puntuaciones de coeficiente intelectual (verbal, manipulativo y total, a partir de los dos anteriores), otros cuatro índices de forma independiente, tres de los cuales revisten interés para nuestra disciplina: la escala de memoria operativa, con los subtests de aritmética, dígitos y letras y números;

la escala de velocidad de procesamiento, con los subtests de clave de números y búsqueda de símbolos, y la escala de comprensión verbal, con los subtests de vocabulario, semejanzas e información. Dadas las limitaciones de espacio de este artículo, remitimos al lector al manual de Kaufman & Lichtenberger (1999) sobre la aplicación e interpretación de la prueba y a la visión de conjunto ofrecida por Muñoz Martín (en prensa) sobre la aplicabilidad de estos tres factores en la investigación traductológica.

La Torre de Hanoi es un test utilizado para medir el funcionamiento ejecutivo, mediante la manipulación de diferentes discos de radio creciente que se apilan en una estaca hasta insertarlos en otra en el mismo orden siguiendo ciertas reglas. En traductología, este test se ha utilizado en investigación empírica sobre creatividad y automatización al traducir, para comprobar la relación entre el potencial de resolución de problemas bien definidos y el de problemas mal definidos (Marín García 2008: 23).

Allport acuñó el término «estilo cognitivo» en 1937, pero fueron los psicólogos cognitivos de los años cincuenta y sesenta del siglo pasado, quienes mostraron que existen claras diferencias individuales en el modo habitual en que los sujetos procesan la información y utilizan sus recursos, de cara a un proceso de toma de decisiones o de resolución de problemas (Sánchez-Elvira Paniagua 2003); aunque se mantienen de manera consistente y estable a lo largo de la vida, pueden modificarse mediante entrenamiento.

Desde la psicología, dependiendo del autor, se han identificado diversas categorías de estilos cognitivos, de acuerdo con la polaridad extrema hacia la que tienden. Entre las dimensiones más estudiadas de interés para la traductología, se encuentran las siguientes: impulsividad-reflexividad, según tienda un individuo a manifestar la primera solución que viene a su mente aunque sea incorrecta, desconociendo los errores y tomando decisiones rápidamente o dejen transcurrir un tiempo antes de proponer una solución evitando cometer errores; independiente de campo-dependiente de campo, al valorar si se atiende en mayor medida a un aspecto determinado de la información o se percibe un fenómeno como un todo unitario y se recurra más al contexto; serial-holístico, dependiendo de si se analizan en detalle todos los elementos de un problema y se ordenan de forma secuencial o se procesan varios simultáneamente para formar una unidad compleja (Fernández-Jiménez 2003).

Entre los tests más usados para determinar los estilos cognitivos se encuentran los siguientes: el Test de emparejamiento de figuras familiares o MFFT, del inglés *Matching Familiar Figures Test* (Kagan 1965), para medir la dimensión reflexividad-impulsividad, en el que el sujeto selecciona entre varias figuras alternativas la que se ajusta a la de referencia y obtiene dos índices, la rapidez de respuesta y el número de errores; el EFT, del inglés *Embedded Figures Test*,

de Witkin et al. (1971) para calibrar la dependencia e independencia de campo mediante la identificación de una figura geométrica simple en el contexto de otra más compleja; y el Test de Stroop (Golden 1975) para valorar la flexibilidad-rigidez cognitiva, al tener en cuenta la capacidad de cambio de una estrategia al inhibir la respuesta habitual y ofrecer una nueva ante ciertos estímulos.

No muy a menudo se han establecido empíricamente vínculos entre personalidad y procesos cognitivos del traductor, lo que no implica que este ámbito no revista cierto interés y pueda ser objeto de estudio en los próximos años. Así Jääskeläinen (2000: 73) sugiere que ciertas diferencias halladas en el rendimiento de algunos sujetos experimentales podrían deberse a rasgos de personalidad. Según Tirkkonen-Condit (2000: 141), la tolerancia a la incertidumbre, que forma parte de todo proceso cognitivo que implica una toma de decisiones, es un rasgo de la personalidad que merece ser tenido en cuenta en la formación de traductores. Se trata de un factor poco estudiado, en gran medida por la falta de instrumentos adecuados; sin embargo, la escala de intolerancia a la incertidumbre de Freeston et al. (1994) ha sido recientemente adaptada al español (González Rodríguez et al. 2006) con índices apropiados de validez y fiabilidad.

Desde que Bandura (1977) introdujo el concepto de autoeficacia o creencia en la propia capacidad de organizar y ejecutar los cursos de acción necesarios para producir logros concretos, ésta se convirtió en pieza importante de la investigación de la Psicología Social Cognitiva aplicada a campos tales como el aprendizaje, el rendimiento y la motivación (Salanova et al. 2004). Se trata de una variable mediadora en procesos tan variados como la toma de decisiones o el éxito de la enseñanza y el aprendizaje, objetos de interés de la traductología. En la actualidad existe una adaptación española de la Escala de autoeficacia general (Baessler y Shwarcer 1996), con datos de fiabilidad adecuada y capacidad predictiva considerable (Sanjuán Suárez et al. 2000).

La atención de muchos investigadores se ha dirigido recientemente hacia otro constructo estrechamente relacionado con el anterior: la eficacia colectiva percibida o el juicio o valoración de los miembros del grupo de su capacidad de rendimiento a la hora de realizar una tarea concreta. Se trata de una propiedad que emerge a nivel grupal, no simplemente de la suma de las creencias de eficacia de cada uno de sus miembros, y afecta, entre otros, a la forma en que el grupo utiliza sus recursos, al esfuerzo invertido en conseguir sus metas e incluso a su vulnerabilidad al descontento ante las dificultades (Bandura 2000). Un hallazgo importante es el hecho de que los grupos altos en eficacia colectiva son capaces de persistir incluso después de tener fallos, mostrándose todavía más eficaces (Little y Madigan 1997). Aunque aún estén en proceso de validación, existen instrumentos pilotos disponibles para medir la autoeficacia colectiva (Huici et al. 2004).

Por último, y siguiendo en el ámbito de estudio de los grupos de trabajo, su complejidad estructural y funcional se manifiesta también en el fenómeno de la cohesión grupal, expresión singular de su integración afectiva, valorativa y funcional. Tradicionalmente, en la investigación aplicada, la cohesión grupal se ha operativizado como la atracción que los miembros del grupo sienten entre sí y con el propio grupo, parámetro que se suele medir mediante autoinformes y cuestionarios. Carron et al. (1985) distinguieron dos bloques fundamentales de aspectos relacionados con la cohesión: la dimensiones individuo-grupo y tarea-relación; esta es la base de su cuestionario de Ambiente de Grupo (GEQ). En esta línea, un estudio preliminar realizado en el aula universitaria apunta la existencia de una estrecha vinculación entre cohesión grupal y eficacia colectiva percibida en estudiantes de traducción (Bolaños Medina 2008).

4. FACTORES CLAVE PARA SELECCIONAR LA PRUEBA PSICOMÉTRICA ADECUADA

El punto de partida a la hora de elegir la prueba psicométrica más adecuada para un diseño de investigación, es definir con exactitud qué constructo se pretende evaluar; con qué objetivo, por ejemplo, si se trata, en una primera fase de la investigación, de preseleccionar a los sujetos en función de sus puntuaciones para su inclusión en diferentes grupos de tratamiento, o si es necesario obtener una medida lo más exacta posible de cada uno de ellos para triangular los resultados; y, por último, las principales características de la muestra poblacional concreta a la que lo vamos a aplicar, como son la edad, su disponibilidad, nivel sociocultural, etc.

A continuación, es necesaria una fase de documentación sobre las pruebas, de preferencia estandarizadas, disponibles para medir dicho parámetro. Se debe atender, en particular, a si pueden aplicarse a sujetos de edades similares a las de la muestra seleccionada, si son de aplicación individual o grupal, el tiempo de administración, el formato, y su adecuación al objetivo principal del estudio; por ejemplo, si se trata de seleccionar a los individuos que obtengan las puntuaciones inferiores o superiores, se puede aplicar una prueba grupal para su detección y luego una individual a los sujetos preseleccionados. Asimismo, esta fase implica también recopilar información sobre la validez y la fiabilidad de las pruebas psicométricas para poder seleccionar las que, siendo aplicables dadas las condiciones de nuestro experimento, presenten los criterios de calidad más sólidos.

La validez se relaciona directamente con el hecho de que una prueba mida efectivamente la dimensión que se supone que ha de calibrar. Se trata de una estimación sobre la veracidad de las inferencias realizadas a partir de las puntuaciones obtenidas (AERA, APA & NCME 1999: 9), y en su evaluación se suelen utilizar

tres estrategias que no son mutuamente excluyentes: el análisis del contenido, el análisis del constructo medido y el contraste de las puntuaciones de la prueba con otras medidas o criterios de interés.

El objetivo de la validez de contenido es «analizar hasta qué punto los elementos o ítems que componen el test son una muestra relevante y representativa del constructo sobre el que se van a realizar las inferencias» (Barbero García 2006: 311) y se suele aplicar con mayor frecuencia en pruebas de rendimiento o conocimiento. Se recurre a un grupo de expertos para garantizar que la prueba no incluya aspectos irrelevantes pero sí todos los más importantes.

La validez de constructo pretende comprobar que la variable medida por el test es aceptable dentro de un sistema teórico. Tras definir el concepto científico de interés, a partir de las bases teóricas existentes, se establecen hipótesis sobre la relación, por un lado, entre éste y ciertas conductas directamente observables, y, por el otro, entre él y otros constructos dados; a continuación se diseña un instrumento de medida para obtener datos empíricos de las interrelaciones entre las puntuaciones de la prueba y las variables hipotetizadas.

La validez referida al criterio o predictiva implica averiguar en qué medida las puntuaciones de la prueba pueden ser utilizadas para inferir el comportamiento de los sujetos en una variable independiente diferente, denominada criterio; es decir, se trata de «evaluar la hipótesis de relación entre test y criterio» (Barbero García *et al.* 2006: 321) para expresar «la convergencia de indicadores» (Moreno Rosset 2005: 118).

Los valores del coeficiente de validez están comprendidos en el intervalo -1 y +1; los valores cercanos a 1 indican una fuerte asociación lineal positiva entre las dos variables, y los valores próximos a -1 señalan una relación negativa muy intensa, no existiendo relación alguna entre las variables cuando el coeficiente de correlación es igual a 0. Por lo tanto, cuanto mayor sea éste, mayor será la validez, pero hay que tener en cuenta que suelen ser más bajos que los de fiabilidad (entre 0,20 y 0,30 se consideran muy bajos y los superiores a 0,90, muy altos).

Por su parte, la fiabilidad está directamente relacionada con la estabilidad y precisión de la medida, es decir, en qué grado se obtendrán puntuaciones similares cuando se evalúe un mismo rasgo, con el mismo test en condiciones parecidas, y hasta qué punto se puede afirmar que dichos valores están libres de errores de medida; es decir, es la proporción de la varianza observada que se explica por la varianza real, al descontar la varianza de errores no sistemáticos de medición.

Los procedimientos más utilizados para el cálculo de la fiabilidad son cuatro. El método conocido como «test-retest» mide la estabilidad de la prueba y calcula el coeficiente de fiabilidad mediante la correlación entre las puntuaciones obtenidas por un grupo de sujetos en una primera aplicación y las halladas en una segunda; de este modo se toman en consideración posibles errores de

medida provenientes de las diferencias entre las dos condiciones y tiene la gran ventaja de no requerir más formas distintas del mismo test. El método de formas paralelas implica, como su nombre indica, construir dos versiones del test, de manera que los errores derivados de los reactivos y de las ocasiones diferentes de aplicación se puedan evitar. Esta vez, se administra a la mitad de la muestra el primer test y a la otra mitad el segundo en un primer momento, para cruzarlos más tarde en la segunda aplicación.

El tercer método se basa en la consistencia interna y sólo requiere una aplicación del test, por lo que permite controlar factores como el intervalo entre pruebas y la memoria o el aprendizaje. Una vez obtenidas las puntuaciones, se divide el test en dos mitades de similar dificultad y contenido para calcular la correlación entre las puntuaciones de los sujetos en ambas partes mediante diferentes métodos estadísticos. Por último, también se puede evaluar la consistencia entre dos o más jueces o evaluadores a la hora de corregir la prueba, aunque este método se suele considerar como complementario de los anteriormente descritos, no como sustitutorio.

Forns i Santacana et al. (2002) recomiendan rechazar las pruebas que presenten un coeficiente de fiabilidad bajo (inferior al 0,65), utilizar las de fiabilidad media (entre el 0,75 y el 0,85) únicamente como pruebas de cribado o preliminares y, siempre que sea posible, recurrir a pruebas de alta fiabilidad (por encima del 0,90) cuando el objetivo sea tomar una decisión relevante. En cualquier caso, no hay que olvidar que «una prueba puede ser fiable sin ser válida, pero no puede ser válida sin ser fiable» (Moreno Rosset 2005: 116).

Por último, a la hora de seleccionar la prueba psicométrica estandarizada adecuada, cabe analizar si obtener puntuaciones referidas al criterio, que muestren un determinado grado de habilidad en términos absolutos, es todo lo que necesitamos, o sería preferible puntuaciones referidas a la norma. En psicometría, «norma» se refiere a la información estadística que resume la distribución (variabilidad) de las puntuaciones obtenidas en un test por un grupo específico de sujetos. La norma, es, en pocas palabras, la puntuación de referencia de un parámetro concreto (el que medimos) para un grupo poblacional. Si comparamos la puntuación obtenida por un sujeto dado con la norma de su grupo de referencia, tendremos información valiosa para la interpretación del resultado del test. Si optamos por éste último tipo, será necesario también considerar la información normativa acumulada de la investigación previa realizada. Por ejemplo, en el caso de los tests provenientes de otros países, es importante que hayan sido convenientemente adaptados a nuestra cultura; además de haber sido traducidos, deben haber sido aplicados a diferentes grupos poblacionales con su consiguiente análisis estadístico.

5. RECOMENDACIONES PARA LA ADMINISTRACIÓN DE LOS TESTS ESTANDARIZADOS

Cuando las pruebas psicométricas se aplican en un contexto docente, existen ciertas restricciones impuestas por la ética, por ejemplo, la imposibilidad de utilizar un grupo de control durante un tiempo prolongado, ya que todos los estudiantes deben recibir la misma metodología, contenidos y carga docentes o el número limitado de horas lectivas, que impide dedicar el tiempo teóricamente necesario a la administración de las pruebas psicométricas y puede llevar a soluciones tales como la administración de tests creados para una aplicación individual de forma grupal.

Siguiendo los estándares sobre la aplicación de tests educativos y psicológicos de la American Educational Research Association, la American Psychological Association y el National Council on Measurement in Education, (AERA, APA & NCME 1999: 18), los principales organismos profesionales reguladores de este campo, si se va a utilizar un test de un modo que no ha sido enteramente validado o se va a modificar en algún aspecto el procedimiento estandarizado de administración, es necesario que se justifique su nuevo uso y se recojan tanto la información pertinente como aquellos datos que justifiquen su validez.

En aquellos casos en los que, dada la inexistencia o falta de disponibilidad de pruebas psicométricas de validez contrastada, se utilicen otras cuya fiabilidad o validez aún no se ha comprobado en una muestra suficientemente amplia, se deben interpretar los resultados con precaución; siempre que sea posible, se intentará triangular los datos obtenidos.

Especial atención requiere el uso de tests psicológicos en un idioma que no coincida con la lengua materna del estudiante, aún en el caso de que tengan cierto dominio del idioma o sean supuestamente bilingües, algo relativamente frecuente en el ámbito de la traductología. El perfil de rendimiento y capacidades que se obtenga puede verse afectado negativamente, por lo que, según sea el objetivo de investigación, cabría desechar estas puntuaciones o utilizar una prueba alternativa.

Asimismo, algunos tests requieren que la persona que los administre o interprete posea cierta cualificación profesional, formación o experiencia, y así lo hacen constar en sus manuales de uso. Existe una clasificación de los instrumentos en tres niveles, utilizada por los editores de las pruebas para su comercialización: los que requieren únicamente formación y experiencia en el ámbito concreto de aplicación, los que necesitan de ciertos conocimientos sobre la teoría de los tests y métodos estadísticos y los que requieren poseer una titulación superior en psicología, psiquiatría o psicopedagogía y experiencia profesional (Muñiz Fernández 1997: 307-345). En estos casos, se recomienda respetar dicha norma y,

en general, siempre que se considere necesario, se debe solicitar el consejo de un profesional.

Por último, es necesario informar a los sujetos evaluados sobre el objetivo del test, su forma de administración, los factores que se tienen en cuenta a la hora de puntuar sus respuestas, en qué forma se utilizarán dichas puntuaciones, durante cuánto tiempo se guardarán los resultados y en qué condiciones se van a dar a conocer.

REFERENCIAS BIBLIOGRÁFICAS

- AMERICAN EDUCATIONAL RESEARCH ASSOCIATION, AMERICAN PSYCHOLOGICAL ASSOCIATION & NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION [1999] 2008. *Standards for Educational and Psychological Testing*. Washington: American Educational Research Association.
- ALLPORT, G. W. 1937. *Personality: A psychological interpretation*. New York: Holt & Co.
- BAESSLER, J. & SCHWARCER, R. 1996. «Evaluación de la autoeficacia: Adaptación española de la escala de autoeficacia general». En *Ansiedad y estrés*. Vol. 2, pp. 1-8.
- BANDURA, A. 1977. «Self-efficacy: Toward a unifying theory of behavioural change». En *Psychological Review*. Vol 84, pp. 191-215.
- BANDURA, A. 2000. «Exercise of human agency through collective efficacy». En *Current Directions in Psychological Science*. Vol. 9, pp. 75-78.
- BARBERO GARCÍA, M. I. 2006. «La validez de las inferencias». En BARBERO GARCÍA, M. I., VILA ABAD, E. & SUÁREZ FALCÓN, J. Ca. (eds.) *Psicometría*. Madrid: Universidad Nacional de Educación a Distancia.
- BARBERO GARCÍA, M. I.; VILA ABAD, E.; SUÁREZ FALCÓN, J. C. 2006. *Psicometría*. Madrid: Universidad Nacional de Educación a Distancia.
- BOLAÑOS MEDINA, A. 2008. «Cohesión y sentimiento de eficacia colectiva en los grupos virtuales de aprendizaje colaborativo. Implicaciones didácticas». En *I Jornadas Internacionales de Teleformación*. Las Palmas de Gran Canaria: Universidad de Las Palmas de Gran Canaria.
- CARRON, A. V.; WIDMEYER, W. & BRAWLEY, L. 1985. «The development of an instrument to assess cohesion in sport teams: The Group Environment Questionnaire». En *Journal of Sport Psychology*. Vol. 7, pp 244-267.
- DAM-JENSEN, H. & HEINE, C. 2009. «Process Research Methods and Their Application in the Didactics of Text Production and Translation». En *Trans-kom*. Vol. 2, pp. 1-25. Artículo disponible en red: [http://www.trans-kom.eu/bd02nr01/trans-kom_02_01_01_Dam-Jensen_Heine_Process_Research.20090721.pdf] (Consulta 15/09/09).
- FERNÁNDEZ JIMÉNEZ, E. 2003. «Enfoque integral de las diferencias individuales en la inteligencia y personalidad». En SANCHEZ-ELVIRA PANIAGUA, A. (ed.). *Introducción al estudio de las diferencias individuales*. Madrid: Sanz y Torres.
- FERNÁNDEZ SÁNCHEZ, M. M. & MUÑOZ MARTÍN, R. 2008. *Aproximaciones cognitivas al estudio de la traducción y la interpretación*. Granada: Comares.
- FORNS I SANTACANA, M.; ABAD GIL, J.; AMADOR CAMPOS, J. A.; KIRCHNER NEBOT, T.; ROIG LÓPEZ, F. 2002. *Avaluació psicològica*. Barcelona: UOC.

- FREESTON, M. H.; RHÉAUME, J.; LETARTE, H.; DUGAS, M. J. & LADOUCEUR, R. 1994. «Why do people worry?». En *Personality and individual differences*. Vol. 17, pp. 791-802.
- GOLDEN, C. J. 1975. «A group version of the Stroop color and Word test». En *Journal of Personality Assessment*. Vol. 39, pp. 386-388.
- GONZÁLEZ RODRÍGUEZ, M.; CUBAS LEÓN, R.; ROVELLA, A. T.; DARIAS HERRERA, M. 2006. *Adaptación española de la escala de intolerancia a la incertidumbre: procesos cognitivos, ansiedad y depresión*. En *Psicología y Salud*. Vol.16, pp. 219-233.
- GÖPFERICH, S.; JAKOBSEN, A. L. & MEES, I. M. 2008. *Looking at Eyes: Eye-Tracking Studies of Reading and Translation Processing*. Copenhagen: Samfundslitteratur Press.
- HALVERSON, S. 2009. «Psycholinguistic and cognitive approaches». En BAKER, M. & SALDANHA, G. *Routledge Encyclopedia of Translation Studies*. New York: Routledge.
- HANSEN, G 2003. «Controlling the process. Theoretical and methodological reflections on research into translation processes». En ALVES, F. (ed.) *Triangulating Translation*. Amsterdam/Philadelphia: John Benjamins Publishing.
- HUICI, C. & MORALES, J. F. 2004. *Psicología de Grupos. Estructura y Procesos*. Madrid: UNED.
- JÄÄSKELÄINEN, R. 2000. «Focus on Methodology in Think-aloud Studies on Translating». En TIRKKONEN-CONDIT, S. & JÄÄSKELÄINEN, R. (eds.) *Tapping and Mapping the Processes of Translation and Interpreting*. Amsterdam/Philadelphia: John Benjamins Publishing.
- KAGAN, J. 1965. *Matching Familiar Figures Test*. Cambridge: Harvard University.
- KAUFMAN, A.; LICHTENBERGER, E. O. 1999. *Claves para la evaluación con el WAIS-III*. Madrid: TEA Ediciones.
- LITTLE, B. L.; MADIGAN, R. M. 1997. «The relationship between collective efficacy and performance in manufacturing work teams». En *Small Group Research*. Vol. 28, pp. 517-534.
- MARÍN GARCÍA, Á. 2005. *Creatividad y automatización al traducir*. Proyecto de investigación dirigido por el Dr. Muñoz Martín. Inédito. Universidad de Granada.
- MORENO ROSSET, C. (ed.) 2005. *Evaluación psicológica. Concepto, proceso y aplicación en las áreas de desarrollo y de la inteligencia*. Madrid: Sanz y Torres.
- MUÑIZ FERNÁNDEZ, J. 1997. «Aspectos éticos y deontológicos de la evaluación psicológica». En CORDERO, A. (coord.) *La evaluación psicológica en el año 2000*. Madrid: Tea ediciones.
- MUÑOZ MARTÍN, R. [en prensa]. «Subject profiling in translation process research». En MEEZ, I.; GÖPFERICH, S.; ALVES, F. [Título no disponible]. Copenhagen: Samfundslitteratur.
- PRESAS, M. 2006. «El estudio de la competencia traductora: de la observación empírica a las tareas experimentales». *XII Congreso Latinoamericano de Estudios Germanísticos*. La Habana. Artículo disponible en red: [http://www.fti.uab.es/pacte/Ultimo_congreso/Cuba/Pres_Cuba.pdf] (Consulta 2/09/09).

- SALANOVA, M.; GRAU, R.; MARTÍNEZ, I.; CIFRE, E.; LLORENS, S. & GARCÍA RENEDO, M. 2004. *Nuevos horizontes en la investigación sobre autoeficacia*, Castellón: Universitat Jaume I.
- SANCHEZ-ELVIRA PANIAGUA, Á. (ed.) 2003. *Introducción al estudio de las diferencias individuales*. Madrid: Sanz y Torres.
- SANJUÁN SUÁREZ, P.; PÉREZ GARCÍA, A. M. & BERMÚDEZ MORENO, J. 2000. «Escala de autoeficacia general: datos psicométricos de la adaptación para la población española». *Psicothema*. Vol. 12, pp 509-513.
- TIRKKONEN-CONDIT, S. 2000. «Uncertainty in Translation Processes». En TIRKKONEN-CONDIT, S. & JÄÄSKELÄINEN, R. (eds.) *Tapping and Mapping the Processes of Translation and Interpreting*. Amsterdam/Philadelphia: John Benjamins Publishing.
- WITKING, H.; OLTMAN, P.; RASKIN, E. & KARP, S. 1971. *Manual of the Embedded Figures Test*. Palo Alto: Consulting Psychologists Press.