

Stuart Campbell Sandra Hale

# Translation and Interpreting Assessment in the Context of Educational Measurement

*Davide De Giorgi*

**FONDAZIONE SCUOLE CIVICHE MILANO  
Scuola Superiore per Mediatori Linguistici  
via Alex Visconti 18, 20151 MILANO**

Relatore: Prof. Bruno Osimo  
Correlatrice: Prof.ssa Elena Berlot

Diploma in Scienze della Mediazione Linguistica  
24 Novembre 2003

## Sommario

Abstract (English).....pag. V

Abstract (français).....pag. VII

Prefazione.....pag. IX

Problematica Traduttiva.....pag. XIV

Traduzione con testo a fronte.....pag. 1

Bibliografia.....pag. 47



## **Abstract (English)**

The candidate has carried out the translation from English into Italian of a recent contribution by Stuart Campbell and Sandra Hale published in the book «Translation Today: Trends and Perspectives». The contribution aims at providing a global analysis of a number of works, published over the last decades, concerned with assessment procedures in educational contexts (including accreditation procedures). The analysis shows that testing procedures designed and adopted by scholars or academic/non-academic institutions may range widely and that the numerous interpreting/translation competences required are generally quite well verified; the most serious gap in today's interpreting and translation assessment is to be found in the lack of validity and reliability in the current testing procedures. And as new modes of translation emerge, the need for clearly formulated and uniformly accepted – and therefore more reliable – methods of assessment of translation and interpretation competence becomes greater.

In the Preface to the work the candidate examines the topic in detail, closely analyses the discussions and trends in language testing found in the contribution and expresses his views on the subject. The Preface is completed by a further section entitled «Problemativa traduttiva» in which the candidate presents a comprehensive translation-oriented analysis of his work: he comments on his choice of words, expressions and explains how he has solved the main translation problems encountered. This section includes a discussion focusing on the main translational aspects: the candidate describes the translated text as an open text and highlights the importance of connotation/connotative meaning of words – as opposed to denotation/denotative meaning – and of intertextuality (e.g. quotations) in defining the degree of openness of the text. In close relation to this topic, he draws the discussion on polysemy and on how the connotative meaning of words is essentially determined by the context and co-text. The candidate analyses the two key elements of his «translation strategy» – the «dominant» and the «model reader» – establishing a comparison with the «narrative strategy»

adopted by the authors of the source text. He then explains, by means of examples, how choosing a particular «translation strategy» can affect or even radically alter the way a message or a whole text is received and decoded by the «empirical readers».

As a postscript, the candidate deals with the issue of «communication loss» and explains the difficulty involved in reaching an acceptable compromise between a «linguistic» and a «cultural» translation.

## Abstract (français)

Le candidat a réalisé la traduction de l'anglais en l'italien d'une récente étude de Stuart Campbell et Sandra Hale publiée dans le livre «Translation Today: Trends and Perspectives». Cette étude contient une analyse générale d'une série d'ouvrages, publiés ces dernières décennies, qui ont pour objet les procédures et les méthodes d'évaluation adoptées dans des contextes académiques (y compris les procédures d'accréditation). De cette analyse il résulte que les tests (d'aptitude par exemple) conçus et choisis par les experts ainsi que par les institutions universitaires et non-universitaires, peuvent être très variés; par ailleurs les nombreuses compétences en interprétation/traduction sont souvent bien cernées. Cependant, les méthodes d'évaluation ne sont pas suffisamment standardisées et manquent souvent de fiabilité. Avec l'apparition de nouvelles spécialisations en traduction/interprétation, une nouvelle nécessité se fait jour, celle de disposer de méthodes d'évaluation des compétences plus claires et qui soient largement acceptées et utilisées – et de ce fait plus fiables.

Dans la préface de sa traduction, le candidat donne un aperçu détaillé des sujets abordés par les auteurs et du débat ouvert en matière d'évaluation des compétences linguistiques et il exprime son point de vue à cet égard. La préface est complétée par une section intitulée «Problematica Traduttiva» dans laquelle le candidat présente une analyse traductologique de son travail, commentant le choix qu'il a fait de certains termes et justifiant certaines expressions et les solutions adoptées pour les principaux problèmes de traduction rencontrés dans le texte. Cette section inclut un examen des aspects traductologiques fondamentaux: le candidat aborde la question de l'«ouverture sémantique» du texte et explique comment celle-ci évolue en fonction de la valeur connotative des mots – par opposition avec la dénotation – et de la présence de références intertextuelles (par exemple les citations). Ce sujet se rattache au thème que le candidat développe sur la polysémie des mots et sur l'importance que le «contexte» et le «cotexte» (le texte autour d'un énoncé) revêtent afin d'aboutir au sens «connotatif» des mots. Ensuite il analyse les deux éléments clefs de sa «stratégie

traductive» – soit la recherche de la «fonction dominante» et l'identification du «lecteur modèle» – en établissant une comparaison avec la «stratégie narrative» adoptée par les auteurs du texte de départ. En s'aidant d'exemples, le candidat explique dans quelle mesure le choix d'une «stratégie traductive» peut affecter, voire altérer radicalement la façon dont un message ou un texte entier est reçu et décodé par les «lecteurs empiriques». En dernier lieu le candidat aborde le problème de la «perte de communication» et il fait part de son espoir d'être parvenu à un équilibre acceptable entre traduction «linguistique» et traduction «culturelle».

## Prefazione

Il tema della valutazione della traduzione e dell'interpretazione in ambito accademico è da molto tempo al centro di studi, ricerche, dibattiti che coinvolgono studiosi e docenti di tutto il mondo. Al fine di comprendere il motivo di tanto interesse per questo tema è necessario tenere presente che maggiore conoscenza e padronanza delle principali lingue in uso nel mondo e maggiore duttilità nell'utilizzo delle stesse sono capacità divenute, negli ultimi decenni, sempre più richieste in molti ambiti lavorativi, primo fra tutti quello della traduzione e dell'interpretazione. In realtà, parlare oggi di traduzione e interpretazione *tout court* sarebbe piuttosto limitativo, poiché la tendenza generale nel mondo del lavoro è richiedere competenze e conoscenze sempre più specialistiche e ad alto livello. Basti pensare che già da tempo esistono figure professionali di interpreti/traduttori che operano in ambiti specifici quali il settore giuridico/giudiziario e medico/sanitario, l'assistenza sociale e di comunità, la salute mentale, la localizzazione di software, la pubblicità etc. Si tratta di attività di mediazione linguistica specializzate destinate a divenire sempre più diffuse e richieste nella prospettiva di un mondo sempre più «globalizzato», multietnico e multiculturale in cui il problema principale sarà fondamentalmente la difficoltà di comunicazione. Sempre più spesso, quindi, essere dei buoni traduttori o interpreti può non bastare; sempre più spesso si richiedono conoscenze che vanno al di là della semplice padronanza della lingua. Negli ultimi decenni sono fiorite, in tutto il mondo, moltissime scuole per interpreti e traduttori e oramai molte università offrono corsi di traduzione e interpretazione ai quali è possibile accedere con un diploma e il superamento di test di ammissione. La necessità di sfornare professionisti in mediazione linguistica qualitativamente migliori ha portato gli esperti del settore (linguisti, docenti, valutatori, ideatori dei test) a porre in primo

piano e a riesaminare alcune questioni fondamentali quali la formazione di interpreti e traduttori, l'attendibilità e l'utilità dei più disparati modelli di test (dai test attitudinali e di profitto agli esami di accreditamento e di trasferimento di crediti), le modalità di valutazione della prestazione in traduzione e interpretazione e, non ultimo, l'attendibilità e la coerenza valutativa degli esaminatori. Il saggio di Stuart Campbell e Sandra Hale si propone infatti di comprendere e di farci comprendere quale sia lo stato dell'arte degli studi condotti e delle opere pubblicate fino ad oggi riguardanti tali questioni. In particolar modo grande attenzione è riservata al tema della valutazione in traduzione e in interpretazione. La valutazione della prestazione rappresenta infatti un anello fondamentale della catena formativa in traduzione e in interpretazione; se da un lato la preparazione alla professione di mediatore linguistico deve essere costantemente aggiornata per restare al passo con l'evolversi stesso della professione e del mondo del lavoro, dall'altro è necessario adottare metodi di valutazione più precisi e obiettivi possibili, che forniscano dunque gli stessi risultati o comunque risultati confrontabili da valutatore a valutatore, da candidato a candidato e nel corso del tempo. In realtà oggi questo non è ancora possibile in quanto ogni istituzione, ogni università adotta modelli di test, metodi di valutazione e approcci valutativi differenti (basti pensare alla differenza fondamentale tra i test basati sulla norma, dove i risultati del singolo vengono rapportati a quelli di altri individui dello stesso gruppo e i test basati su criteri, in cui le competenze del singolo sono rapportate a criteri prefissati). Ovviamente bisogna ricordarsi che qualsiasi test, che si prefigga di «misurare» la competenza linguistica di un individuo, non potrà mai essere uno strumento totalmente esatto e preciso, in quanto rappresenta un campione di comportamento che non può prendere in considerazione tutte le variabili umane e psicologiche (attitudine all'apprendimento, diverso approccio personale, motivazione e atteggiamento psicologico ecc.). I vari autori, presi in rassegna nello studio, trattano di valutazione della traduzione e dell'interpretazione in maniera diversa e, spesso, concentrano la loro attenzione solamente su alcuni tipi di test o su alcuni aspetti dei test (ad esempio lo scopo o le competenze da valutare) mentre rimangono sul vago o addirittura non affrontano altre questioni altrettanto fondamentali ai fini di una valutazione attendibile. In altre parole non esistono dei

criteri di valutazione oggettivi prestabiliti e universalmente accettati e applicati e, quindi, nell'ideazione dei test e dei metodi valutativi (soprattutto in interpretazione) molto spazio è lasciato alla soggettività, all'intuito degli ideatori dei test e dei docenti. Nell'ambito della traduzione si parla molto di valutazione finalizzata all'accreditamento (es. presso le Nazioni Unite) o alla didattica, mentre raramente si discute dei test attitudinali per traduttori. Molti autori addirittura parlano di valutazione della traduzione senza fare alcun riferimento allo scopo. I tipi di competenze richieste in traduzione variano enormemente, dato che in pratica ogni autore elabora un proprio schema o una tabella delle competenze (principalmente linguistiche e tecniche) o dei tipi di errore nella lingua di arrivo. A eccezione di un autore, la traduzione di un testo sembra essere la modalità standard per la valutazione delle capacità traduttive. Resta quasi del tutto insoluto il problema della lunghezza ideale dei test di traduzione o del tempo massimo consentito per il loro svolgimento: ogni istituzione (es. NAATI, l'ente australiano deputato all'accreditamento di interpreti e traduttori) adotta le proprie procedure e ogni autore propone modelli di test differenti. Per quanto riguarda gli approcci base degli strumenti dei test sembra esserci una netta preferenza per lo schema di valutazione *upside-down*, cioè capovolto, nel quale da un punteggio pieno si detraggono i punti degli errori. In realtà, si tratta di un metodo che ben poco si adatta alla valutazione di una traduzione se si considera che questa può essere fatta in infiniti modi e che esistono molti possibili errori; il numero di punti massimo è arbitrario e non è in alcun modo collegato al possibile numero di errori. Da un punto di vista teorico un individuo potrebbe totalizzare un numero così elevato di errori da ottenere un punteggio di valutazione inferiore allo zero (es. in una scala di valutazione da 0 a 100 punti); è dunque impossibile fissare un minimo (nel nostro esempio 0) alla scala di valutazione e perciò valutare con precisione e attendibilità la prestazione di tutti i candidati.

Sulla valutazione dell'interpretazione si è pubblicato ancora meno. Quasi tutta la letteratura esistente è dedicata ai test attitudinali di accesso ai corsi di interpretariato di conferenza, agli esami di accreditamento per esercitare la professione di interprete e alla valutazione qualitativa di interpreti professionisti, soprattutto di conferenza. I test attitudinali e i test d'ingresso mirano più o meno

tutti a saggiare le stesse capacità e competenze e seguono, in genere, il modello basato su criteri in cui i candidati sono chiamati a soddisfare tutti i criteri prestabiliti. Anche le forme dei test sono più o meno comuni (es. *shadowing*, traduzione a vista, ecc.). Mancano però dei criteri valutativi oggettivi e dunque risulta oltremodo difficile valutare l'attendibilità di tali test.

Un'altra lacuna della letteratura presa in esame può essere individuata nello scarso approfondimento di questioni relative alla formazione dei valutatori, alle scale di valutazione, al punteggiamento e alle procedure di equiparazione dei test.

Esiste un esiguo numero di test che si occupano di esami di accreditamento o certificazione per ottenere una qualifica professionale (ad esempio gli esami dell'australiana NAATI); tuttavia anche le linee guida, fissate dalla NAATI per la valutazione della prestazione interpretativa, lasciano ampia discrezionalità valutativa agli esaminatori a discapito di una maggiore obiettività e attendibilità dei risultati. Nonostante la NAATI si servi di esaminatori qualificati e cerchi di aggiornare i suoi esami, siamo ancora lontani da un'analisi attenta e sistematica della validità e dell'attendibilità dei test.

In generale si può affermare che vi sia un certo accordo sui gruppi di competenze da valutare, sia in traduzione sia in interpretazione, ma quasi mai si discute dell'efficacia di un particolare strumento di valutazione nel giudicare tali competenze. Se da un canto si parla in maniera abbastanza diffusa di scopi, competenze e forma dello strumento di valutazione, quasi mai l'approccio di base (basato sulla norma o su criteri) è trattato in maniera esplicita, nonostante vi sia una tacita preferenza per l'approccio basato su criteri. Poca attenzione è inoltre riservata al tipo di risultati, all'utilità di uno strumento basato sulla norma nel differenziare i candidati e ai meccanismi di feedback.

Gli autori della ricerca concludono la loro analisi affrontando la mancanza più evidente in tema di misurazione e valutazione accademica, cioè l'attendibilità. Bachmann, citata più volte dagli autori dell'articolo, ribadisce la necessità di adottare criteri più chiari per una resa corretta. In effetti, spesso, i giovani studenti di traduzione e di interpretazione si trovano a dover reagire in maniera istintiva al momento di dover scegliere tra una resa più fedele o una meno fedele al messaggio originale ma più vicina e comprensibile per il ricevente; per quale

soluzione bisogna optare? Per di più, se il candidato compie un ipotetico «atto di coraggio» scegliendo una soluzione ardita e meno fedele all'originale, verrà premiato o penalizzato? L'assenza di indicazioni chiare ed esplicite in merito alla fedeltà può influire sull'attendibilità del test. Anche il grado di rapidità è determinante; di conseguenza il tempo massimo, per lo svolgimento di un test, dovrebbe essere fissato non in base a criteri puramente arbitrari bensì solo dopo aver stabilito se esiste una velocità alla quale sia possibile ottenere, dalla maggior parte dei candidati, una prestazione ottimale e quindi più vicina alle loro reali capacità. Un esempio pratico che mette in luce la carenza di attendibilità dei test è offerto dalle ripetizioni. Se un candidato sbaglia la traduzione di un elemento ripetuto (per questioni formali o stilistiche) va penalizzato una o più volte? La ripetizione potrebbe essere vista come un elemento che riduce il grado discriminatorio del test, compromettendo l'attendibilità del test stesso; d'altro canto la presenza di ripetizioni, in alcuni casi, potrebbe fungere da stimolo alla ricerca di soluzioni creative che possano rappresentare un valore aggiunto alla prestazione del candidato. Ma non sono solamente i test ad essere oggetto di studio sull'attendibilità; spesso si parla infatti di attendibilità dei singoli esaminatori e di attendibilità delle valutazioni tra più esaminatori. Alcune istituzioni, come la NAATI e l'American Translator Association (ATA), sembrano dare molto peso al comportamento professionale degli esaminatori al fine di raggiungere un maggior grado di attendibilità. In ogni caso, in nessuna delle pubblicazioni prese in esame compare uno studio serio sul tema dell'attendibilità e della coerenza dei valutatori. In conclusione, dalla ricerca effettuata da Stuart Campbell e Sandra Hale, si può facilmente comprendere quanto lavoro vi sia ancora da compiere nel campo della valutazione e quanta strada bisogna ancora percorrere per arrivare a test estremamente attendibili. Impiegare test più attendibili e obiettivi significherebbe essenzialmente saper valutare e formare meglio le nuove schiere di giovani che si apprestano ad entrare nel mondo della traduzione o dell'interpretariato e che, forse, un giorno, saranno chiamati a loro volta a dover valutare e formare nuovi professionisti della mediazione linguistica.

## Problematica traduttiva

La traduzione del testo originale ha presentato una serie di problemi traduttivi. La prima e principale questione riguarda la traduzione in italiano del termine inglese «interpreting»; come molte volte accade nella lingua inglese, questa parola ammette in italiano una serie di possibili traduttori diversi. In funzione di sostantivo può essere tradotta come «interpretariato» o «interpretazione»; in funzione di aggettivo si traduce con «interpretativo». Gli autori del saggio l'hanno impiegata per tradurre indistintamente «interpretariato», «interpretazione» e «interpretativo». Solamente in due sporadici casi si ricorre alla parola inglese «interpretation»: nel primo caso si parla di «...Graduate Institute of Translation and *Interpretation* Studies...», nel secondo di «judicial *interpretation*». Soprattutto in quest'ultimo caso non risulta ben chiaro il motivo per cui gli estensori abbiano optato per questa variante lessicale dato che nello stesso paragrafo si fa riferimento alla stessa attività con l'espressione «judicial *interpreting*» e «court *interpreting*». Probabilmente si tratta di una scelta dettata da motivi estetici, stilistici, quindi per evitare di ripetere troppe volte la stessa parola.

In italiano il discorso sembra più complesso e meno chiaro. I termini «interpretariato» e «interpretazione» possono significare la stessa cosa, ovvero l'attività, la funzione svolta dall'interprete; tuttavia nella mia traduzione mi sono servito di entrambi i traduttori. In realtà, la scelta di una parola piuttosto che l'altra è stata a volte intuitiva ed è dipesa in gran parte dal contesto in cui la stessa è collocata: se da un canto parlo di «valutazione dell'*interpretazione*» (intesa come «prestazione interpretativa») dall'altro parlo di «corso di *interpretariato*» o di «*interpretariato* di conferenza» o di «*interpretariato* di comunità o in campo medico/sanitario e giuridico/giudiziario». (Tuttavia al punto

n. 8 della pagina web

[www.club.it/culture/culture2001/giuliana.garzone/note.garzone.html](http://www.club.it/culture/culture2001/giuliana.garzone/note.garzone.html) si parla di «*interpretazione* di comunità nel campo medico e sanitario e giuridico/giudiziario»). Ho preferito utilizzare il vocabolo «interpretazione» in riferimento all'aspetto della prestazione interpretativa in sé o alla generica traduzione orale di un messaggio, e il termine «interpretariato» in riferimento all'attività dell'interprete vera e propria (soprattutto se si specifica il tipo di interpretariato: ad esempio di trattativa, di conferenza ecc...). Si tratta, come già detto, di una scelta da un lato arbitraria e intuitiva e dall'altro dettata dal fatto che «interpretazione» è una parola generica ampiamente usata anche in psicologia, ermeneutica, critica d'arte, semiotica ecc., mentre «interpretariato» è un termine settoriale univoco.

Una ricerca condotta con il motore di ricerca Google ha comunque confermato una tendenza ad usare indistintamente i vocaboli «interpretazione» e «interpretariato»: si parla infatti ad esempio di «corsi di laurea in traduzione e *interpretazione*» come pure, con maggior precisione terminologica, di «corsi di laurea in traduzione e *interpretariato*».

Restando nell'ambito dell'interpretariato, Roseann Dueñas Gonzáles, un'autrice citata dagli estensori del saggio, si serve dell'espressione «...simultaneous (unseen or spontaneous)...translation...» in riferimento all'interpretariato giuridico. Si tratta di espressioni specifiche relative al settore dell'interpretariato giuridico e, con tutta probabilità, connotate geograficamente (in uso negli Stati Uniti ma poco note in Europa). Per tale motivo trovare una traduzione accettabile, in un primo momento, ha comportato notevoli difficoltà. In effetti è stata proprio Sandra Hale, uno degli autori, a fornirmi la soluzione al problema. In genere si parla di «unseen simultaneous translation» allorché l'interprete giurato, nel corso di un processo, traduce con il metodo dello «chuchotage», cioè bisbigliando la traduzione (quindi senza l'ausilio di apparecchiature audiovisive), alla parte in causa non inglese – o comunque di lingua diversa da quella del processo – da un luogo o postazione non visibile al pubblico dell'aula (per questioni di sicurezza o privacy). Nell'interpretazione consecutiva l'interprete si trova invece a fianco al testimone di fronte alla corte al momento della deposizione. L'espressione «spontaneous» sta semplicemente a significare che l'interprete deve saper tradurre velocemente

senza avere tempo per riflettere o prendere appunti; probabilmente l'autrice fa riferimento a quel tipo di interpretariato che normalmente, nel contesto europeo, prende il nome di «instant translation» o «liaison interpreting» ovvero una interpretazione «frase per frase» o «di collegamento» o «di trattativa». Anche quest'ultima forma di interpretariato non prevede l'utilizzo di attrezzature audiovisive o la possibilità di prendere appunti dato che l'interprete è fisicamente vicino ai suoi uditori e spesso si pone tra le parti che beneficiano della traduzione. Da notare inoltre che, normalmente, la «unseen translation» sta a indicare ciò che noi chiamiamo «traduzione a vista».

Il vocabolo inglese «testing» ha posto qualche problema di traduzione. In alcuni casi (es. «aptitude testing») ho optato per il traduttore «test»; in altre situazioni ho preferito una resa differente a seconda del contesto, ad esempio: «modalità di valutazione», «procedure di verifica» (per «testing procedures»), «valutazione linguistica» o «modalità di valutazione del linguaggio» (per «language testing»), «test di verifica» o «prove».

Nel quadro della valutazione, qualche difficoltà ha comportato la scelta del giusto traduttore per i vocaboli «rater», «marker», «grader» ed «examiner»; per una questione di coerenza e maggiore semplicità ho preferito servirmi del traduttore «esaminatore» per la resa di «examiner» e del traduttore «valutatore» nei restanti casi. In realtà tra i due vocaboli esiste una certa differenza: la parola «esaminatore» ha un più ristretto campo di utilizzo, poiché in genere è applicabile solamente ad un contesto d'esame/concorso; la «valutazione» invece può avvenire anche al di fuori di un esame. È pur vero che, generalmente, chi svolge il ruolo di esaminatore debba anche emettere un giudizio, una valutazione; in questo caso, il significato e l'impiego dei due traduttori tenderanno a sovrapporsi.

In linea di massima ho usato lo stesso traduttore, cioè «valutazione», per tradurre sia «assessment» sia «evaluation».

Per quanto riguarda la resa del verbo «to measure» ho impiegato, a seconda dei casi e del contesto linguistico, diversi traduttori: i più frequenti sono «giudicare», «misurare», «valutare», «calcolare».

La traduzione dei termini «norm-referenced test» e «criterion-referenced test» ha comportato qualche problema. Da una veloce ricerca condotta con Google sui siti

italiani è emerso che si tende ad usare le stesse espressioni inglesi. Per una maggiore chiarezza io ho preferito invece trovare traduttori italiani che fossero comprensibili e avessero lo stesso significato, ovvero «test basato sulla norma» e «test basato su criteri».

Per la traduzione del termine «performance» ho scartato quasi subito l'idea di lasciarlo in inglese e ho optato, nella maggior parte dei casi, per il traduttore italiano «prestazione» e talvolta per il traduttore «rendimento», anche se il primo mi è sembrato più adatto al contesto della valutazione delle capacità traduttive.

Inizialmente ho avuto dubbi anche sulla resa dell'onnipresente concetto della «reliability» nell'ambito della valutazione. I possibili traduttori erano «attendibilità» e «affidabilità». In ultima analisi ho scelto il traduttore «attendibilità» poiché, a mio parere, si addice di più a un metodo di valutazione; invece, in genere si tende a parlare di «affidabilità» in riferimento a una persona o a un mezzo meccanico.

Anche il vocabolo inglese «scholarship», in realtà poco frequente nell'originale, ha inizialmente posto qualche problema di resa. Si tratta di una parola che abbraccia tutta una serie di traduttori differenti, ovvero ha un campo semantico denotativo abbastanza ampio: infatti, consultando il dizionario bilingue inglese-italiano, alla voce «scholarship» ho riscontrato i seguenti traduttori italiani:

borsa di studio (come primo significato)

cultura, erudizione, sapere, dottrina (come secondo significato)

studiosi (pl) (come ultima scelta)

Mi sono dunque trovato di fronte alla necessità di operare una scelta lessicale importante. Si tratta, in altre parole, di un caso, del resto abbastanza frequente, in cui l'intervento interpretativo del traduttore gioca un ruolo considerevole nel mantenimento di una coerenza lessicale. Io ho optato per i traduttori «conoscenze», «dottrina» e «studi» poiché «scholarship» è usato in riferimento al campo di conoscenze e alla ricerca alla base della formazione di interpreti e traduttori.

Di un certo interesse anche la traduzione del termine «sub-skills»; nei siti internet italiani non compare quasi mai il termine inglese e risultano poco diffusi anche i traduttori «sub-capacità» e «sub-abilità»; si parla più comunemente di

«sottocapacità» o «sub-competenze» e soprattutto di «sottoabilità» e «sottocompetenze». Quest'ultimo risulta il più comune ed è quello che ho deciso di adottare nella traduzione.

La traduzione del termine inglese «sub-test» è stata più agevole; infatti nei siti internet italiani quella più diffusa è «subtest» o «sub-test» ma ho riscontrato anche l'uso – meno frequente però – del termine «sottotest». Anche in questo frangente ho adottato la versione più comunemente accettata, ovvero «subtest».

La traduzione non ha presentato ulteriori problemi traduttivi degni di nota. L'esistenza di uno studio della problematica traduttiva, di una semiosi del testo è comunque indicativo dell'apertura del testo stesso. Tale apertura è dovuta sia alla presenza di una certa connotazione, sia ai continui riferimenti intertestuali. Il testo è infatti ricco di rimandi e citazioni che, comunque, sono spesso ben identificabili (esplicitezza) grazie all'uso di delimitatori grafici (virgolette) mentre le fonti sono sempre citate e specificate (esplicitezza della fonte); anche il motivo delle citazioni (esplicitezza della funzione) mi è sembrato spesso facilmente comprensibile. In ogni caso la decodifica delle citazioni è risultata a volte problematica per me a causa della disconoscenza sia degli autori citati che delle loro teorie; in una certa misura è un problema che potrebbe aver riguardato anche i lettori della cultura emittente e che potrebbe riguardare parte dei lettori della cultura ricevente.

Leggendo il testo è facile imbattersi in vocaboli polisemici, cioè che ammettono molteplici interpretazioni e possono avere uno spettro semantico più o meno esteso; per tale ragione essi sono comprensibili solamente se teniamo conto del contesto e del co-testo in cui si trovano; casi emblematici sono, ad esempio, le parole inglesi «portrayal» o «stakeholders». La prima, da un punto di vista denotativo, ammette tre significati: «raffigurazione», «presentazione» e «ritratto». Nel testo sarebbe stato tuttavia difficile stabilire quale significato attribuire alla parola, se non avessi considerato con attenzione il suo contesto e in modo particolare il suo co-testo; infatti è stato proprio grazie allo studio del contesto linguistico della parola che sono stato in grado di comprenderne il significato connotativo, ovvero «(qualità) della resa linguistica». Lo stesso ragionamento vale per il vocabolo «stakeholders»; esso infatti normalmente indica l'«azionista», il

«partecipante» ma è risultato subito ovvio che il termine non andava preso nel suo senso strettamente letterale ma in senso figurato, ovvero «soggetti coinvolti». Anche in questa situazione il contesto ed il co-testo sono stati fondamentali al fine della comprensione.

In fase di stesura del testo tradotto, una delle mie maggiori preoccupazioni è stata la gestione del «residuo comunicativo», che inevitabilmente accompagna ogni traduzione e più in generale ogni atto comunicativo. Il residuo comunicativo consiste essenzialmente nella progressiva perdita del contenuto o del significato del messaggio originale attraverso i vari passaggi di decodifica in materiale psichico e di ricodifica, e quindi di riverbalizzazione dello stesso messaggio nel codice ricevente. A questo proposito ha giocato un ruolo rilevante la scelta della «strategia traduttiva». Per poterla elaborare ho dovuto procedere all'individuazione della «strategia narrativa» degli autori dell'originale – la «dominante» e il «lettore modello» del prototesto (primo grado della ricostruzione abduzione secondo Peirce). Trattandosi di un testo saggistico la funzione dominante che ho individuato è di carattere informativo (cosa che si evince in maniera inequivocabile già dal titolo): gli autori si propongono innanzitutto di divulgare e commentare i risultati della loro indagine. È comunque possibile individuare marche a livello lessicale, dovute essenzialmente al diffuso impiego di termini ed espressioni settoriali proprie del vocabolario tecnico della misurazione accademica e della interpretazione / traduzione. Per di più nel caso specifico dei termini «unseen» e «spontaneous» (di cui sopra) è presente un'ulteriore marca di carattere geografico; sembrerebbe trattarsi infatti di espressioni proprie dell'intepretariato giuridico statunitense e per tale motivo non condivise da molti interpreti italiani e probabilmente europei. Qui sono dovuto ricorrere a espressioni forse semanticamente non del tutto coincidenti con quelle originarie («instant translation» come analogo culturale di «spontaneous translation») ma che risultano per lo meno comprensibili ai lettori del metatesto. Volendo invece prendere in prestito la terminologia dello scienziato della traduzione Toury, potrei dire di aver sacrificato, in questo frangente, l'«adeguatezza» della mia traduzione a vantaggio dell'«accettabilità». Esiste inoltre un altro caso in cui ho optato per una «traduzione culturale» e «accettabile»: nella fattispecie, ho deciso di tradurre

l'espressione inglese «postgraduate...course» ricorrendo al suo equivalente nella cultura italiana del dopo riforma dei cicli di studio, ovvero «corso di laurea di secondo livello». Si tratta di una palese manipolazione del testo originale che sortisce l'effetto di annullare la distanza cronotopica tra prototesto e lettore del metatesto, avvicinando, pertanto, il prototesto alla metacultura. I miei lettori non dovranno compiere alcuno sforzo per comprendere ciò di cui parlo, molti di loro saranno ignari del mio «intervento culturale» e solamente i più attenti e smaliziati potrebbero accorgersi, o per lo meno sospettare una tale manipolazione.

Più in generale credo di essere comunque riuscito a mantenere un certo equilibrio tra «adeguatezza» e una «accettabilità» traduttive. Ad esempio, nel caso di espressioni settoriali ho scelto di conservare il termine inglese solo se esso non ha traduzioni in italiano o se, pur avendone, è con buona probabilità ampiamente conosciuto o facilmente comprensibile per il mio lettore modello (es. «default»); per contro ho tradotto le espressioni che ammettono uno specifico traduttore in italiano (es. «aptitude test») e ho cercato un traduttore «accettabile» anche per quelle espressioni di più difficile e meno intuitiva comprensione (si vedano, ad esempio, le espressioni «criterion-referenced» e «norm-referenced»).

Ho riscontrato ulteriori elementi di marcatezza a livello lessicale. In particolare, in una citazione, si fa riferimento al criterio di «scoreability»; all'inizio ho pensato che si potesse trattare di un vocabolo creato *ad hoc* dall'autore citato per descrivere le caratteristiche di un determinato tipo di test attitudinale. In realtà, attraverso un'attenta analisi dei riscontri sui siti internet, ho potuto constatare un utilizzo abbastanza diffuso del vocabolo, seppur limitato a pochi settori specifici. Uno di questi è proprio quello della valutazione: sul sito <http://www.ed.psu.edu/insys/ESD/darling/Assess.html>, Linda Darling-Hammond, un'esperta americana impegnata sul fronte della riforma del sistema scolastico e dei metodi di valutazione, parla di «...efficiency and and easy *scoreability*...» in riferimento ai test di verifica. In questo contesto mi sembra abbastanza chiaro che per «scoreability» si debba intendere la possibilità di attribuire un punteggio o una votazione. Nella mia traduzione la scelta del giusto traduttore ha posto una difficoltà aggiuntiva: dato che il vocabolo è inserito in un'enumerazione, ho ritenuto che l'adozione, nel testo tradotto, di una perifrasi avrebbe appesantito

troppo l'enunciato rallentandone il ritmo; pertanto la scelta è caduta sul traducevole «valutabilità».

Il vocabolo «scoreability» trova anche ampia applicazione in un altro campo: l'industria della carta. Secondo il glossario della CE, Eurodicautom, esso sta a indicare l'«attitudine alla fustellatura» della carta o del cartone (presumibilmente usati per realizzare scatole), ovvero la capacità di tali materiali di resistere, senza subire danni o rotture, ad un processo di sagomatura eseguita con attrezzi specifici (fustella).

In un contesto totalmente diverso, quello sportivo e, nello specifico, nel mondo del bowling, il vocabolo «scoreability» in genere indica l'ottimizzazione di una boccia o della corsia – ad esempio, grazie all'impiego di materiali e tecnologie costruttive particolari – al fine di accrescere la percentuale di successo dei colpi, quindi di fare più punti. In senso lato la «scoreability» di una boccia è direttamente proporzionale alle sue qualità dinamiche e alla precisione che essa garantisce durante il gioco.

In altri contesti (es. medico-scientifico) si parla invece di «scoreability» in riferimento alla possibilità di quantificare o di attribuire un punteggio o all'affidabilità/precisione di dati forniti.

Considerando la natura strettamente tecnica e settoriale della trattazione, non è stato difficile ipotizzare quale fosse il «lettore modello» che gli autori avevano in mente al momento di scegliere la strategia testuale da adottare; con tutta probabilità era ed è l'esperto, lo studioso nel campo della misurazione o il mediatore linguistico specialmente anglofono. Sono in realtà le stesse figure professionali che io ho individuato come probabili destinatari dell'opera da me tradotta - ovvero i «lettori modello» del metatesto; è comunque ragionevole pensare che essa, in quanto prodotto di una cultura emittente extraeuropea, ottenga un successo più limitato nella cultura ricevente (italiana) e si rivolga dunque a quella ristretta cerchia di specialisti più colti o che mostrano più interesse e attenzione per le tendenze e i fenomeni che avvengono anche al di fuori dell'Italia o dell'Europa. Non ho effettuato cambiamenti traduttivi di rilievo, almeno per quanto concerne lo stile e il registro. A livello culturale, ho optato per un cambiamento generalizzante in un caso: ho tradotto «...California Court

certification examination...» con «esame di certificazione per gli interpreti giurati californiani», ritenendo che tale scelta non avrebbe pregiudicato la comprensione del testo da parte del mio lettore modello.

Nelle note conclusive del testo in inglese, la valutazione, ambito scientifico ancora quasi del tutto inesplorato ma che ha un grosso potenziale di sviluppo, viene paragonata ad un bambino che muove i primi passi; anche in questo caso ho preferito una traduzione generalizzante: «...assessment does need to...realize that there are some bigger kids on the block for it to learn from...» diventa dunque «...la valutazione deve...comprendere che ci sono discipline più evolute e mature dalle quali imparare...». Come si può notare ho compiuto una manipolazione stilistica eliminando l'immagine del bambino, poiché ho ritenuto che nel testo italiano avrebbe perso parte dell'efficacia e della bellezza che invece assume nell'originale. Si tratta comunque di una scelta personale e pertanto opinabile.

In conclusione, posso affermare che vi sia quindi una sostanziale corrispondenza tra la «strategia narrativa» adottata dagli autori e la mia «strategia traduttiva». Ho cercato di limitare il residuo comunicativo mantenendomi, per quanto possibile, fedele al testo originale, operando delle scelte traduttive che privilegiassero l'adeguatezza del lessico (ad eccezione dei casi descritti sopra) ma anche della sintassi e del registro e affidando a questo apparato metatestuale la spiegazione delle scelte traduttive meno comprensibili per la metacultura.

## Introduction

Translator and interpreter education is now widely practised around the world and is supported by an increasingly sophisticated body of research and scholarship. Much of this work is concerned with identifying the components of competence and proposing curriculum models that incorporate these components and suitable teaching strategies. The scholarship supporting translation and interpreting education necessarily entails discussions of assessment and there has been some encouraging work in this area. However, there has been little recognition in translation and interpreting circles that educational measurement as a broader field has its own tradition of scholarship, a widely accepted body of knowledge and terminology, and a range of approaches. Notions like reliability and validity are part of the basic architecture of educational measurement.

Test designers need to ensure that test results are reliable, for example, yielding the same results with different groups of candidates and at different points in time; and they need to construct tests that are valid in that they, for instance, reflect the model of learning that underpins the curriculum and are relevant to the professional behaviour taught in the curriculum. A major issue in educational measurement of relevance to translation and interpreting assessment is the fundamental difference of approach between norm- and criterion-referenced testing. Norm-referenced tests are designed to rank candidates against each other; criterion-referenced tests require candidates to demonstrate that they have satisfied a set criterion. These fundamental issues are comprehensively dealt with in standard works on educational measurement such as Ebel (1972) and Thorndike *et al.* (1991). Closer to our discipline, Bachmann (1991) represents a comprehensive discussion on language testing, firmly grounded in measurement theory.

## Introduzione

Oggi la formazione di traduttori e interpreti è un'attività ampiamente diffusa nel mondo ed è supportata da un corpus sempre più vasto e complesso di ricerche e conoscenze a livello accademico. Si tratta prevalentemente di studi che si concentrano sull'individuazione delle componenti della competenza e che propongono modelli di curricula che incorporano queste componenti nonché delle opportune strategie d'insegnamento. La dottrina alla base dell'insegnamento della traduzione e dell'interpretazione non può prescindere da discorsi relativi alla valutazione ed è proprio in questo ambito che sono stati compiuti incoraggianti passi avanti. Eppure nelle cerchie di interpreti e traduttori sono in pochi a riconoscere che la misurazione accademica intesa come più vasta materia abbia una propria tradizione di studi, un corpus di conoscenze e di terminologia ampiamente condiviso e una serie di approcci differenti.

Nozioni quali attendibilità e validità sono parte integrante della struttura di base della misurazione accademica.

Gli ideatori di test devono assicurarsi che i risultati dei test siano attendibili e che pertanto essi, per esempio, producano gli stessi risultati con diversi gruppi di candidati e in momenti diversi nel tempo. Allo stesso modo i test devono essere validi e quindi, ad esempio, rispecchiare il modello di apprendimento sotteso al programma di studi e l'atteggiamento professionale insegnato. Nell'ambito della misurazione accademica una delle questioni di primaria importanza relative alla valutazione della traduzione e dell'interpretazione è la differenza fondamentale di approccio tra test basati sulla norma e i test basati su criteri. I primi mirano a stabilire un raffronto tra i candidati; i secondi prevedono che i candidati dimostrino di aver soddisfatto una serie di criteri prefissati. Tali questioni fondamentali sono trattate in maniera esaustiva nelle opere di base sulla misurazione accademica di autori come Ebel (1972) e Thorndike *et alia* (1991). Più strettamente legata al nostro argomento risulta essere l'esauriente disamina di Bachmann (1991) sulle modalità di valutazione del linguaggio, profondamente radicata nella teoria della misurazione.

The discussion in this contribution sets itself outside a current and vital issue in Translation Studies (and to a much lesser extent in work on interpreting) – the instability of notions such as *quality*, *value* and *assessment*. A recent volume of *The Translator* was dedicated to this issue, with an introduction by Carol Maier that points out the difficulty of defining these concepts on the basis of theories about the nature of translation. Maier observes that «one sees a shared emphasis on defining and assessing quality in the context of specific situations, especially pedagogical ones» (Maier, 2000: 140). While we acknowledge the complexity and importance of defining these notions, we confess that we sidestep the issue and jump straight into Maier’s pedagogical context; our approach has been to scrutinize translation and interpreting assessment with the broader perspective of educational measurement. Using some fundamental criteria from educational measurement as a framework, we ask how current translation and interpreting assessment practice stands up to broader scrutiny, and what directions we need to take in the future.

La disamina in questo saggio si colloca al di fuori di una questione quantomai attuale e vitale nei *Translation Studies* (e in misura minore negli studi sull'interpretazione) – l'instabilità di concetti come *qualità*, *valore* e *valutazione*. Un recente numero del *Translator* è stato dedicato a questo argomento; nell'introduzione di Carol Maier si mette in evidenza come sia difficile definire questi concetti sulla base di teorie sulla natura della traduzione. Maier fa notare che «è possibile cogliere un comune interesse nel definire e nel valutare la qualità nel contesto di situazioni specifiche, soprattutto quelle pedagogiche» (Maier, 2000: 140). Se, da un canto, riconosciamo la complessità e l'importanza di definire tali nozioni, dall'altro confessiamo di eludere il problema e di tuffarci nel contesto pedagogico di Maier; il nostro approccio è consistito nell'analizzare attentamente la valutazione della traduzione e dell'interpretazione nella più ampia prospettiva della misurazione accademica. Prendendo come riferimento alcuni criteri fondamentali della misurazione accademica ci chiediamo quante volte la pratica della valutazione della traduzione e dell'interpretazione regga il confronto con un'analisi più ampia e approfondita e in quali direzioni dovremo muoverci in futuro.

## Basic Approach

Our basic approach has been to propose a checklist of criteria against which an assessment procedure might be measured. We have then examined a selection of published works that deal with translation and interpreting assessment procedures in some fashion, and weighed their findings against some of the criteria on the checklist. The works were collected through a search of the Linguistics and Language Behaviour Abstracts (LLBA) and Modern Language Association (MLA) databases, as well as our private collections. It is important to note that we limited our choice of works to those that deal specifically with assessment procedures in an educational context (including accreditation), for which reason the absence of seminal works like House (1981) should come as no surprise. We concede that the published works examined are by no means a comprehensive collection, but we maintain that they are a fair representation of the state of art over the last decades, as published; however, there is no doubt a good deal of interesting practice locked away in the internal documentation of teaching institutions.

The checklist is not intended as a definitive taxonomy of the characteristics of assessment procedures, and we acknowledge that there are overlaps between some of the items. For example, a procedure that aims at summative assessment may generate information that can be used for credit transfer (cf. item 2 below); but of course credit transfer information requires the additional potential for translatability between education systems or institutions. In an Australian educational institution, for instance, test procedures can have the purpose of producing (a) summative information so that the institution can award grades, (b) information that will allow accreditation by the external accrediting authority, and (c) information that will allow another institution to calculate the amount of credit to be granted. A single programme might include units whose assessment procedures do one of these things or several at once. Similarly, items 6 and 8 below overlap to an extent, but differ in their focus; item 6 is oriented towards the institution and its assessment policies while item 8 is oriented towards the broader constituency of stakeholders in the assessment process.

## Approccio di Base

Il nostro approccio di base è stato quello di proporre un elenco di criteri secondo i quali una procedura di valutazione potrebbe essere giudicata. In seguito abbiamo preso in esame una selezione di pubblicazioni che in qualche modo trattano di procedure di valutazione della traduzione e dell'interpretazione e abbiamo quindi confrontato i loro risultati sulla base di alcuni parametri dell'elenco. Per la scelta delle pubblicazioni abbiamo consultato, oltre alle nostre raccolte private, gli *Abstracts* del Linguistic and Language Behaviour (LLBA) e i data base della Modern Language Association (MLA). È importante sottolineare che ci si è limitati a scegliere quei testi che trattano specificamente di procedure di valutazione in un contesto accademico (compreso l'accreditamento), ragion per cui l'assenza di opere autorevoli come House (1981) non deve sorprendere. Sappiamo che le pubblicazioni esaminate non costituiscono in alcun modo una raccolta completa ma riteniamo che formino un quadro sufficientemente rappresentativo dello stato dell'arte del materiale pubblicato su questo argomento negli ultimi decenni; comunque non vi è dubbio che una gran quantità di interessanti procedure sia conservata nei documenti interni delle istituzioni accademiche.

La lista non va considerata una tassonomia definitiva delle caratteristiche delle procedure di valutazione e ammettiamo che vi siano delle sovrapposizioni tra alcune voci. Ad esempio, una procedura finalizzata a una valutazione sommativa può generare informazioni che possono essere usate per il trasferimento dei crediti (cfr. voce 2 sotto); ma ovviamente i dati relativi al trasferimento dei crediti necessitano del potenziale aggiuntivo per la loro traducibilità tra sistemi accademici o istituzioni. In una istituzione accademica australiana, per esempio, le procedure dei test possono servire a ottenere (a) informazioni sommative che permettano alle istituzioni di attribuire voti, (b) informazioni che permettano l'accreditamento da parte dell'autorità esterna preposta, e (c) informazioni che permetteranno a un'altra istituzione di calcolare l'ammontare del credito da assegnare. Un singolo programma può includere unità le cui procedure di valutazione svolgano una di queste funzioni o più funzioni allo stesso tempo. Allo stesso modo i punti 6 e 8 sotto si sovrappongono in parte, ma si concentrano su aspetti differenti; il punto 6 si rivolge alle istituzioni e alle relative politiche valutative mentre il punto 8 si rivolge alla più ampia cerchia di soggetti coinvolti nel processo valutativo.

The checklist follows:

- (1). What broad area is being assessed? For example, interpreting, translation, subtitling, specific language combinations, etc.
- (2). What is the purpose of the assessment instrument? For example, is aimed at:
  - Measuring aptitude (e.g. to enter a training course);
  - Determining placement (e.g. at a particular starting point in a training course);
  - Providing formative assessment (i.e. the skills and knowledge attained at points during a training course);
  - Providing summative assessment (i.e. the skills and knowledge attained at the end of a training course);
  - Accreditation (e.g. for entry into a professional body);
  - Credit transfer (e.g. to allow student mobility between universities)?
- (3). What competencies are assessed, e.g. language 1 and language 2 knowledge, transfer competence, speed, accuracy, memory, terminology, cultural knowledge, etc.?
- (4). What is the form of the assessment instrument? For example, a timed translation, an interpreting role play, a multiple choice test, etc
- (5). What is the basic approach of the instrument? For example, is it norm-referenced, i.e. ranking candidates from best to worst; or is it criterion-referenced, i.e. measuring performance against a known criterion? Or does it assess skills learned on the job?
- (6). What kind of results does the instrument generate? For example, does it generate a qualitative description of performance, a numerical score based on the objective items, a pass/fail result, etc?
- (7). How well does a norm-referenced instrument discriminate among candidates?

La lista è la seguente:

- (1) Quale area generica viene valutata? Ad esempio, interpretazione, traduzione, sottotitolaggio, combinazioni linguistiche specifiche ecc.
- (2) Quale scopo ha lo strumento valutativo? Ad esempio, è finalizzato a:
  - Misurare l'attitudine (es. a partecipare a un corso di formazione);
  - Stabilire l'inserimento (es. ad un particolare punto di partenza in un corso di formazione);
  - Fornire una valutazione formativa (cioè le capacità e le conoscenze acquisite in diversi momenti durante un corso di formazione);
  - Fornire una valutazione sommativa (cioè le capacità e le conoscenze acquisite al termine di un corso di formazione);
  - L'accreditamento (es. per aver accesso a un ordine professionale);
  - Trasferimento dei crediti (es. che permetta la mobilità degli studenti tra università).
- (3) Quali competenze sono valutate, es. conoscenza della lingua 1 e della lingua 2, capacità traduttive, velocità, accuratezza, memoria, terminologia, conoscenze culturali, ecc.?
- (4) Quale forma deve avere lo strumento di valutazione? Ad esempio, una traduzione con limite di tempo, un gioco di ruolo interpretativo, un test a scelta multipla, ecc.
- (5) Qual è l'approccio di base dello strumento? Ad esempio, è un test basato sulla norma, cioè finalizzato a classificare i candidati dal migliore al peggiore; o un test basato su criteri, cioè finalizzato a misurare la prestazione sulla base di un dato criterio? Oppure valuta capacità acquisite al lavoro?
- (6) Che tipo di risultati fornisce lo strumento? Ad esempio, fornisce una descrizione qualitativa della prestazione, un punteggio numerico basato su elementi oggettivi, indica il superamento o il non superamento della prova, ecc.?
- (7) Quale utilità può avere uno strumento basato sulla norma nel differenziare i candidati?

- (8) What are the reporting mechanisms? For example:
- Who receives feedback (e.g. candidate, instructor, institution)?
  - When does feedback occur (e.g. immediately, months later)?
  - How is feedback given (e.g. qualitatively, quantitatively)?
- (9) How valid is the assessment instrument?
- (10) How reliable is the assessment instrument?

We note in advance of the discussion that some of the items in the list are simply not discussed in the materials that we examined. We will return to these gaps later in this chapter.

- (8) Quali sono i meccanismi relazionali? Ad esempio:
- Chi riceve il feedback (es. il candidato, l'istruttore, l'istituzione)?
  - Quando viene recepito il feedback (es. immediatamente, dopo alcuni mesi)?
  - Come viene dato il feedback (es. qualitativamente, quantitativamente)?
- (9) Quale validità ha lo strumento di valutazione?
- (10) Quale attendibilità ha lo strumento di valutazione?

Prima di addentrarci nel discorso segnaliamo che alcuni punti della lista non sono trattati nel materiale da noi esaminato. Torneremo a parlare di queste lacune più avanti in questo capitolo.

## Translation Assessment

Works on assessment in translation can be divided into two broad categories of assessment purpose: *accreditation* and *pedagogy*, reflecting the two broad constituencies of recruitment and training. In the accreditation area, Schäffner (1998) provides a critique of the Institute of Linguists syllabus in German. A UN accreditation perspective from Beijing is provided by Wu (1994), while Bell (1997), Martin (1997) and Ozolins (1998) discuss national accreditation in Australia. The offerings from Beijing and Australia each deal with both interpreting and translation, while all the works deal in some fashion with tests that bestow a public validation of competence. The pedagogy area in translation is less clear cut in terms of purpose: Brunette (2000) makes some reference to translation didactics in her attempt to establish a terminology for translation quality assessment, but is not clear about purpose, for example, diagnostic, formative or summative assessment. Dollerup (1993), Kussmaul (1995) and Sainz (1993) are clearly concerned about formative assessment, while Farahzad (1992) and Ivanova (1998) discuss summative assessment in the form of final translation examinations at university. James *et al.* (1995) is the only work in our selection to examine credit transfer (in the area of screen translation), while Campbell (1999) makes some small inroads into diagnostic assessment. Interestingly, we came across very little discussion of aptitude testing for translator education, although Cestac (1987) describes selection tests for recruitment at UN Headquarters; conversely there is a good deal of discussion of aptitude for interpreter education (cf. below). What is also interesting is that a number of writers discussed translation assessment without making any reference to purpose (for example, Bowker, 2000). Something of a hybrid is the Institute of Linguists New Diploma in English and Chinese described by Ostarhild (1994), which appears to be an attempt to move an accreditation instrument from an earlier test of bilingualism to one that also tests translation.

## Valutazione della Traduzione

È possibile suddividere la letteratura sull'argomento della valutazione in traduzione in due ampie categorie a seconda che lo scopo della valutazione sia l'*accreditamento* o la *pedagogia*: il primo scopo si collega all'ambito dell'assunzione, il secondo a quello della formazione. In materia di accreditamento, Schäffner (1998) ci lascia una critica in tedesco del programma d'insegnamento dell'Institute of Linguists, Wu (1994), da Pechino, esprime il suo punto di vista sull'accREDITamento presso le Nazioni Unite mentre Bell (1997), Martin (1997) e Ozolins (1998) disquisiscono di accREDITamento nazionale in Australia. I contributi provenienti da Pechino e dall'Australia trattano sia di interpretazione sia di traduzione, mentre più in generale tutti i testi presi in rassegna si occupano in qualche modo dei test che certificano pubblicamente delle competenze. Lo scopo pedagogico nella traduzione è meno chiaro: Brunette (2000) fa qualche accenno alla didattica della traduzione nel tentativo di stabilire una terminologia per la valutazione della qualità della traduzione, ma non si esprime chiaramente in merito allo scopo, quale può essere, ad esempio, quello diagnostico, formativo o di una valutazione sommativa. Dollerup (1993) e Sainz (1993) dimostrano un chiaro interesse per la valutazione formativa, mentre Farahzad (1992) e Ivanova (1998) parlano di valutazione sommativa negli esami finali di traduzione all'università. L'opera di James *et alia* (1995) è l'unica, fra quelle prese in esame, ad analizzare il trasferimento dei crediti (nel campo della traduzione per i media), mentre Campbell (1991) fa qualche breve excursus nel campo della valutazione diagnostica. È interessante osservare che raramente ci siamo imbattuti in qualche discorso sui test attitudinali per traduttori, anche se Cestac (1987) parla di test di selezione per l'assunzione presso i Quartieri Generali delle Nazioni Unite; per contro abbiamo trovato molto materiale sull'attitudine all'interpretazione (cfr. sotto). Altrettanto degno di nota è il fatto che diversi scrittori abbiano parlato di valutazione della traduzione senza fare riferimento allo scopo (si veda ad esempio Bowker, 2000). Risulta ibrido invece il New Diploma in English and Chinese dell'Institute of Linguists descritto da Ostarhild (1994), che assomiglia a un tentativo di trasferire uno strumento di accREDITamento da un precedente test di bilinguismo ad un altro che valuta anche la traduzione.

The types of translation competencies discussed range widely, but a crucial factor seems to be the extent to which translation is integrated into a socio-communicative framework. Where translation is not linked to such a framework, a default position seems to operate, in which competencies are largely *target language focused*. An example of this type is Dollerup's (1993) assessment scheme for translation in the framework of language study, which works empirically from target language (TL) error analysis in order to construct student feedback form that assesses detailed competencies grouped under *text, spelling, punctuation, words/word knowledge, syntax/grammar* and *expression*. Sainz (1993) develops a similar feedback chart that allows students to critique their own work, but does not specify the competencies, other than to suggest that teachers can compile a «chart of "Types of Mistakes"»; she suggests that for a particular text it might include *connectors, grammar, lexical items, misunderstanding, nouns (agreement), omission, prepositions, punctuation, style, register, syntax, and tenses*. Farahzad's (1992) list is somewhat different: *accuracy, appropriateness, naturalness, cohesion, style of discourse/choice of words*. Ivanova (1998) tells us a little about translation assessment at the University of Sophia; although she provides a review of literature on translation competence, the final examination marking scheme described simply deals with *lexical infelicities, lexical error, grammatical mistake* and *stylistic inappropriateness*.

Scholars working within a communicative framework grounded in theory tend to go beyond the classification of TL errors. An example is the approach taken by Hatim and Williams (1998), who, although they do not mention assessment in their discussion of a university translation programme in Morocco, do outline a syllabus based on a sophisticated model of communication which aims to have students «negotiate the transaction and exploit the signs...which surround them». Very detailed objectives – presumably reflecting the competencies to be assessed – cascade from these broad aims. Similar is the approach of Kussmaul (1995), who lists a number of «categories of evaluation» of texts, which seem to us to reflect competencies (he is after all dealing with translator education).

I tipi di competenze richieste in traduzione finora discusse variano enormemente, ma un fattore di fondamentale importanza sembra essere il grado di integrazione di una traduzione in un quadro socio-comunicativo. Laddove la traduzione non è collegata a questo quadro, sembra intervenire una posizione di *default* che vede le competenze concentrarsi prevalentemente sulla lingua ricevente. Un esempio di questo tipo è lo schema di valutazione della traduzione nel quadro dello studio della lingua elaborato da Dollerup (1993); di fatto tale schema parte dall'analisi dell'errore nella lingua ricevente al fine di costruire una forma di feedback dello studente che valuti una serie dettagliata di competenze raggruppate sotto le seguenti voci: *testo, spelling, punteggiatura, conoscenza di parole/parola, sintassi/grammatica ed espressione*. Sainz (1993) elabora uno schema di feedback simile che permetta agli studenti di analizzare in modo critico il proprio lavoro; egli non specifica le competenze, ma lascia intendere la possibilità che siano gli insegnanti stessi a compilare un «grafico dei “Tipi di errore”». Tale grafico potrebbe includere per uno specifico testo i *connettori, la grammatica, gli elementi lessicali, gli errori di comprensione, i nomi (concordanza), le omissioni, le preposizioni, la punteggiatura, lo stile, il registro, la sintassi e i tempi*. Piuttosto differente è la lista proposta da Farahzad, che comprende: *accuratezza, appropriatezza, naturalezza, coesione, stile del discorso/scelta lessicale*. Ivanova (1998) ci parla brevemente della valutazione della traduzione presso l'Università di Sofia; sebbene fornisca un riesame della letteratura sulla competenza traduttiva, lo schema di valutazione dell'esame finale da lei presentato tratta semplicemente di *scelte lessicali infelici, errori lessicali e grammaticali e di stile non appropriato*.

Gli studiosi che operano in un quadro comunicativo teorico tendono ad andare oltre la classificazione degli errori nella lingua ricevente. Ne è un esempio l'approccio di Hatim e di Williams (1998), che, sebbene non parlino di valutazione nella loro analisi di un programma di traduzione universitario in Marocco, delineano un programma di insegnamento basato su un sofisticato modello comunicativo finalizzato a spingere gli studenti a «negoziare la transazione e a sfruttare i segni...che li circondano». Da questi intenti generali scaturisce una serie molto dettagliata di obiettivi – i quali presumibilmente riflettono le competenze da valutare. Simile è l'approccio di Kussmaul (1995), che elenca una serie di «categorie di valutazione» dei testi che sembrerebbero riflettere delle competenze (dopo tutto si sta occupando della formazione del traduttore).

These are: *cultural adequacy, situational adequacy, speech acts, meaning of words, «language errors»* (Kussamaul's quotation marks). Integrated into a professional context is the scheme of James *et al.* (1995), where a blend of linguistic and technical competencies is achieved in a discussion of screen translation assessment. The groups of competencies are *portrayal, language quality, grammar, spelling, punctuation* and *time-coding, synchronisation, positioning, colour, breaks between subtitles* respectively. A professional framework also informs the competencies assessed by Australia's National Accreditation Authority for Translators and Interpreters (NAATI) (Bell, 1997). An approach to competencies beyond the mere listing of TL criteria is also found in the findings of experimental tests reported by Niedzielski and Chernovaty (1993) (dealing with both translation and interpreting in technical fields). The authors claimed (1993: 139) that «maturity and experience in some technical field(s)» and «original and creative thinking» were «factors found to achieve success in translating», on the basis of measuring information errors, lexical errors, grammatical errors, referential errors, style mistakes, and other criteria (cf. 1993: Tables 1-5, 144-6).

The translation of a text appears to be the standard form for translation assessment, although Ostarhild (1994) describes such tasks as skimming and scanning material in English and Chinese and producing «written commentaries in the other language» (1994: 53). The test described – the Institute of Linguists New Diploma in English and Chinese – is, as mentioned above, a kind of hybrid test of translation and bilingualism.

Surprisingly there seems to be very little discussion of the ideal length of translation tests or the time allowed for their completion, let alone any theoretically or empirically based findings on the subject. In the accreditation area, NAATI follows the curious practice of a strict time constraint on examinations at the basic Professional level (500 words in two hours), but a much more generous allowance at the Advanced level. Dollerup (1993) uses texts ranging from 50 to 700 words in his classroom-based model, presumably on the basis that students can handle longer texts as skill increases. Farahzad (1992) is braver, describing a range of test types including single sentences for translation and whole texts of 200 words.

Nello specifico: *adeguatezza culturale e situazionale, enunciazioni, significato delle parole, «errori linguistici»* (le virgolette sono di Kussmaul). Inserito in un contesto professionale è lo schema di James *et alia* (1995), i quali, in una disamina sulla valutazione della traduzione per i media, arrivano a fondere una serie di competenze *linguistiche e tecniche*. I gruppi di competenze sono rispettivamente *qualità della resa linguistica, qualità del linguaggio, grammatica, spelling, punteggiatura e codifica temporale, sincronizzazione, posizione, colore, intervalli tra i sottotitoli*. Il contesto professionale è alla base anche delle competenze valutate dall'australiana National Accreditation Authority for Translators and Interpreters (NAATI) (Bell, 1997). È possibile riscontrare un approccio alle competenze che va oltre il mero elenco di criteri della lingua ricevente anche nei risultati di alcuni test sperimentali riportati da Niedzielski e Chernovaty (1993) (con risvolti negli aspetti tecnici della traduzione e dell'interpretazione). Gli autori affermavano (1993: 139) che «la maturità e l'esperienza in alcuni ambiti tecnici» e una «forma mentis originale e creativa» fossero «fattori chiave per avere successo nella traduzione», basandosi sulla valutazione degli errori di informazione, lessicali, grammaticali, referenziali, di stile e altri criteri (cfr. 1993: Tavole 1-5, 144-6).

La traduzione di un testo sembra essere la modalità standard per la valutazione della traduzione, anche se Ostarhild (1994) parla di attività quali l'esame superficiale e approfondito di materiale in inglese e cinese seguita dalla realizzazione di «commenti scritti nell'altra lingua» (1994: 53). Il test descritto – il New Diploma in English and Chinese dell'Institute of Linguists – è, come già detto, una specie di test ibrido di traduzione e bilinguismo.

Desto sorpresa il fatto che, apparentemente, il problema della lunghezza ideale dei test di traduzione o del tempo massimo consentito per il loro svolgimento non sia stato quasi per niente affrontato e che in merito non esistano risultati su basi empiriche o teoriche. Nell'ambito dell'accreditamento, le procedure seguite dalla NAATI risultano alquanto insolite, poiché prevedono ristretti margini di tempo per gli esami al gradino più basso del livello *Professional* (500 parole in due ore) e condizioni decisamente più favorevoli al livello *Advanced*. Dollerup (1993) impiega testi che variano dalle 50 alle 700 parole nel suo modello basato sulla classe, presumibilmente partendo dal presupposto che gli studenti riescono a gestire testi più lunghi man mano che le loro capacità aumentano. Farahzad (1992) è più coraggioso, e presenta una gamma di modelli di test che spaziano dalla traduzione di singole frasi a interi testi di 200 parole.

Cestac (1987) describes the various UN examination papers, which include a 700-word general translation in three hours, a 2000-word summary in two hours, two 400-word specialized translations in three hours, and two 300-word translations from the candidates' non-main language in two hours. Farahzad stands out in suggesting «limited response» items, where students are faced with, for example, several translations of a sentence and are asked to select the error-free version.

Little is written about the basic approaches of test instruments, and it is difficult to ascertain whether norm-referenced or criterion-referenced approaches are generally favoured. The upside-down marking scheme that seems to be commonly used (error marks being deducted from a perfect score) is so odd as to defy categorisation. Admittedly it is possible to establish a rank order of candidates using error marking (i.e. the top candidate is the one with the least errors), just as one can establish criteria for passing (i.e. every candidate with less than  $n$  errors passes). But the fundamental mathematics are so peculiar that we would have to be careful in determining whether it reflects a norm-referenced or criterion-referenced approach. Error marking works very well for TV quiz shows, because the number of correct responses equals the perfect score. But for translation the number of correct responses is infinite (on the reckoning that any translation can be done in an infinite number of ways) or very large (on the reckoning that there is a very large number of possible errors in any translation). The theoretical consequence is a ranked scale with an infinitely long tail. Let us say that the «perfect score» is 100, and that the two top candidates score 90 and 95. Now if the bottom candidate scores 0 and there is a normal distribution of scores in the candidature we have some sense of the relativities and we can compute means, standard deviations, z-scores, and the like – the tools of the trade in norm-referenced assessment. The problem comes when poor candidates score below zero (even though the marker may report the result as zero) – perhaps minus 20, minus 30, minus 80, or minus anything at all. Because there is no bottom to the scale, we have no way to assess the relative achievement of the top scoring candidates; depending on where the bottom of the scale finds itself, one may be very good and one exceptional, or perhaps they are separated by a whisker.

Cestac (1987) descrive i vari documenti sugli esami delle Nazioni Unite tra i quali figura una traduzione generica di 700 parole con limite di tempo di tre ore, un riassunto di 2000 parole con limite di tempo di 2 ore, due traduzioni specialistiche di 400 parole in tre ore e due traduzioni di 300 parole dalla lingua straniera dei candidati con limite di 2 ore. Farahzad si segnala per la sua proposta di introdurre prove «a risposta breve» nelle quali gli studenti sono chiamati, ad esempio, a selezionare, tra le varie elencate, la versione corretta della traduzione di una frase.

Sugli approcci di base degli strumenti dei test si è scritto poco, e pertanto risulta difficile stabilire se in genere si propenda per approcci basati sulla norma o su criteri. Lo schema di valutazione *upside-down* che sembra essere comunemente usato (si parte da un punteggio pieno dal quale si detraggono i punti degli errori) è così strano da essere difficilmente classificabile. Ovviamente è possibile stabilire una classifica dei candidati usando il calcolo degli errori (cioè il candidato migliore è quello che ha fatto meno errori) come anche stabilire i criteri per il superamento del test (cioè tutti i candidati che hanno fatto meno di  $n$  errori superano il test). Tuttavia i calcoli matematici di base sono talmente particolari che dovremmo stare molto attenti a stabilire se rifletta un approccio basato sulla norma o su criteri. La valutazione basata sul calcolo degli errori funziona molto bene per i quiz televisivi perché il numero delle risposte corrette equivale al punteggio massimo. Ma nella traduzione il numero di risposte corrette è infinito (se si considera che una traduzione può essere fatta in infiniti modi) o comunque molto vasto (se si considera che esistono molti possibili errori in una traduzione). Ne conseguirebbe, da un punto di vista teorico, una scala di valutazione che non ha mai fine. Supponiamo che il «punteggio massimo» sia 100 punti e che i due migliori candidati totalizzino rispettivamente 90 e 95 punti. Ora, se il candidato peggiore totalizza 0 punti e vi è una normale distribuzione dei voti tra i candidati, abbiamo un senso della relatività tra i candidati e possiamo calcolare medie, deviazioni standard, *z-scores* e così via tutti gli elementi per una valutazione basata sulla norma. I problemi nascono quando vi sono dei candidati che ottengono un punteggio inferiore a 0 (anche se il valutatore può riportare un punteggio di 0) – come meno 20, meno 30, meno 80 o meno qualsiasi punteggio. Poiché la scala di valutazione non ha un minimo, non c'è modo di valutare i risultati relativi dei candidati migliori; a seconda di dove si trovi il minimo stesso della scala, un candidato potrebbe essere molto buono, un altro addirittura eccezionale, o magari i due differiscono di poco.

In fact the balance of evidence shows that error deduction marking is really a criterion-referenced system, in which the number of marks in a perfect score is arbitrary and bears no relation to the possible number of errors. A pass mark (i.e. the perfect score less the maximum number of errors tolerated) is simply an indication of a criterion. If this is true, then a list of ranked scores based on error deduction is no more than a kind of statistical window dressing. Teague (1987), in describing the accreditation marking scheme of the American Translators Association, confirms this. Although «the grader...totals up the errors, and applies a final scale to get a final mark», the result is simply «fail» or «pass» (1987: 22). As a postscript, Bastin (2000) emphasises that «trainees must be taught how to do things right rather than being punished for what they have done wrong» (2000: 236); as both university teachers and accreditation examiners, the present authors are deeply unhappy about the practice of importing error deduction techniques into the educational context.

In effetti, i fatti indicano come la valutazione basata sulla deduzione degli errori sia proprio un sistema basato su criteri, nel quale il numero di punti massimo è arbitrario e non è in alcun modo collegato al possibile numero di errori. Un voto che permetta di superare l'esame (es. il punteggio massimo meno il massimo numero di errori tollerati) è semplicemente un'indicazione di un criterio. Se ciò fosse vero, classificare una serie di punteggi in base al criterio della deduzione dei voti non differirebbe molto dal compiere una operazione statistica di facciata. Teague (1987), nel descrivere lo schema di valutazione della American Translators Association conferma quanto appena detto. Sebbene «il valutatore...sommi tutti gli errori, e applichi una scala di valutazione finale per ottenere il voto finale» il risultato sarà semplicemente il superamento o il non superamento dell'esame. (1987: 22). In conclusione, Bastin (2000) fa notare che «sarebbe meglio insegnare agli studenti come evitare gli errori piuttosto che punirli per quelli che commettono» (2000: 236); sia come docenti universitari che come esaminatori nell'ambito dell'accREDITAMENTO, questi autori sono estremamente scontenti dell'impiego di tecniche di deduzione degli errori nel contesto accademico.

## Interpreting Assessment

There is very little written on interpreting assessment (Hatim & Mason, 1997). This may be partly due to the relatively few formal courses in the field worldwide, to the limited research in the area, and to the intuitive nature of test design and assessment criteria. The little literature that exists on interpreting assessment is dominated by discussions on aptitude tests for entry to conference interpreting courses (Keiser, 1978; Gerver *et al.*, 1984; Gerver *et al.*, 1989; Longley, 1989; Bowen & Bowen, 1989; Lambert, 1991; Moser-Mercer, 1994; Arjona-Tseng, 1994). The other categories include: accreditation or certification examinations to enter the profession, in particular community interpreting and court interpreting (Bell, 1997; Gentile, 1997; Scweda Nicholson & Martinsen, 1997; Miguélez, 1999; Vidal, 2000); testing that is related to interpreter training courses, most of which train conference interpreters (Longley, 1978; Macintosh, 1995; Schjoldager, 1995); and quality assessment of interpreting performance, mainly of professional conference interpreters (Pöchhacker, 1993; Bühler, 1986; Kopczynski, 1992; Dejean Lefeal, 1990; Kalina, 2001). The last category will not be discussed here given our focus on educational contexts.

Common to all aptitude tests described in the literature are the competencies the tests aim to assess, the subjective marking criteria, and the high failure rate. There is general agreement on the skills and abilities necessary of a trainee interpreter to succeed in a conference interpreting course or in the profession (Lambert, 1991), although this is not based on any empirical data, but rather on intuitive judgements by trainers who are mostly practising interpreters. These competencies include: good knowledge of the relevant languages, speed of comprehension and production, good general knowledge of the world, good public speaking skills, good memory, stress tolerance and ability to work as a team. The tests tend to be criterion-referenced, with candidates required to reach each criterion in order to pass the test. In some of the tests, the initial components act as eliminatory components, where a candidate cannot progress to the next phase of the examination, if he or she fails any of the preceding phases.

## Valutazione dell'Interpretazione

Molto esigua è la quantità di testi scritti sulla valutazione dell'interpretazione (Hatim & Mason, 1997). Ciò può essere in parte ascritto alla relativa carenza di corsi ufficiali in questo campo a livello mondiale, alla limitata attività di ricerca nel campo e alla natura intuitiva dell'ideazione dei test e dei criteri di valutazione. La poca letteratura esistente sulla valutazione dell'interpretazione è dominata da discussioni sui test attitudinali per accedere a corsi di interpretariato di conferenza (Kaiser, 1978; Gerver *et alia*, 1984; Gerver *et alia*, 1989; Longley, 1989; Bowen & Bowen, 1989; Lambert, 1991; Moser-Mercer, 1994; Arjona-Tseng, 1994). Tra le altre categorie di test figurano: gli esami di accreditamento o certificazione per avere accesso alla professione, in modo particolare all'interpretariato di comunità o in campo giuridico/giudiziario (Bell, 1997; Gentile, 1997; Schweda Nicholson & Martinsen, 1997; Miguélez, 1999; Vidal, 2000); i test relativi a corsi di formazione d'interpretariato, in special modo quelli dedicati agli interpreti di conferenza (Longley, 1978; Macintosh, 1995; Schjoldager, 1995); e la valutazione qualitativa della prestazione interpretativa in modo particolare di interpreti di conferenza professionisti (Pöchhacker, 1993; Bühler, 1986; Kopczynski, 1992; Dejean Lefeal, 1990; Kalina, 2001). Avendo circoscritto la nostra analisi ai contesti accademici non tratteremo in questa sede l'ultima categoria.

Comuni a tutti i test attitudinali descritti nella letteratura presa in esame sono le competenze che i test mirano a valutare, i criteri soggettivi di valutazione, e l'alto tasso di insuccesso. Vi è consenso generale sulle capacità e le competenze che un apprendista interprete deve necessariamente possedere per poter riuscire in un corso di interpretariato di conferenza o nella professione stessa (Lambert, 1991), anche se tutto ciò non trova riscontro nei dati empirici quanto piuttosto nei giudizi intuitivi di istruttori che per lo più formano interpreti. Tra queste competenze figurano: una buona conoscenza delle lingue in questione, velocità di comprensione e produzione, una buona conoscenza generale del mondo, buone capacità di esprimersi in pubblico, una buona memoria, la capacità di sopportare lo stress e di lavorare in gruppo. I test seguono generalmente il modello basato su criteri in cui i candidati sono chiamati a soddisfare uno per uno tutti i criteri per poterli superare. In alcuni tipi di test, le componenti iniziali fungono da componenti selettive e quindi un candidato non può accedere a una fase successiva se non ha superato quelle precedenti.

The forms of the assessment instruments are also shared by most entrance/aptitude tests. These include shadowing, cloze tests (both oral and written), written translation, sight translation, memory tests, and interviews. The rigour of these entrance tests and their high failure rates have led some to question the appropriateness of these assessment instruments which seem to expect applicants to perform almost at the level of professional interpreters before they even commence the training course (Gerver *et al.*, 1984). The predictive power of the tests and the lack of objectives assessment criteria used have also been criticised by some, who advocate research to correct these deficiencies (Gerver *et al.*, 1989; Arjona-Tseng, 1994; Moser-Mercer, 1994).

The reliability of the test results is very difficult to ascertain. As Moser-Mercer (1994) points out, there are no standardized interpreting aptitude tests. In spite of the advances made in language testing, little of that knowledge has been adopted by interpreter educators in the design of their testing (Moser-Mercer, 1994; Hatim & Mason, 1997). Bowen and Bowen (1989; 111) state that their aptitude tests are based on «Robert Lado, then Dean of Georgetown University's School of Languages and Linguistics and his criteria of validity...reliability...scoreability...economy...and administrability» but, apart from mentioning a standardised English terminology test recommended by the University's Psychology Counselling Centre, there is no other mention of how the tests are assessed for validity and reliability. Moser-Mercer (1994: 65) comments that Bowen and Bowen's standardisation «in no way meets the criteria for true standardisation». Two apparently well-motivated testing procedures are reported in the literature, by Gerver *et al.* (1984: 1989) and Arjona-Tseng (1994). Gerver *et al.* (1984: 1989) report the results of a research project which developed and assessed a set of psychometric aptitude tests. The aim of the study was to lead to the establishment of objective criteria for the entrance tests used for the postgraduate conference interpreting course run by the Polytechnic of Central London. At the time of the study, only two thirds of students who passed the initial aptitude test successfully completed the intensive six-month course. The final examination comprised language specific interpreting tests in both the consecutive and simultaneous modes.

Anche le forme degli strumenti valutativi sono comuni alla maggior parte dei test d'ingresso o attitudinali. Citiamo ad esempio lo *shadowing*, i *cloze test* (orali e scritti), la traduzione scritta, la traduzione a vista, i test mnemonici e i colloqui. La severità di questi test d'ingresso e l'alto tasso di insuccessi ha portato alcuni a mettere in dubbio la validità di tali strumenti valutativi che sembrerebbero dare per scontato che i candidati, ancora prima di cominciare il corso di formazione, rendano quasi al livello degli interpreti professionali (Gerver *et alia*, 1984). Quest'ultimo aspetto dei test e la mancanza di criteri valutativi oggettivi impiegati sono pure stati oggetto di critiche da parte di alcuni per i quali la ricerca dovrebbe colmare queste lacune. (Gerver *et alia*, 1989; Arjona-Tseng, 1994; Moser-Mercer, 1994).

L'attendibilità dei risultati dei test è molto difficile da appurare. Come Moser-Mercer (1994) ha fatto notare, non esistono modelli standardizzati di test attitudinali d'interpretazione. Nonostante i progressi compiuti nel campo della valutazione linguistica, solo una piccola parte delle conoscenze acquisite è stata impiegata dagli insegnanti di interpretazione nel mettere a punto i loro test (Moser-Mercer, 1994; Hatim & Mason, 1997). Bowen & Bowen (1989: 111) dichiarano che i loro test attitudinali si basano su «Robert Lado, allora preside della School of Languages and Linguistics della Georgetown University e su i suoi criteri di validità...attendibilità...valutabilità...economia...e amministrabilità» ma, al di là di una menzione al test standardizzato di terminologia inglese raccomandato dallo University's Psychology Counselling Centre, non vi sono altri riferimenti su come valutare la validità e l'attendibilità dei test. Moser-Mercer (1994: 65) commenta la standardizzazione di Bowen & Bowen sostenendo che «non soddisfa in alcun modo i criteri per una reale standardizzazione».

Due procedure di verifica, apparentemente con una buona motivazione, sono menzionate nella letteratura analizzata da Gerver *et alia* (1984; 1989) e da Arjona-Tseng (1994). Gerver *et alia* (1994; 1989) riportano i risultati di un progetto di ricerca finalizzato a sviluppare e valutare una serie di test attitudinali psicometrici. L'obiettivo dello studio era stabilire dei criteri oggettivi per i test d'ingresso per il corso di laurea di secondo livello d'interpretariato di conferenza organizzato dal Polytechnic of Central London. All'epoca dello studio, solo due terzi degli studenti che avevano passato il test attitudinale iniziale erano stati in grado di completare con successo il corso intensivo di sei mesi. L'esame finale comprendeva specifici test interpretativi linguistici in consecutiva e in simultanea.

The study looked at three types of tests: text-based, drawing on work done in the area of text processing (Kintsch, 1974); sub-skill based, drawing on the work on cognitive tests (Eckstrom *et al.*, 1976), and stress-based, drawing on the work done on speed testing (Furneaux, 1956). Sub-tests were conducted under each of these broad categories. Under the text-based test there were the following sub-tests: recall-text memory, recall-logical memory, completion/deletion – cloze, completion/deletion – error detection. Under the sub-skill-based test there were: a synonyms test, an expressional fluency test where candidates had to rewrite a test, and a verbal comprehension test. For the stress-based test, the team used an existing instrument, the Nufferno test (Furneaux, 1965), which measures the effect of speed stress on a cognitive task. The results of these tests were compared with the results of the final examinations. The study found that candidates who passed the final interpreting examination had scored higher on all the entrance tests than those who failed. The researchers conclude that «the tests appear to have been successful in reflecting generally the abilities required for interpreting» (Gerver *et al.*, 1984: 27).

Arjona-Tseng emphasises the dearth of literature on «rater-training issues, decision-making rules, reliability and validity issues, scaling, scoring, and test-equating procedures» (1994: 69). She attempts to address this need by proving a psychometrically-based approach to the development of entrance tests, with a standardised set of administration procedures, a tighter set of assessment criteria, appropriate rater training, and pilot testing. These new tests have been used at the Graduate Institute of Translation and Interpretation Studies at Fu Jen Catholic University with a 91% success rate for those selected to complete the course. Arjona-Tseng stresses the need for valid and reliable aptitude tests for admission to interpreter training courses.

Although aptitude testing dominates the interpreting assessment literature, a small literature exists on accreditation or certification examinations for professional recognition. Most accreditation or certification examinations are conducted in the area of community interpreting in general, or specifically for court interpreting. Few countries train interpreters in community interpreting or use university courses as the only entry path to the profession.

Lo studio prendeva in esame tre tipi di test: quelli basati sul testo, che si ispiravano al lavoro svolto nell'ambito della elaborazione di testi (Kintsch, 1974); quelli basati sulle sottocompetenze, ispirati ai test cognitivi (Eckstrom *et alia*, 1976); e quelli basati sullo stress, ispirati ai test di verifica della velocità (Furneaux, 1956). Dei subtest erano stati condotti per ognuna di queste tre ampie categorie. I test basati sul testo comprendevano le seguenti sottocategorie di test: capacità di memorizzare testi o nessi logici, *cloze test* con possibilità di completamento/eliminazione, individuazione di errori con possibilità di completamento/eliminazione. I test basati sulle sottocompetenze comprendevano: test sui sinonimi, test sulla scorrevolezza espressiva in cui i candidati erano chiamati a riscrivere una testo, e un test sulla comprensione verbale. Per il test sullo stress, il team si era servito di uno strumento già esistente, il test Nufferno (Furneaux, 1965) che misura gli effetti dello stress dovuto alla velocità su un compito cognitivo. I risultati dei test erano stati confrontati con gli esiti degli esami conclusivi. Ne era emerso che i candidati che avevano superato l'esame finale di interpretazione erano gli stessi che avevano ottenuto i punteggi più alti in tutti i test di ammissione. I ricercatori avevano concluso che «i test sembrano aver riflesso in linea di massima le capacità richieste per l'interpretazione» (Gerver *et alia*, 1984: 27).

Arjona-Tseng pone l'accento sulla penuria di test su questioni relative alla formazione dei valutatori o riguardanti le regole del processo decisionale, i problemi di attendibilità e validità, le scale di valutazione, il punteggio e le procedure di equiparazione dei test (1994: 69). La studiosa cerca di sopperire a queste mancanze fornendo un approccio fondato sulla psicomelia per lo sviluppo di test d'ingresso, il quale prevede una serie di procedure amministrative standardizzate, un più esiguo numero di criteri valutativi, un'appropriata formazione dei valutatori e test pilota. Questi nuovi test sono stati impiegati presso il Graduate Institute of Translation and Interpretation Studies dell'Università Cattolica di Fu Jen e il risultato è stato che il 91% degli studenti selezionati ha concluso positivamente il corso. Arjona-Tseng rileva la necessità di test attitudinali validi e attendibili per l'ammissione a corsi di formazione d'interpretariato.

Sebbene siano le questioni relative ai test attitudinali a monopolizzare la letteratura esistente sulla valutazione dell'interpretazione, esiste un esiguo numero di test sugli esami di accreditamento o certificazione per il riconoscimento professionale. Buona parte di questi esami sono condotti nell'ambito generale dell'interpretariato di comunità o più specificatamente in quello dell'interpretariato in campo giuridico/giudiziario. Sono pochi i paesi che formano interpreti specializzati in interpretariato di comunità o che usano i corsi universitari come unica via d'accesso alla professione d'interprete.

On the contrary, however, entry to the conference interpreting profession normally depends on successful completion of a university course.

The National Accreditation Authority for Translators and Interpreters is the accrediting body in Australia. Although there is accreditation for conference interpreting, examinations are not available for this skill, and conference interpreters gain accreditation on the basis of recognition of qualifications. The bulk of examinations is at the Professional level (formerly Level Three) and the Paraprofessional level (formerly Level Two). Courses in Australia that are approved by NAATI must adhere to NAATI guidelines when conducting their students' final examinations, which must reflect the NAATI format, content, and assessment criteria. Bell (1997: 98) describes NAATI examinations as «skills-based (performance assessments)». The Paraprofessional examination contains two dialogues of approximately 300 words in length each, and four questions on ethics of the profession and sociocultural aspects of interpreting. These examinations aim to assess the candidates' ability to practice as «paraprofessional» interpreters, mainly in the areas of welfare and education. The Professional interpreter examination comprises two dialogues of approximately 450 words each in length, with questions on ethics of the profession and sociocultural aspects of interpreting, and two 300-30 word passages, normally speeches, to be used for consecutive interpretation. These examinations aim to accredit interpreters to work in all areas of community interpreting, including medical and legal settings.

Candidates must pass each component with a minimum seventy marks out of one hundred, although, because of the error deduction marking scheme used, this cannot be interpreted as a percentage (cf. the discussion of error deduction marking earlier in this chapter). All examinations are marked by two examiners using NAATI's marking guidelines, which allow a good deal of subjective latitude. When discussing issues of accreditation for community interpreters, Gentile (1997) makes the point that evaluation criteria are usually vague, with specific meanings being left to the interpretation of each individual. He also comments on the difficulty of achieving standardisation across language pairs.

Per contro, per accedere alla carriera professionale di interprete di conferenza è in genere necessario portare a termine un corso universitario.

La National Accreditation Authority for Translators and Interpreters è l'ente preposto all'accREDITAMENTO in Australia. Sebbene l'accREDITAMENTO esista per l'interpretariato di conferenza, non esistono esami per questa specializzazione; per questo motivo gli interpreti di conferenza acquisiscono crediti sulla base del riconoscimento delle loro qualifiche. La maggior parte degli esami avvengono al livello *Professional* (in passato *Level Three*) e al livello *Paraprofessional* (in passato *Level Two*). In Australia, gli esami conclusivi dei corsi approvati dalla NAATI devono conformarsi alle linee guida NAATI per quanto riguarda la struttura, il contenuto e i criteri di valutazione impiegati. Bell (1997: 98) descrive gli esami NAATI come «(valutazioni del rendimento) basate sulle capacità». L'esame al livello *Paraprofessional* prevede due dialoghi di circa 300 parole ciascuno e quattro quesiti sull'etica della professione e sugli aspetti socioculturali dell'interpretariato.

Questi esami mirano a valutare la capacità dei candidati di lavorare come interpreti «paraprofessionali», segnatamente nell'ambito dell'assistenza sociale e in quello accademico. Gli esami al livello *Professional* prevedono due dialoghi di circa 450 parole ciascuno, con quesiti sugli aspetti etici e socioculturali della professione e l'interpretazione consecutiva di due brani, in genere discorsi, di 300-30 parole. Tali esami sono finalizzati a fornire agli interpreti la qualifica necessaria per poter lavorare in tutti gli ambiti dell'interpretariato di comunità, compresi i settori medico/sanitario e giuridico/giudiziario.

I candidati devono ottenere come minimo una votazione di 70/100, sebbene questa non vada letta in termini percentuali per via del metodo di deduzione dei punti utilizzato (cfr. la discussione sulla sottrazione dei punti già presentata in questo capitolo). Due esaminatori assegnano una votazione a tutti gli esami basandosi sulle linee guida della NAATI inerenti alla valutazione, che comunque permettono un'ampia discrezionalità personale. In merito alle questioni relative all'accREDITAMENTO per gli interpreti di comunità, Gentile (1997) spiega che i criteri di valutazione sono di solito vaghi, lasciando a ognuno il compito di interpretarne il significato specifico. Egli inoltre si esprime sulla difficoltà di raggiungere la standardizzazione tra coppie linguistiche.

These examinations have never been systematically scrutinised from the point of view of validity and reliability, although Bell states that: «In order to conduct valid and reliable tests, NAATI contracts more than 250 examiners on 46 different Examiners' Panels...In order to keep the examinations relevant to the development of the profession and the requirements of the employers, NAATI consults regularly with related individuals and organizations» (1997: 98). Our assumption is that these measures are intended to generate debate between the profession and NAATI which will help it improve the general quality of its tests; but this is of course a far cry from systematic scrutiny of the testing regime. The validity of the examinations has been questioned by Dueñas Gonzáles, who criticised their capacity to assess the skills and competencies required by court interpreters, stating that:

the test should not be used to examine court interpreters for three reasons: (1) it does not reflect the rigorous demands of the three modes used in judicial interpreting: simultaneous (unseen or spontaneous), legal consecutive and sight translation; (2) it does not test for mastery on all the linguistic registers encountered in the legal context,...and (3) it would not be a valid instrument to determine ability in judicial interpretation because its format, content and assessment methods are not sufficiently refined to measure the unique elements of court interpreting.

(Dueñas Gonzáles *et al.*, 1991: 91)

Anecdotal evidence shows that most practitioners are also dissatisfied with the tests' validity in other areas of community interpreting, especially with regard to the long consecutive passages which do not reflect the practice. In response to such criticism, NAATI is currently conducting a complete review of its examinations, the results of which will not be available for some time.

Unfortunately we were unable to access any literature on the California Court certification examination and cannot report on it. Such information would have allowed for a useful comparison with accreditation/certification examinations in other countries.

The court interpreter examination conducted by the Ministry of Justice in Spain comprises two main components: the translation of two texts, one into each language, without the use of dictionaries and with a one-hour time limit.

La validità e l'attendibilità di questi esami non sono mai state valutate in maniera attenta e sistematica, sebbene Bell affermi che: «al fine di condurre test validi e affidabili, la NAATI assume a contratto più di 250 esaminatori provenienti da 46 Panel di esaminatori differenti...Allo scopo di garantire che gli esami siano sempre al passo con l'evolversi della professione e con i requisiti imposti dai datori di lavoro, la NAATI consulta regolarmente i singoli e le organizzazioni ad essa affiliati» (1997: 98). Presumiamo che tali misure abbiano lo scopo di instaurare un dialogo tra il mondo professionale e la NAATI utile a migliorare la qualità generale dei suoi test; ma, com'è ovvio, siamo ancora decisamente lontani da un esame approfondito delle procedure dei test. A mettere in dubbio la validità degli esami è stato Dueñas Gonzáles, secondo cui non permetterebbero di valutare le capacità e le competenze richieste agli interpreti giurati:

il test non deve essere impiegato per esaminare interpreti giurati per tre ragioni: (1) non soddisfa i rigorosi requisiti delle tre modalità impiegate nell'interpretariato giuridico: la traduzione simultanea (*chuchotage* o *instant translation*), consecutiva di carattere giuridico e a vista; (2) non verifica la completa padronanza di tutti i registri linguistici presenti nel contesto giuridico,...e (3) non sarebbe un valido strumento per stabilire l'abilità nell'interpretariato giuridico poiché la struttura, il contenuto e i metodi di valutazione non consentono di giudicare con sufficiente precisione gli elementi peculiari dell'interpretariato giuridico.

(Dueñas Gonzáles *et alia*, 1991: 91)

L'evidenza aneddotica rivela come la maggior parte dei professionisti non siano soddisfatti della validità dei test in altri ambiti dell'interpretariato di comunità, soprattutto in riferimento ai lunghi brani di consecutiva che non trovano riscontro nella realtà. In risposta a queste critiche la NAATI sta attualmente conducendo una completa revisione dei suoi esami e i risultati di tale operazione non sono ancora disponibili.

Sfortunatamente non abbiamo avuto modo di consultare alcun testo inerente all'esame di certificazione per gli interpreti giurati californiani e non possiamo riferirvi nulla in merito. Peccato, perché sarebbe stato interessante e utile confrontare questi dati con gli esami di accreditamento/certificazione in altri paesi.

L'esame per interpreti giurati condotto dal Ministero di Giustizia spagnolo si compone di due parti principali: la traduzione di due testi, uno verso ciascuna lingua, senza l'aiuto di dizionari e con un limite tempo di un'ora.

Those who pass this phase with at least 50% can take the second component, a one-hour written examination on the government, the Ministry of Justice, the court system, and the laws and regulations surrounding workers' rights. There is no examination of any interpreting skill whatsoever, or of interpreter role or ethics. The only prerequisite for sitting the examination is a secondary school certificate. Miguélez strongly criticises this examination on the basis of lack of reliability and validity (Miguélez, 1999: 2). The certification examination which sworn interpreters take has currently been modified. The old examination consisted of two timed translations into Spanish. The first translation exercise is eliminatory and consists of texts ranging from 299-500 words in length, taken from magazines or newspapers and with no standard guidelines on level of difficulty. The text for the second exercise is always on a legal or economic/commercial topic, with a length ranging from 472-794 words. Two hours are allocated per exercise. Once again, Miguélez criticises this examination, making the observation that «it is reasonable to think that the same candidate sitting for different versions of the exam could get very different results» (1993: 3).

The new certification examination does not improve much on the old one. It maintains the translation exercises as described above and adds two components: a translation from Spanish and an oral exercise, where the candidate reads a text in the foreign language and then summarises it and answers questions on it to a panel of examiners. Miguélez attacks the new examination by stating that «the most obvious problem with this new test format is that it does not in any way test a candidate's ability to translate a legal document into the language of certification or to interpret in any of the three modes. The exam...lacks even the most basic standards of validity and reliability» (Miguélez, 1999: 4).

Nicholson and Martinsen (1997) describe the examination used in Denmark for interpreters to become members of the Authorized Interpreters Panel, approved by the National Commission of the Danish Police. Candidates must either possess a degree in a foreign language or be a native speaker of a foreign language. The only testing conducted is an oral test to assess the candidate's knowledge of Danish. The other language is not tested nor are any interpreting skills (1997: 262-3).

If little has been written on interpreting assessment in general, even less is found on any type of assessment as part of training courses.

Coloro che superano questa fase con almeno il 50% passano alla seconda fase, ovvero un esame scritto della durata di un'ora sul governo, il Ministero di Giustizia, il sistema giuridico, le leggi e i regolamenti a tutela dei diritti dei lavoratori. Non sono previsti esami sulle capacità interpretative o sul ruolo o l'etica dell'interprete. L'unico requisito necessario per accedere all'esame è un diploma di scuola secondaria. Miguélez critica pesantemente questo tipo di esame in quanto mancherebbe di attendibilità e validità (Miguélez, 1999: 2). L'esame di certificazione che gli interpreti giurati sostengono è attualmente in fase di modifica. Il vecchio esame consisteva nella realizzazione di due traduzioni a tempo verso lo spagnolo. Il primo esercizio di traduzione ha valore eliminatorio e consta di un testo che varia dalle 299 alle 500 parole, tratto da riviste o quotidiani e che non deve rispettare linee guida prestabili riguardo al grado di difficoltà. Il testo del secondo esercizio, la cui lunghezza varia dalle 472 alle 794 parole, tratta sempre argomenti di carattere economico o commerciale. Si hanno a disposizione due ore di tempo per svolgere ciascun esercizio. Ancora una volta Miguélez critica questa prova facendo notare che «è ragionevole poter pensare che lo stesso candidato, sostenendo diverse versioni dell'esame possa ottenere risultati anche molto discrepanti» (1993: 3).

Il nuovo esame di certificazione non è poi molto migliorato rispetto al precedente. Gli esercizi di traduzione appena descritti non cambiano, ma sono state aggiunte due prove: una traduzione dallo spagnolo e un esercizio orale in cui il candidato legge un testo in una lingua straniera e poi lo riassume e risponde ad alcuni quesiti su di esso di fronte ad un panel di esaminatori. Miguélez contesta questo tipo d'esame sostenendo che «il problema più evidente del nuovo tipo di test è che non permette di valutare in alcun modo la capacità del candidato di tradurre un documento legale nella lingua in cui deve ottenere la certificazione né d'interpretare in una delle tre modalità. L'esame...non soddisfa neppure gli standard di base quanto a validità e attendibilità» (Miguélez, 1999: 4).

Nicholson e Martinsen (1997) descrivono il tipo d'esame che, in Danimarca, gli interpreti devono superare per entrare a far parte dell'Authorized Interpreters Panel, approvato dalla National Commission of the Danish Police. I candidati devono essere laureati in una lingua straniera oppure madrelingua stranieri. L'unica prova condotta è un test orale che mira a saggiare la conoscenza della lingua danese da parte del candidato. Non sono valutate né la conoscenza dell'altra lingua né le capacità interpretative. (1997: 128).

Se si è scritto poco sulla valutazione dell'interpretazione in generale, ancora meno esiste in merito a qualsiasi tipo di valutazione nei corsi di formazione.

Macintosh points out, however, that although there seems to be little published on assessment systems, performance measurement is an area that has long been recognised as in need of systematic study: «some courses (e.g. ETI Geneva) have developed comprehensive and detailed marking schemes for final examinations, which attach different weightings to different components of a candidate's performance» (1995: 128). This may very well be so, and a survey of assessment procedures used by interpreting courses worldwide might produce very interesting results.

We have already explained that, in Australia, training courses that are NAATI approved must adhere to NAATI guidelines. Hence the description of the NAATI accreditation examination also applies to the final examinations conducted in educational programmes (units taken prior to final examinations are not assessed under NAATI guidelines). Gerver *et al.* make a brief mention of the final examination for the conference interpreting course at the former Polytechnic of Central London, mentioning that it tests for consecutive and simultaneous interpreting skills. Longley (1978) mentions that they use professional interpreters as raters in their London six-month intensive conference interpreting course. Longley makes one interesting observation about the difference between intuitive marking and more systematic marking. As part of a government funded course conducted by her institution, weighted marks were requested for specific types of errors. The raters had made an intuitive assessment of each candidate's performance at the end of the examination and were then faced with the time-consuming task of allocating marks for each specific component, or deducting marks for each type or error. Surprisingly, the results were very similar under both systems (1978: 54).

Schjoldager (1995) provides us with a marking sheet to assess simultaneous interpreting, which can be used by interpreters and students to self evaluate their performance, as well as by interpreter trainers. The sheet provides a set of criteria under four major categories: Comprehensibility and delivery, Language, Coherence and plausibility and Loyalty, with arguments and examples for each criterion. Schjoldager states that her «intention is merely to offer an explicit, systematic alternative to intuitive assessment procedures, whose criteria are not only implicit but also, I feel, arbitrary. Only explicit criteria can be useful to learners» (1995: 194).

Macintosh fa comunque notare che, sebbene la letteratura sui sistemi di valutazione sia piuttosto scarna, da lungo tempo si avverte la necessità di uno studio sistematico nel campo della misurazione della prestazione: «in alcuni corsi (come ad esempio all'ETI di Ginevra) sono state sviluppate tabelle di valutazione complete e dettagliate per gli esami conclusivi, nelle quali viene dato peso diverso ai vari aspetti della prestazione del candidato» (1995: 128).

Potrebbe benissimo essere così e un'indagine condotta sulle procedure di valutazione usate nei corsi d'interpretariato in tutto il mondo potrebbe fornire risultati estremamente interessanti.

Abbiamo già spiegato come, in Australia, i corsi di formazione riconosciuti dalla NAATI debbano conformarsi alle linee guida NAATI. Pertanto la descrizione degli esami di accreditamento della NAATI vale anche per gli esami finali condotti nei programmi accademici (le unità precedenti ai test conclusivi non seguono il modello valutativo NAATI). Gerver *et alia* fanno un breve accenno alla prova conclusiva del corso di interpretariato di conferenza dell'ex Polytechnic of Central London, dicendo che valuta le capacità interpretative sia in simultanea sia in consecutiva. Longley (1978) fa notare che nel loro corso intensivo d'interpretariato di conferenza della durata di sei mesi a Londra si servono di interpreti professionisti come valutatori. Longley inoltre compie un'interessante osservazione in merito alla differenza tra una valutazione intuitiva e una valutazione più sistematica. Nell'ambito di un corso finanziato dallo stato e diretto dalla sua istituzione, i voti ponderati erano necessari per alcuni tipi di errore. I valutatori, al termine della prova, esprimevano una valutazione intuitiva della prestazione di ciascun candidato e si trovavano poi a dover svolgere il lungo lavoro di assegnazione dei voti per ogni specifico elemento o di sottrazione dei punti per ciascun tipo di errore. I risultati ottenuti con entrambi i sistemi erano sorprendentemente molto simili (1978: 54).

Schjoldager (1995) ha elaborato una tabella di valutazione per l'interpretazione simultanea, utile sia per gli interpreti e gli studenti che desiderano autovalutare la propria prestazione sia per i docenti di interpretazione. La tabella stabilisce una serie di criteri inseriti in quattro categorie principali: Comprensibilità e resa, Linguaggio, Coerenza e plausibilità e Fedeltà. Ogni criterio è accompagnato da argomentazioni ed esempi. Schjoldager afferma di «voler semplicemente fornire un'alternativa esplicita e sistematica alle procedure di valutazione intuitive, contraddistinte da criteri, a mio parere, non soltanto impliciti ma anche arbitrari. Solo dei criteri espliciti possono essere utili agli studenti» (1995: 194).

## **Knowledge Gaps in Translation and Interpreting Assessment**

It will be evident from comparing our checklist with our survey that there exists a number of knowledge gaps in translation and interpreting assessment. In this section we briefly mention some of the less crucial gaps before a somewhat lengthier discussion of a fundamental omission in the literature – reliability. We will argue that this issue above all is in need of serious work.

The first four items on our checklist are reasonably well covered in the literature, at least as far as the traditional modes of interpreting and translation are concerned; assessment in newer or more peripheral modes of work such as interpreting in mental health settings, software localisation, and multilingual advertising has barely been discussed. Nevertheless, we have a fair understanding of the state of the art in the domains of the areas and purpose of assessment, the competencies assessed, and the forms of assessment. There are, however, differing amounts of emphasis with, for example, a preponderance of work on aptitude testing for interpreting, and a spread of work across achievement and accreditation testing in translation. Generally speaking, there is some agreement on the sets of competencies assessed in both translation and interpreting, but little explicit discussion of the efficacy of particular assessment instruments to measure those competencies. The basic forms of both translation and interpreting tests reflect a philosophy that the tests should resemble the real-world task, although in conference interpreting aptitude testing there are attempts to separately measure underlying competencies.

The fifth item – the basic approach – is rarely if ever explicitly discussed, but there seems to be tacit adoption of a criterion-referenced approach (although with no solid discussion of the actual criteria). The next three items – types of results, discrimination, and reporting mechanisms – are only minimally discussed.

These less crucial issues contrast starkly with the paucity of discussion on the central topics of validity and reliability. The knowledge gap in these areas is so large that we can do no more here than sketch the problem.

## Lacune conoscitive nella Valutazione della Traduzione e dell'Interpretazione

Mettendo a confronto la nostra lista iniziale con la ricerca condotta noterete chiaramente la presenza di alcune lacune conoscitive nell'ambito della valutazione della traduzione e dell'interpretazione. In questo paragrafo tratteremo brevemente di alcuni dei gap meno gravi prima di affrontare un discorso più approfondito relativo a una grande mancanza della letteratura analizzata – l'attendibilità. Ed è soprattutto su questo aspetto che secondo noi bisognerà condurre un serio lavoro.

I primi quattro punti della nostra lista sono trattati con sufficiente completezza nei testi esaminati, per lo meno per quanto concerne le tradizionali modalità di traduzione e interpretazione; in effetti, non si discute quasi per nulla di valutazione in modalità lavorative più recenti o secondarie come l'interpretariato nell'ambito della salute mentale, della localizzazione di software, della pubblicità in più lingue. Nonostante ciò abbiamo un quadro generale abbastanza chiaro dello stato dell'arte nell'ambito delle aree e dello scopo della valutazione, delle competenze valutate nonché delle forme di valutazione. Tuttavia ad alcuni temi è riservata una maggiore attenzione rispetto ad altri, ad esempio si parla molto di più di test attitudinali di interpretariato che non dei test conclusivi o di accreditamento in traduzione. In generale, possiamo dire che vi è un certo accordo sui gruppi di competenze da valutare sia in traduzione sia in interpretazione, ma non si discute quasi per nulla dell'efficacia di un particolare strumento di valutazione nel giudicare tali competenze. Le forme base dei test d'interpretariato e di traduzione rispecchiano la concezione che il test debba riprodurre una situazione di lavoro reale, sebbene nei test attitudinali per l'interpretariato di conferenza si cerchi di valutare le competenze intrinseche in modo separato.

Il quinto punto – l'approccio di base – non è quasi mai trattato in maniera esplicita, ma sembra esserci una tacita preferenza per l'approccio basato su criteri (sebbene manchi un'analisi esaustiva dei criteri veri e propri). I tre punti successivi – tipo di risultati, differenziazione e meccanismi relazionali – sono affrontati in modo superficiale.

Questi aspetti meno rilevanti si pongono in netto contrasto con lo scarso approfondimento delle questioni cruciali di validità e attendibilità. In questi ambiti il gap conoscitivo è talmente ampio che non possiamo fare altro che accennare il problema.

Indeed we will say very little at all about validity given that the consensus in measurement and evaluation circles is that tests cannot be valid unless they (or more accurately their scores) are reliable. Validity in interpreting and translation testing is tied up with knotty issues such as the nature of the competencies assessed, the models of learning underpinning educational programmes, and the extent to which tests should reflect professional tasks. Reliability stands out as the priority problem, and we devote the remainder of this section to a sketch of what we see as the main issues.

While reliability is extensively discussed in standard manuals on educational measurement, we have drawn on Bachmann (1991) to frame our discussion given that this work on language testing is a little closer to home than more general works.

According to Bachmann:

The investigation of reliability is concerned with answering the question, «How much of an individual's test performance is due to measurement error, or to factors other than the language ability we want to measure?» and with minimizing the effects of these factors on test scores. (Bachmann, 1991: 163)

These factors can be grouped into «test method facets», «attributes of the test taker that are not considered part of the language capabilities that we want to measure», and «random factors that are largely unpredictable and temporary» (1991: 164). Given that the latter two groups apply to tests of any kind, we will focus on «test method facets» as criteria affecting the reliability of interpreting and translation assessment. Chapter 5 of Bachmann (1991) is dedicated to test methods, and the summary of test method facets on page 119 could, we feel, be adapted to the interpreting and translation context. For example, *explicitness of criteria for correctness* resonates with the frequent query from translation and interpreting examinees about fidelity to the source text (ST); how closely, one is often asked, do I need to stick to the original? An inexplicit translation test instruction could affect the reliability of the test if one candidate believes that the target text must owe its loyalty to the ST rather than the target reader, while another candidate believes the opposite.

In effetti, parleremo molto poco di validità dato che gli esperti di misurazione e valutazione comunemente ritengono che i test non possono essere validi a meno che questi (e i loro punteggi nello specifico) non siano attendibili. Per i test di traduzione e interpretazione la validità è legata a questioni spinose quali la natura delle competenze da valutare, i modelli di apprendimento alla base dei programmi accademici e il grado di somiglianza dei test con le reali situazioni professionali. L'attendibilità rappresenta il problema principale, pertanto nel resto del paragrafo cercheremo di delineare quelli che noi riteniamo gli aspetti più importanti della questione.

Sebbene il tema dell'attendibilità sia ampiamente trattato nei manuali di base sulla misurazione accademica, abbiamo fatto riferimento a Bachmann (1991) per delineare il nostro discorso dato che il suo lavoro sulla valutazione linguistica è un po' più vicino al nostro campo d'indagine rispetto ad altri testi più generali.

Secondo Bachmann:

Lo studio dell'affidabilità non può prescindere dalla risposta a questa domanda, «in che misura il rendimento di un individuo in un test è dovuto all'errore di misurazione, o a fattori diversi dall'abilità che si vuole valutare?», come pure dall'attenuazione, per quanto possibile, degli effetti di tali fattori sui punteggi dei test. (Bachmann, 1991: 163)

È possibile raggruppare questi fattori negli «aspetti metodologici del test», «gli attributi dei candidati che non fanno parte delle capacità linguistiche che vogliamo giudicare», e «i fattori accidentali spesso imprevedibili e temporanei» (1991: 164). Dato che gli ultimi due gruppi sono validi per test di qualsiasi tipo, ci concentreremo sugli «aspetti metodologici del test» in quanto criteri che influiscono sull'affidabilità della valutazione nella traduzione e nell'interpretazione. Il quinto capitolo dell'opera di Bachmann (1991) è dedicato ai metodi dei test, e riteniamo che il riassunto degli aspetti metodologici del test presente a pagina 119 possa essere applicato al contesto dell'interpretazione e della traduzione. Ad esempio, la necessità di *criteri chiari per una resa corretta* trova riscontro nei dubbi che spesso nutrono i candidati interpreti o traduttori riguardo alla fedeltà al prototesto; spesso ci si domanda, in che misura bisogna restare fedeli all'originale? La mancanza di istruzioni chiare ed esplicite in un test di traduzione può influire sull'attendibilità del test se un candidato ritiene di dover restare fedele all'originale piuttosto che al lettore del metatesto mentre un altro candidato compie il ragionamento opposto.

*Degree of speededness* is highly relevant: when we impose a time limit on a test, do we know from empirical investigation the extent to which the speededness affects performance quality? Is there a speed at which we will get the optimum performance from the majority of candidates, and therefore have an optimally reliable test (at least on this facet)?

For interpreting and translation, a very significant test method facet is the degree of difficulty of the source material. Despite some inroads into the question of translation text difficulty (Campbell, 1999; Campbell & Hale, 1999), this remains a major barrier to improving test reliability. We would assert that in the absence of convincing methods for assessing ST difficulty, any testing regime that regularly introduces fresh STs and passages (for example, for security reasons) will potentially generate highly unreliable scores.

A basic concept in considering reliability is *parallel tests* (Bachmann, 1991: 168), from which can be derived a «definition of reliability as the correlation between the observed scores on two parallel tests». In other words, the most reliable test is one where parallel versions yield the same scores (i.e. a perfect correlation). In translation, this would involve finding or composing two examination texts of exactly the same degree of complexity in lexis, grammar, content, style and rhetorical structure. The lack of any real discussion of even this most basic measure of test reliability is a serious indictment of the present state of translation assessment. While occasional statements of intent are made (for example, Bell, 1997), we know of no serious work on basic questions such as the reliability of translation test scores over time, from language to language, or from text to text. Campbell (1991) makes a preliminary foray into the discriminatory power of items in translation tests in an attempt to launch a discussion about the internal consistency of such tests.

Much work, then, needs to be done. Again, we rely on Bachmann to frame the following discussion, highlighting some of the specific problems encountered in assessing translation and interpreting.

Il *grado di rapidità* è determinante; se viene stabilito un tempo massimo per un test, è possibile comprendere da un esame empirico in che misura la velocità influisca sulla qualità della prestazione? Esiste una velocità alla quale sarà possibile ottenere dalla maggior parte dei candidati una prestazione ottimale e pertanto un test pienamente attendibile (per lo meno da questo punto di vista)?

Per quanto riguarda la traduzione e l'interpretazione, un aspetto metodologico del test estremamente significativo è il grado di difficoltà del materiale originale. Sebbene qualche approfondimento sul tema della difficoltà di traduzione dei testi sia stato compiuto (Campbell, 1999; Campbell & Hale, 1999), questo aspetto rimane uno dei principali ostacoli alla creazione di test più attendibili. Potremmo dire che in mancanza di metodi convincenti per stabilire la difficoltà del prototesto, qualsiasi sistema che regolarmente introduca nuovi prototesti e brani (ad esempio per motivi di sicurezza) rischierà di fornire punteggi altamente inattendibili.

Uno dei concetti chiave legati all'attendibilità è il *test parallelo* (Bachmann, 1991: 168); l'attendibilità viene così ad essere presentata come «la correlazione tra i punteggi di due test paralleli». In altri termini, il test più attendibile è quello in cui versioni parallele forniscono gli stessi punteggi (cioè una correlazione perfetta). Allo stesso modo per la traduzione bisognerebbe trovare e assemblare due testi d'esame caratterizzati da uno stesso grado di complessità del lessico, della grammatica, del contenuto, dello stile e della struttura retorica. Il fatto che non si discuta nemmeno di questo basilare metodo di calcolo dell'attendibilità dei test la dice lunga sull'attuale situazione della valutazione della traduzione. Se occasionalmente vengono pronunciate dichiarazioni d'intenti in merito (si veda Bell, 1997), non siamo a conoscenza di alcun testo che si occupi seriamente di questioni basilari quali l'attendibilità dei punteggi dei test di traduzione nel tempo, da lingua a lingua e da testo a testo. Campbell (1991) compie un'iniziale incursione nell'ambito del potere discriminante delle voci nei test di traduzione con l'intento di promuovere un dibattito sulla coerenza intrinseca di tali test.

In breve c'è ancora molto da lavorare. Ancora una volta ci rifacciamo a Bachmann per presentare il prossimo argomento nel quale cercheremo di mettere in luce alcuni dei problemi specifici riscontrati nella valutazione della traduzione e dell'interpretazione.

## Internal consistency

If we assume that the basic test format is to translate or interpret, then investigation is needed into the way that candidates perform on different parts of the written or spoken input, and the extent to which those parts may be differentially weighted. A simple example is that of repeated material in a written or spoken passage. How, for instance, do we deal with passages with repeated chunks (for example, formulaic expressions introducing clauses in a treaty)? The implications for test reliability are profound: if a candidate mistranslates a repeated chunk, do we penalize multiple times? This is a common dilemma in translation test marking that goes to the heart of reliability because it may be argued that the candidate's performance could have been more reliably measured if he or she had been given a chance to be tested on a number of *different* items; the repeats may be interpreted as a test method facet that diminishes the discriminatory power of the test and therefore reduces its reliability. On the other hand, the repetitions may call for a creative solution that draws out the competence of the candidate. Arabic, for example, often employs a degree of parallelism that is not tolerated in English, and we might reward the candidate who manages to convey the rhetorical effect through a more natural English device. Internal consistency is also an issue tied up with text development and is particularly critical when we try to construct parallel tests. Let us say that we want to base a test on a 1000-word press article, using, say 500 words. In the first 250 words the writer is likely to be laying the groundwork for his or her argument, perhaps using irony or humour. The next 500 words may contain detailed exposition based on a technical account of the issue, and the last 250 a concluding summary that picks up the rhetorical flavour of the introduction, or even introduces a new note of warning. While it would be tempting to think that the most efficient way to create parallel tests is to cut one text into two, it is obvious that in this example neither half would reflect the rhetorical structure of the other and thus both would have different internal consistency.

## Coerenza intrinseca

Dando per scontato che la struttura base del test preveda un lavoro di traduzione/interpretazione, è necessario studiare come i candidati si disimpegnano nelle diverse parti del messaggio scritto od orale e in che misura queste parti possono essere valutate in maniera differente. Una semplice prova consiste nel riproporre più volte lo stesso messaggio in un brano scritto od orale. Come bisogna comportarsi, per esempio, in presenza di brani in cui compaiono delle ripetizioni (come accade nel caso di espressioni stereotipate che introducono le clausole di un trattato)? Ancora una volta emergono le profonde implicazioni dell'attendibilità dei test: il candidato che sbaglia la traduzione dell'elemento ripetuto va penalizzato una o più volte? Si tratta di un vero e proprio dilemma per tutti coloro che devono valutare dei test di traduzione, un dilemma che va dritto al cuore dell'attendibilità in quanto qualcuno potrebbe dire che la prestazione del candidato sarebbe stata giudicata in maniera più affidabile se questi avesse avuto la possibilità di essere giudicato su una serie di elementi *diversi*; le ripetizioni potrebbero essere interpretate come un aspetto metodologico del test che diminuisce il potere discriminatorio del test e di conseguenza ne riduce il grado di attendibilità. D'altro canto, però, la presenza di ripetizioni può rappresentare uno stimolo alla ricerca di soluzioni creative che possano rivelare le capacità e le conoscenze del candidato. La lingua araba, ad esempio, spesso si serve di un grado di parallelismo che non è accettato nell'inglese e potremmo quindi premiare il candidato che riesce a restituire l'effetto retorico per mezzo di un artificio che suona meglio in inglese. Il problema della coerenza intrinseca è anche strettamente correlato allo sviluppo del testo e si pone con maggiore urgenza nella costruzione di test paralleli. Ipotizziamo di voler basare il nostro test su un articolo di giornale di 1000 parole, usandone, diciamo, 500. Le prime 250 parole probabilmente serviranno all'autore per porre le basi della propria tesi, magari usando un tono ironico o umoristico. Le successive 500 parole potrebbero contenere una dettagliata esposizione basata su di un resoconto tecnico della questione e le ultime 250 una sintesi conclusiva che riprende il sapore retorico dell'introduzione o che introduce persino una nota d'ammonimento. Se da un canto sarebbe bello poter pensare che il modo migliore di creare dei test paralleli sia quello di dividere un testo in due metà, dall'altro, risulta chiaro che in quest'esempio nessuna delle due metà rifletterebe la stessa struttura retorica e pertanto avrebbero una diversa coerenza intrinseca.

## Estimating Reliability

Those lucky enough to use multiple choice and other brief response test item types have the luxury of measuring test reliability through *split-half* methods, where «we divide the test into two halves and then determine the extent to which scores on these two halves are consistent with each other» ( Bachmann, 1991: 172). The crucial requirement of split-half measures is that performance on one half must be independent of performance on the other half. Even if we could find ways to split interpreting and translation tests (for example, odd versus even paragraphs, first half versus second half), there is no way that the two halves can be independent; if they were, they would not constitute a text. Split-half methods appear, then, to be ruled out. An alternative approach – the Kuder-Richardson reliability coefficients – suffer the same fate for different reasons. The KR formulae are based on the means and variances of the items in the test, and assume that all items «are of nearly equal difficulty and independent of each other» (1991: 176); even if discrete items could be identified, the criteria of equal difficulty and independence would be impossible to achieve.

Indeed, interpreting and translation tests seem to have much more in common with open-ended instruments like essays, where statistical methods of estimating reliability on the basis of individual test items are extremely difficult to apply. The practice of «second markers», «trial marking», etc. indicates a focus on the marker rather than the items as a source of information about reliability. Bachmann speaks of intra- and inter-rater reliability (1991: 178-81). Estimates on intra-rater reliability are made by having a marker rate the same group of subjects twice – on two separate occasions and in different orders – and calculating a correlation coefficient of some kind. Anyone who has spent a day on an interpreting assessment jury or marking a pile of translation examinations will be aware of the potential shifts in rater behaviour through fatigue, or through recency effects as markedly different candidates present. Similarly, a correlation coefficient can be calculated to estimate how consistently two or more markers rate the same candidates.

## Valutare l'Attendibilità

Coloro che sono abbastanza fortunati da usare la scelta multipla o altri tipi di test a risposta breve si possono permettere il lusso di calcolare l'affidabilità dei test con metodi *split-half* ovvero di divisione a metà, nei quali «dividiamo il test in due parti e stabiliamo fino a che punto i punteggi di queste due metà siano coerenti tra loro» (Bachmann, 1991: 172). Il requisito fondamentale delle misurazioni *split-half* è che la prestazione raggiunta per la prima parte sia indipendente da quella della seconda parte. Anche se trovassimo un modo per suddividere i test di traduzione e interpretazione (ad esempio, paragrafi pari/dispari; prima parte/seconda parte), le due parti non potrebbero mai essere indipendenti; se lo fossero, non formerebbero mai un testo. Per questa ragione i metodi *split-half* sembrerebbero da scartare. Un approccio alternativo – i coefficienti di affidabilità di Kuder-Richardson – è altrettanto inutilizzabile per altre ragioni. Le formule KR si basano sulle medie e le varianze delle voci di un test e presuppongono che tutte le voci «siano all'incirca della stessa difficoltà e che siano indipendenti tra loro» (1991: 176); sebbene sia possibile ravvisare voci distinte, sarebbe impossibile soddisfare i criteri di pari difficoltà e indipendenza.

In effetti, i test di traduzione e interpretazione sembrano avere molti più aspetti in comune con gli strumenti di verifica aperti come i temi, ai quali è molto difficile poter applicare metodi statistici di valutazione dell'attendibilità sulla base delle singole voci del test. L'impiego di «indici secondari», della «valutazione di prova» ecc. indica che sono i valutatori e non gli aspetti del test ad essere ritenuti una fonte di informazioni sull'affidabilità. Bachmann fa un distinguo tra attendibilità dei valutatori e attendibilità tra valutatori (1991: 178-81). È possibile elaborare delle stime sull'attendibilità dei valutatori facendo loro giudicare lo stesso gruppo di soggetti due volte – in due momenti diversi e in ordine differente – e calcolando un coefficiente di correlazione di qualche tipo. Chiunque, almeno per un giorno, abbia fatto parte di una commissione di valutazione d'interpretazione o abbia dovuto valutare un gran numero di prove di traduzione sarà sicuramente a conoscenza dei possibili cambi di atteggiamento del valutatore dovuti alla stanchezza o all'avvicinarsi di candidati estremamente differenti tra di loro. Allo stesso modo, si può calcolare un coefficiente di correlazione per stimare in quale misura due o più valutatori sono coerenti nel giudicare gli stessi candidati.

Organisations like NAATI and American Translators Association (ATA) appear to depend heavily on intra- and inter-rater behaviour to achieve reliability. We can only guess at the extent to which educational institutions take rater reliability seriously in achievement tests, final examinations, and the like. It is somewhat surprising to note, then, that our sample of reading contained not a single major published study on the issue or rater consistency.

## **Concluding Remarks**

The translation and interpreting research world asks a great deal of itself. With major current research pushes in areas as diverse as cognitive processing, cultural studies, lexicography and machine translation, it is perhaps not surprising that the field of assessment is in its infancy. But assessment does need to grow up a little and realise that there are some bigger kids on the block for it to learn from; the wider field of measurement and evaluation represents a solid source of knowledge that we can use to understand and improve our assessment practice. It is not just a question of filling in the knowledge gaps, but a question of profession building. As an applied discipline, translation and interpreting puts people into real and important jobs; better assessment means better translators and interpreters.

### Note

1. The assistance of Adriana Weissen in undertaking the literature search is acknowledged.

Alcune organizzazioni come la NAATI e l'American Translator Association (ATA) sembrano dare molto peso al comportamento dei valutatori e tra i valutatori al fine di raggiungere l'affidabilità. Noi non possiamo fare altro che ipotizzare in quale misura delle istituzioni accademiche prendano sul serio l'attendibilità dei valutatori nei test di profitto, negli esami finali ecc. Tuttavia è abbastanza sorprendente il fatto che di tutte le letture prese in esame nessuna di esse contenga un importante studio pubblicato sul tema della coerenza degli valutatori.

## Note conclusive

Il mondo della ricerca nel campo della traduzione e dell'interpretazione nutre in se grandi aspettative. Ma visto che attualmente gli stimoli alla ricerca sono più forti nei campi più disparati, quali l'elaborazione cognitiva, i *cultural studies*, la lessicografia e la traduzione automatica forse non ci si deve sorprendere se il campo della valutazione stia muovendo solo ora i primi passi. Ma la valutazione deve assolutamente crescere e comprendere che ci sono discipline più evolute e mature dalle quali imparare; il più vasto campo della misurazione e della valutazione rappresenta una fonte consolidata di conoscenze che possiamo utilizzare per capire e migliorare i nostri metodi di valutazione. Non si tratta solo di colmare i gap conoscitivi, ma anche di crescere professionalmente. In quanto discipline applicate, la traduzione e l'interpretazione pongono le persone in contesti professionali reali e importanti; un miglior metodo di valutazione significa migliori traduttori e interpreti.

## Nota

1. Si riconosce l'aiuto di Adriana Weissen nella ricerca della letteratura presa in esame.

## Bibliografia

- Arjona-Tseng, E. (1994). A psychometric approach to the selection of translation and interpreting students in Taiwan. In S. Lambert and B. Moser-Mercer (eds). *Bridging the Gap. Empirical Research in Simultaneous Interpretation* (pp. 69-86). Amsterdam: John Benjamins.
- Bachmann, L. (1991) *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bastin, G.L. (2000) Evaluating beginners' re-expression and creativity: A positive approach. *The Translator* 6/2, 231-45.
- Bell, S. (1997) The challenges of setting and monitoring the standards of community interpreting: An Australian perspective. In S. Carr, R. Roberts, A. Dufour and D. Steyn (eds) *The Critical Link: Interpreters in the Community. Papers from the First International Conference on Interpreting in Legal, Health and Social Services Settings*. Amsterdam: John Benjamins.
- Bowen, D. and Bowen, M. (1989) Aptitude for interpreting. In L. Gran and J. Dodds (eds) *The Theoretical and Practical Aspects of Teaching Conference Interpretation* (pp. 109-25). Udine: Campanotto Editore.
- Bowker, L. (2000) A corpus-based approach to evaluating student translations. *The Translator* 6/2, 183-210.
- Brunette, L. (200) Towards a terminology for translation quality assessment: A comparison of TQA practices. *The Translator* 6/2, 169-82.
- Bühler, H. (1986) Linguistic (semantic) and extra-linguistic (pragmatic) criteria for the evaluation of conference interpretation and interpreters. *Multilingua* 4-5, 231-5.
- Campbell, S. (1991) Towards a model of translation competence. *Meta* 36/2-3, 329-43.
- Campbell, S. (1999) A cognitive approach to source text difficulty in translation. *Target* 11/1, 33-63.
- Campbell, S. and Hale, S. (1999) What makes a text difficult to translate? *Proceedings of the 1998 ALAA Congress*. <http://www.cltr.uq.edu.au/alaa/proceed/camphale.html> (19 April 1999).
- Cestac, F. (1987) The recruitment of translators/précis writers at the United Nations and quality control of translators. In M. Gaddis Rose (ed.) *Translation Excellence: Assessment Achievement Maintenance* (pp. 77-86). Binghamton, NY : SUNY.
- Dejean Lefeal, K. (1990) Some thoughts on the evaluation of simultaneous interpretation. In D. Bowen and M. Bowen (eds) *Interpreting Yesterday, Today and Tomorrow* (pp. 154-60). Binghamton, NY: SUNY.
- Dollerup, C. (1993) Systematic feedback in teaching translation. In C. Dollerup and A. Lindegaard (eds) *Teaching Translation and Interpreting 2. Insights, Aims, Visions, Papers from the Second Language International Conference* (pp. 121-32). Amsterdam/Philadelphia: John Benjamins.
- Duenas Gonzalez, R., Vásquez, V. and Mikkelson, H. (1991). *Fundamentals of Court Interpretation. Theory, Policy and Practice*. North Carolina: Carolina Academic Press.
- Ebel, R.L. (1972) *Essentials of Educational Measurement* (2<sup>nd</sup> edn). Englewood Cliffs, NJ: Prentice-Hall.
- Eckstrom, R.B., French, J.W., Harman, H.H., and Dermen, D. (1976) *Manual for Kit of Factor-Referenced Cognitive Tests*. Princeton, NJ: Educational Testing Service.
- Farahzad, F. (1992) Testing achievements in translation classes. In C. Dollerup and A. Loddengaard (eds) *Teaching Translation and Interpreting. Training, Talent and*

- Experience. Papers from the First Language International Conference* (pp. 271-8). Amsterdam/Philadelphia. John Benjamins.
- Furneaux, W.D. (1956) *Manual of Nufferno Speed Tests and Manual of Nufferno Level Tests*. London: Institute of Psychiatry.
- Gentile, A. (1997) Community interpreting or not? Practices, standards and accreditation. In S. Carr, R. Roberts, A. Dufour and D. Steyn (eds) *The Critical Link: Interpreters in the Community* (pp. 109-18). Amsterdam: John Benjamins.
- Gerver, D., Longley, P., Long, J. and Lambert, S. (1984) Selecting trainee conference interpreters: A preliminary study. *Journal of Occupational Psychology* 57, 17-31.
- Gerver, D., Longley, P., Long, J. and Lambert, S. (1989) Selection tests for trainee conference interpreters. *Meta* 34/4, 724-35.
- Hatim, B. and Mason, I. (1997) *The translator as Communicator*. Routledge: London.
- Hatim, B. and Williams, M. (1998) Course profile. Diploma in Translation (Diplôme de Traducteur). *The Translator* 4/1, 125-33.
- House, J. (1981) *A Model for Translation Quality Assessment*. Tübingen: Narr.
- Ivanova, A. (1998) Educating the "Language Elite". In K. Malmjaer (ed.) *Translation and Language Teaching. Language Teaching and Translation* (pp. 91-109). Manchester: St Jerome.
- James, H., Roffe, I. and Thorne, D. (1995) Assessment and skills in screen translation. In C. Dollerup and V. Appel (eds) *Teaching Translation and Interpreting 3. New Horizons. Papers from the Third Language International Conference* (pp. 271-8). Amsterdam/Philadelphia: John Benjamins.
- Kalina, S. (2001) Quality requirements in conference interpreting. In *The First International Conference on Translation and Interpretation Studies: Theories of Translation and Interpretation and Problems in Korean Translation and Interpretation* (pp. 19-31). Seoul: Hankuk University of Foreign Studies.
- Keiser, W. (1978) Selection and training of conference interpreters. In D. Gerver and H. Wallace Sinaiko (eds) *Language Interpretation and Communication* (pp. 11-24). New York and London: Plenum Press.
- Kintsch, W. (1974) *Representation of Meaning in Memory*. Hillsdale, NJ: Erlbaum.
- Kopczynski, A. (1992) Quality in conference interpreting: Some pragmatic problems. In M. Snell Hornby, F. Pöchhacker and K. Kaindl (eds) *Translation Studies. An Interdiscipline*. Amsterdam: John Benjamins.
- Kusmaul, P. (1995) *Training the Translator*. Amsterdam/Philadelphia: John Benjamins.
- Lambert, S. (1991) Aptitude testing for simultaneous interpretation at the university of Ottawa. *Meta* 36/4, 586-94.
- Longley, P. (1989) The use of aptitude testing in selection of students for conference interpreting training. In L. Gran and J. Dodds (eds) *The Theoretical and Practical Aspects of Teaching Conference Interpretation* (pp. 105-8). Udine: Campanotto Editore.
- Macintosh, J. (1995) A review of conference interpretation: Practice and Training. *Target* 7/1, 119-33.
- Maier, C. (2000) Introduction. *The Translator* 6/2, 137-48.
- Martin, P. (1994) NAATI: Role and functions. In R. Seymour and C.C. Liu (eds) *Translation and Interpreting: Bridging East and West* (pp. 23-32). Hawaii: University of Hawaii and the East-West Center.
- Miguélez, C. (1999) Current issues in court interpreting: Spain, a case study. *Proteus. Newsletter of the National Association of Judiciary Interpreters and Translators* 8/2, 1-4.

- Moser-Mercer, B. (1994) Aptitude testing for conference interpreting: Why, When and How. In S. Lambert and B. Moser-Mercer (eds) *Bridging the Gap: Empirical Research in Simultaneous Interpretation* (pp. 57-67), Amsterdam: John Benjamins.
- Niedzielski, H. and Chernovaty, L. (1993) Linguistic and technical preparation in the training of technical translators and interpreters. In S.E. Wright and L.D. Wright, jun. (eds) *Scientific and Technical Translation* (pp. 123-49). Amsterdam: John Benjamins.
- Ostarhild, E. (1994) The Institute of Linguists New Diploma in English and Chinese – factors influencing syllabus update. In R. Seymour and C.C. Liu (eds) *Translation and Interpreting: Bridging East and West* (pp. 51-4). Hawaii: University of Hawaii and the East-West Center.
- Ozolins, U. (1998) *Interpreting and Translation in Australia. Current Issues and International Comparisons*. Melbourne: Language Australia.
- Pöchhacker, F. (1993) Quality assurance in simultaneous interpreting. In C. Dollerup and A. Lindegaard (eds) *Teaching Translation and Interpreting 2. Insights, Aims, Visions. Papers from the Second Language International Conference* (pp. 233-42). Amsterdam/Philadelphia: John Benjamins.
- Sainz, M. (1993) Student-centred correction of translations. In C. Dollerup and A. Lindegaard (eds) *Teaching Translation and Interpreting 2. Insights, Aims, Visions. Papers from the Second Language International Conference* (pp. 133-41). Amsterdam/Philadelphia: John Benjamins.
- Schaeffner, C. (1998) Qualifications for professional translators: Translation in language teaching versus teaching translation. In K. Malmjaer (ed.) *Translation and Language Teaching. Language Teaching and Translation*. (pp. 117-33). Manchester: St Jerome.
- Schjoldager, A. (1995) Assessment of simultaneous interpreting. In C. Dollerup and V. Appel (eds) *Teaching Translation and Interpreting 3. New Horizons* (pp. 186-95). Amsterdam: John Benjamins.
- Schweda Nicholson, N. and Martinsen, B. (1997) Court interpretation in Denmark. In S. Carr, R. Roberts, A Dufour and D. Steyn (eds) *The Critical Link :Interpreters in the Community* (pp. 259-70). Amsterdam: John Benjamins.
- Teague, B. (1987) ATA accreditation and excellence in practice. In M. Gaddis Rose (ed.) *Translation Excellence: Assessment Achievement Maintenance* (pp. 21-6). Binghamton: SUNY.
- Thorndike, R.M., Cunningham, G.K., Thorndike, R.L. and Hagen, E.P. (1991) *Measurement and Evaluation in Psychology and Evaluation* (5<sup>th</sup> edn.). New York: Macmillan.
- Vidal, M. (2000) NAJIT Certification on the way. *Proteus. Newsletter of the National Association of Judiciary Interpreters and Translators* 9/3, 1-3.
- Wu, J. (1994) Task-oriented and comprehensive training of translators and interpreters. In R. Seymour and C.C. Liu (eds) *Translation and Interpreting: Bridging East and West* (pp. 87-95). Hawaii: University of Hawaii and the East-West Centre.