BENJAMINS CURRENT TOPICS 68

Aptitude for Interpreting

Franz Pöchhacker
and Minhua Liu (eds.)

Aptitude for Interpreting

Benjamins Current Topics

Special issues of established journals tend to circulate within the orbit of the subscribers of those journals. For the Benjamins Current Topics series a number of special issues of various journals have been selected containing salient topics of research with the aim of finding new audiences for topically interesting material, bringing such material to a wider readership in book format.

For an overview of all books published in this series, please see
http://benjamins.com/catalog/bct

Volume 68

Aptitude for Interpreting
Edited by Franz Pöchhacker and Minhua Liu

These materials were previously published in *Interpreting* 13:1 (2011) and 16:1 (2014)

Aptitude for Interpreting

Edited by

Franz Pöchhacker
University of Vienna

Minhua Liu
Monterey Institute of International Studies

John Benjamins Publishing Company
Amsterdam / Philadelphia

Library of Congress Cataloging-in-Publication Data

Aptitude for Interpreting / Edited by Franz Pöchhacker and Minhua Liu.
p. cm. (Benjamins Current Topics, ISSN 1874-0081 ; v. 68)
Includes bibliographical references and index.
1. Translating and interpreting--Study and teaching. I. Pöchhacker, Franz, editor. II. Liu, Minhua.
P306.5.A68 2015
418'.020711--dc23 2014022642
ISBN 978 90 272 4256 3 (Hb ; alk. paper)
ISBN 978 90 272 6954 6 (Eb)

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

Table of contents

INTRODUCTION

Aptitude for interpreting

The abilities and skills required for interpreting have been a topic of special interest ever since the very first scientific investigation into the professional occupation of conference interpreting by Jesus Sanz (1930). In the 1960s and 1970s, AIIC, the International Association of Conference Interpreters, made serious efforts to tackle this issue and come to a better understanding of the prerequisites for a career in professional interpreting (see Keiser 1978). Attempts to put the selection of candidates for interpreter training on a more scientific footing have been made since the 1980s, often drawing on insights from cognitive psychology (e.g. Moser-Mercer 1985). And yet, relatively little empirical research on aptitude for interpreting has been carried out to date, despite recurrent doubts over the reliability, validity and predictive power of tasks designed to test candidates for interpreter training programs (e.g. Dodds 1990). What is more, those studies that do exist do not yet provide us with sufficient guidance on the effective screening of applicants. This dearth of evidence-based methods of admission testing is all the more problematic when one considers the growing demand for highly qualified interpreters as training institutions face increasing pressure to make the most efficient and judicious use of their limited resources.

It was against this backdrop that an initiative to organize an international symposium on aptitude for interpreting took shape. Interpreting scholars in Europe and the US joined forces to convene a group of leading researchers and educators specializing in the study of aptitude testing for interpreter education. The international symposium, "Aptitude for Interpreting: Towards Reliable Admission Testing," was held on May 28–29, 2008, at Lessius University College in Antwerp, Belgium, now part of the KU Leuven Faculty of Arts. Based on a thorough screening process, the organizers — Heidi Salaets and Šárka Timarová of Lessius/KU Leuven, and Sherry Shaw of the University of North Florida — put together a program consisting of ten presentations, all of them centering on the enigmatic connection between means of evaluating aptitude(s) at the entry phase and students' prospects of success in ultimately joining the ranks of the interpreting profession. One of the admirable attributes of the event was its success in bringing together leading scholars from the signed-language interpreting community as well as their spoken-language counterparts.

Seven of the papers presented at Antwerp were made available as a Special Issue of *Interpreting* (13:1, 2011), co-edited by the late and greatly missed Miriam Shlesinger. Two of the scholars present at the Antwerp Symposium went on to publish further research on aptitude in *Interpreting* (16:1, 2014). These papers complement those in the Special Issue and are offered in the present volume together with the set of seven papers published three years earlier.

One can safely claim that the contributions brought together here represent the cutting edge of research into aptitude for interpreting, and the community of interpreting scholars and educators owes a debt of gratitude to the organizers of the landmark symposium in Antwerp. At the same time, the work collected in this volume points to the need for much further research, as, in many ways, the paucity of conclusive findings appears to persist.

Those interested in the state of the art in the field of aptitude testing will appreciate the highly productive integration of work done by interpreter educators, on the one hand, and insights gleaned from other disciplines, chief among them experimental psychology, on the other. Moreover, as reflected by the chapters in this volume, the role of personality traits in the composite profile of a likely candidate has gained prominence alongside the traditional focus on cognitive skills. No less important, the reader will come to see the many similarities (and few differences) between aptitude testing in the fields of spoken- and signed-language interpreting.

The volume opens with a sweeping chronological overview of the field, including a survey of those characteristics that a prospective interpreter is expected to possess and the means of measuring them, by **Mariachiara Russo**. Delivered as the keynote address at the Antwerp Symposium, the paper draws attention to the rather consistent perception of the ideal interpreter profile over the years and offers a succinct review of test designs that allow for the identification of statistically significant predictors of examination results and program completion rates. One point that emerges from this comprehensive presentation of the state of the art is the neglect in most entrance examinations of applicants' 'soft skills' such as motivation and other personality traits. These are effectively explored in some of the following chapters.

In a study which grew out of an ambitious project investigating both cognitive abilities and soft skills, **Šárka Timarová and Heidi Salaets** use a learning style inventory, a test of motivation and a cognitive flexibility task to explore the characteristics of undergraduate students who decide to enrol in an interpreter training program. The authors also investigate the relationship between students' successful completion of training, on the one hand, and their scores on each of these, on the other. They find that students who self-select for interpreting tend to show higher cognitive flexibility and achievement motivation and to be less susceptible to debilitating anxiety.

Anxiety levels and motivation also figure in the study by **Alexandra Rosiers, June Eyckmans and Daniel Bauwens**, which centers on the different profiles of translation and interpreting students. Based on standard tests of self-perceived linguistic competence, particularly oral skills, as well as psychological instruments testing anxiety and integrative motivation, the two groups of students are found to differ with regard to these individual traits. Nevertheless, going on to relate these differences to the students' interpreting performance as measured in a sight translation task, the authors do not find any significant correlations between personality and performance.

The two chapters focusing on spoken-language interpreting students, by authors based in Belgium, are followed by a multi-center study investigating soft skills in both spoken- and signed-language interpreting. Working with interpreting students at training programs in four countries of the European Union, **Sherry Shaw** administered a neurocognitive test battery as well as a psychological instrument measuring personality traits relevant to interpreting skill acquisition. Differences were found in both cognitive performance parameters and motivational personality traits of entry-level and advanced students of interpreting and between participants in the spoken- and signed-language programs.

Yet another aspect of personality, emotional stability, is investigated by **Karen Bontempo and Jemina Napier** in a study of accredited sign language interpreters in Australia. The authors used a questionnaire to measure self-efficacy, goal orientation and negative affectivity, all constructs for which valid and reliable psychometric tools are available and considered predictive of performance in a variety of contexts. While goal orientation was not significantly related with self-rated interpreter competence, and self-efficacy proved to be of only minor interest, negative affectivity correlated significantly with interpreters' own perceptions of their performance.

Following the four empirical chapters primarily concerned with soft skills and personality traits, the remaining four shift the focus to the role of cognitive skills, which have traditionally been foregrounded in the literature on aptitude. The first of these, by **Brooke Macnamara, Adam Moore, Judy Kegl and Andrew Conway**, takes a broad perspective on cognitive components, addressed under the heading of domain-general skills, in conjunction with aspects of personality. In a study conducted among sign language interpreters in the United States, the authors employ instruments designed to measure such cognitive abilities as processing speed, task switching and working memory, as well as certain personality measures. In their comparison of highly skilled and less-skilled professional interpreters, the authors find the two groups to differ in both dimensions. Among other things, the highly skilled interpreters exhibit greater mental flexibility, faster cognitive processing speed and less anxiety about taking risks, supporting the

assumption that both cognitive and personality factors play a role in determining interpreting skill.

Turning entirely to cognitive skills as a core part of aptitude for interpreting, the last three papers in this volume report work designed to examine the predictive value of specific aptitude tests within a given institutional framework. Following up on previous work reported in part in her introductory overview, **Mariachiara Russo** presents the results of a longitudinal study on the 'simultaneous' Italian-to-Italian paraphrase test taken by a total of 64 students as part of their entrance examination at the University of Bologna at Forlì. Two evaluation parameters for the paraphrasing test (synonyms and coherence) were correlated with students' average marks on their final interpreting exams and with the number of exam sessions required for successful completion. The findings show that the ability to use synonyms has considerable predictive validity in classifying students as 'slow' or 'fast' in successfully completing the program.

The ability to find synonyms, as an indicator of expressional fluency, also forms part of the SynCloze test proposed and put to the test by **Franz Pöchhacker** at the University of Vienna. As its name suggests, this test focuses on a cognitive skill that previous studies (e.g. Gerver et al. 1989) have shown to be related to interpreting — that is, the ability to anticipate meaning in context. Using an auditory task requiring sentence completion (cloze) with as many acceptable variants as possible under time pressure, Pöchhacker tested a total of 127 undergraduate students beginning an introductory interpreting course and related their scores to performance on a consecutive interpreting exam administered at the end of the course. His findings show that the SynCloze test discriminates between novices and advanced students of interpreting and that there is a correlation, albeit a moderate one, between the scores on the composite task and on the (intralingual) interpreting exam.

The SynCloze task as well as 'cognitive shadowing' and 'personalized cloze' were in turn put to the test in a longitudinal study by **Catherine Chabasse and Stephanie Kader** at the University of Mainz in Germersheim. Two dozen students with German as their A language and English, French or Italian as their B languages took the three booth-based tests early in their first year of the MA in Conference Interpreting, and their scores were correlated with the grades achieved in the intermediate exam at the end of the second semester. While significant correlations were found between the various tests and certain exam results, the cognitive shadowing task proposed by Chabasse was found to offer the greatest efficiency in scoring.

Such studies represent welcome advances toward more reliable and predictive aptitude testing for would-be interpreters. Beyond the cognitive component, the papers brought together in this volume highlight the complexity of the notion of

aptitude for interpreting and the diverse ways in which cognitive skills and personality traits complement each other. Notwithstanding the extensive efforts reflected in these research endeavors, much of what has been found calls for further investigation and deeper analysis. The insights provided in this book, which has emerged from the concerted efforts of researchers across the spoken- and signed-language interpreting communities, attest to the potential for further progress in this important and challenging aspect of the training and professionalization of interpreters.

Franz Pöchhacker and Minhua Liu

References

Dodds, J. M. (1990). On the aptitude of aptitude testing, *The Interpreters' Newsletter* 3, 17–22.

Gerver, D., Longley, P. E., Long, J. & Lambert, S. (1989). Selection tests for trainee conference interpreters. *Meta* 34 (4), 724–735.

Keiser, W. (1978). Selection and training of conference interpreters. In D. Gerver & H. W. Sinaiko (Eds.), *Language interpretation and communication*. New York: Plenum Press, 11–24.

Moser-Mercer, B. (1985). Screening potential interpreters. *Meta* 30 (1), 97–100.

Sanz, J. (1930). Le travail et les aptitudes des interpretes parlementaires. *Anals d'Orientació Professional* 4, 303–318.

Aptitude testing over the years*

Mariachiara Russo
SSLMIT, University of Bologna

In the present paper I review the existing literature on aptitude testing with a view to highlighting the main emerging themes: which qualities indicate an aptitude in a prospective interpreter, how these qualities may be measured and which types of test should be administered, the issue of valid and reliable testing, proposals for test designs, and, finally, description of aptitude tests which have identified statistically significant predictors. The focus is on spoken language, but signed-language aptitude testing is also partially covered. Available results so far appear to show that interpreting-related cognitive skills and verbal fluency may be measured and may be predictive both for spoken-language and for signed-language interpreting candidates. In particular, the production of synonyms appears to be a strong aptitude predictor from several independent research projects.

> Any aspect of the individual, including some matters untouched by conventional ability and personality measures, can predict response to instruction and hence can be a source of "aptitude". (Cronbach & Snow 1977: 6)

1. Introduction

Before any interpreter training institution was ever set up in the early 1940s and 1950s, it was talented and well educated bilingual or multilingual speakers who played the role of interpreters and became highly reputed professionals (see Delisle & Woodsworth 1995; Roland 1999; Baigorri Jalón 2000). The exceptional interpreting skills these individuals displayed led to the assumption that interpreters possess inborn qualities, an assumption that interpreting professionals, schools and researchers have been challenging over the years.

* The author is very grateful to Prof. Craig Clifford for his suggestions and corrections.

In order to determine whether successful interpreting performance is more a matter of nature or nurture, researchers focused on the issue of aptitude[1] and ways of testing it, an issue that had already been thoroughly investigated in educational and vocational fields (Carroll 1962; Cronbach & Snow 1977). Aptitude, according to Cronbach and Snow (1977), is "any characteristic of a person that forecasts his/her probability of success under a given treatment." For them, "psychologically, aptitude is whatever makes a person ready to learn in a particular situation" (1977: 6) and is therefore a predictor or forecaster. For this very reason, selecting interpreting candidates wisely has become not only a practical necessity for training institutions confronting human and financial restrictions, but also an ethical one, as discussed as early as 1965 at the AIIC Paris Colloque reported by Keiser (1965a). At a pedagogical level, he said, everybody could be trained, but he added:

> Mais il serait anormal de permettre à des candidats dépourvus de dons de se lancer dans des études d'interprétation alors qu'il serait patent qu'ils ne deviendront jamais interprètes, ou qu'ils ne le deviendront qu'après quinze ans d'études, ce qui entraînerait une disproportion des efforts injustifiable. (Keiser 1965a: 3)

Keiser's remarks touched upon two key, but highly elusive concepts relevant to aspiring interpreters, namely the assumption that there are ungifted candidates (candidats dépourvus de dons) and the assumption that it would be evident (qu'il serait patent) that they will never become interpreters unless they are prepared to work hard for many years.

Ever since, the debate about interpreting aptitude and aptitude testing has revolved around these two assumptions and many efforts have been made to address them from a scientific perspective, that is, adopting methodologies that can be checked and replicated and breaking down "interpreting aptitude" and performance into components that can be defined and measured. Against this backdrop, Mackintosh's title "Interpreters are made not born" (1999) has become a manifesto of a brand new approach to interpreting students' selection and training.[2]

The creation of academic institutions to train interpreters helped develop teaching methodologies and scientific research. Indeed, as we will see in the course of this paper, trainers and researchers (themselves often individuals who combined both roles) fruitfully joined efforts to devise tests aimed at "screening potential

1. In the present paper the word "aptitude" is used as an overall term encompassing abilities, skills and personal traits deemed necessary (according to common sense) or reliable predictors (according to scientific research) of successful interpreter training. See Bontempo and Napier (2009) for the debate on the concept of "aptitude" vs. "ability".

2. The same stance against the idea of the "born interpreter" had been voiced earlier by Lambert (1992a) who defined interpreting as a skill to be learnt through a step-wise training.

interpreters" (to quote a seminal paper by Moser-Mercer 1985). However, given the many cognitive, linguistic, pragmatic and motivational implications of becoming an interpreter, what kind of aptitude is expected from a would-be interpreter?

As we will see (§ 2.1), interpreter trainers (most of whom are also professional interpreters) relied on their intuition and personal experience to determine the features and abilities that define "aptitude" for interpreting (the interpreter ideal profile). The first important distinction emerging from the debate on interpreting aptitude (and consequently on selection procedures) is whether greater emphasis should be placed on already acquired skills or on the ability to learn interpreting-related skills (Lederer 1975; Sofr 1976; Pearl 1995 and more recently Tryuk 2002 and Angelelli 2007, who talks about "interpreter readiness"). On the other hand, Seleskovitch and Lederer focus on the concept of "teachability" (Keiser 1965a, 1965b), regarding most training goals as achieveable through exercise, others as unteachable, notwithstanding their vital importance for interpreting practice. Therefore, they maintain, aptitude testing should screen "teachable" candidates because it would be futile to try to teach interpreting to candidates who are not able to think logically.

The issue of defining aptitude and ways of reliably testing it, however, is not confined to academic institutions, but is also a matter of concern for professional interpreter associations. These are concerned with preserving high quality professional standards among its members, and take a particular interest in veryfing the teaching practices implemented by academic centers worldwide. In this field, the International Association of Conference Interpreters (AIIC), set up in 1953, has always played an extremely active role. In particular, AIIC links up training with the profession, and makes a point of according recognition only to those schools that comply with specific requirements, one of them being the use of an entrance examination (AIIC Training Committee 2010). AIIC's first major contribution to the debate on aptitude testing was the organizing of two conferences on the topic, both of them held in Paris (see Bowen & Bowen 1989): the "Colloque sur l'enseignment de l'interprétation" (18–19 December 1965) and the AIIC School Seminar (19–20 October 1974).[3] Subsequent conferences continued the debate: the NATO Symposium on "Language Interpretation and Communication", held in Venice in 1977 (Gerver & Sinaiko 1978), the European Parliament's Colloquium on Interpretation in a Multilingual Institution, organised in May 1979 (Bowen & Bowen 1989) and the conference organised by the SSLMIT of the University of

3. The seminar was held under the auspices of ESIT and is reported by Lederer (1975). Among other things, the suggestion was made that an "entrance examination" would be preferable to aptitude testing because the candidate was not being asked to prove his/her ability to interpret (Keiser 1978: 18).

Trieste in 1986 under the heading of "The Theoretical and Practical Aspects of Teaching Conference Interpretation" (Gran & Dodds 1989). While aptitude testing was on the agenda at all of these events and was clearly seen as relevant to training, it was not until the 2009 Antwerp conference on "Aptitude for Interpreting: Towards Reliable Admission Testing" that an entire event was devoted to it. This conference helped revive interest in aptitude for interpreting, and in the high cost — both for the institutions and for the prospective candidates (whether accepted or rejected) — of ineffective assessment of aptitude.

In the present paper I review the existing literature with a view to highlighting the main emerging themes: which qualities a prospective interpreter should possess that indicate an aptitude (§ 2), how these qualities may be measured (§ 3), the issue of valid and reliable testing (§ 4) and, finally, proposals for test designs (§ 4.1) and aptitude tests which have identified statistically significant predictors (§ 4.2). Although I will concentrate on spoken language aptitude testing, some contributions on sign language aptitude testing will also be mentioned.

2. *What* to look for in an interpreting candidate

The interest in defining aptitude and developing ways to test it has grown over the years as can be seen from Figure 1.

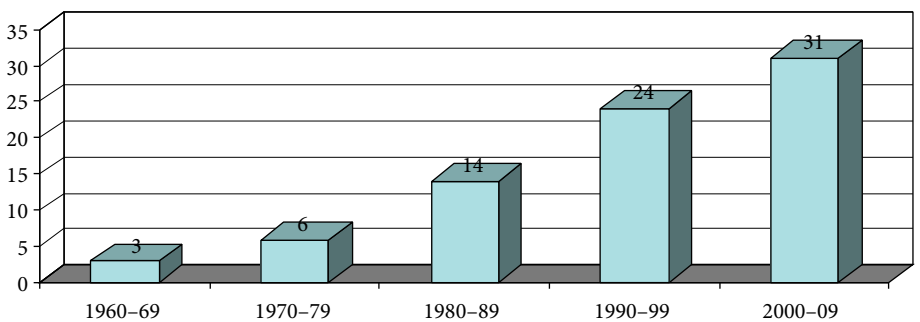

Figure 1. Distribution of publications on aptitude over the years

The efforts devoted to aptitude testing have been impressive, as has the range of publications offering different insights into aptitude and aptitude testing. Notwithstanding their diversity, some general trends can be detected and will be tackled in greater detail in the relevant sections of the present paper. Some authors with an interpreting background collaborated with scholars from other disciplines, such as psychology, to develop multi-component testing batteries which they described in detail (Gerver et al. 1984). Others drew inspiration from different disciplines to suggest tests based on those findings: the test based on research into

foreign language acquisition suggested by Carroll (1978), the shadowing[4] test described by Lambert (1992a) and the test on personality traits described by Schweda Nicholson (2005). Still others have approached aptitude from a more theoretical perspective and produced stimulating reflections: Lederer (1975) and Sofr (1976) drew attention to the concept of interpreting-related aptitude, and Alexieva (1993) focused on the role of short-term memory. Other publications provide descriptions of admission procedures with final-mark correlations (Arjona-Tseng 1994; Donovan 2003; Timarová & Ungoed-Thomas 2009) or of tests on abilities which proved predictive: lexical knowledge (Skaaden 1999), memory capacity[5] and verbal fluency, with particular reference to the production of synonyms (Gerver et al. 1989), cloze tests (Gerver et al. 1984; Moser-Mercer 1985; Lambert 1992b) and text processing strategies (Russo 1993; Russo & Pippa 2004). As to reports on tests, Shaw and Hughes (2006) provide the most recent extensive overview of the characteristics of second-language learners and sign language interpreters based on empirical studies with a view to suggesting evidence-based admission criteria to interpreter schools.

2.1 Ideal interpreter profiles: A consistent view over a 40-year span

As we have seen, the need for some admission criteria to training was stressed as far back as the very first conference on interpreting, the 1965 Paris Colloque. Indeed, the vast majority of participants were in favour of an aptitude test to regulate access to interpreter training institutions.[6] But aptitude at that time was still a very holistic concept, lacking evidence-based indicators or predictors. The first general framework of candidates' prerequisites was inspired by professional and training experience (Keiser 1965b). A review of almost forty years of publications referring to an ideal interpreter's profile as seen by trainers, scholars and professionals[7] (Herbert 1952; Keiser 1965a, 1965b, 1978; Longley 1989; Gerver et al.

4. Alexieva (1993: 8) reports on criticism of shadowing, not only as a reliable exercise for aptitude testing, but also as a training exercise (Dodds 1990), considered even harmful (Thiery 1990: 4) or undesirable (Déjean Le Féal 1998).

5. The vital role of memory as a prerequisite for simultaneous interpreting was also investigated by Darò (1995).

6. According to Keiser (1965a), only a handful supported the darwinistic "natural selection" view.

7. In the early 1990s, Lambert (1992b) analysed articles written by professionals and transcripts of interviews with interpreters, and noted that although empirical research on the skills and abilities required for interpreters was underdeveloped, the professional field was in agreement on the need for these.

Table 1. Ideal interpreter profiles across a 40-year span

Herbert (1952) 1965 AIIC Colloque		**1974 Paris Symposium (Keiser 1978: 17, 20)**	**Gerver et al. (1984: 19); Lambert (1992b: 25)**	**Longley (1989: 106)**
Professional experience perspective	Professional and training experience perspective	Professional and training experience perspective	Professional experience perspective (reported)	Training experience and scholarly perspective
Command of two foreign languages; Lexical fluency; Good memory; Broad general culture; Interpreting attitude	Foreign language command (B+C); Strong mother tongue (A); Love for languages; Ability to comprehend; Language transfer skills	Ability to comprehend and mental rapidity; Expressive ability; General culture; Personal qualities (capacity to adapt to subject, speakers etc.; public speaking, tact and diplomacy); Physical and psychological stamina	Profound knowledge of active and passive languages and cultures; Ability to grasp rapidly and to convey the essential meaning of what is being said; Ability to project information with confidence, coupled with good voice; Wide general knowl-edge and interests, and a willingness to acquire new information; Ability to work as a team member	Excellent knowledge of A, B, C languages and cultures; Ability to grasp rapidly and to convey the essential meaning of discourse, irrespective of the language spoken; A memory which recalls the links between logical sequences of discourse; Ability to convey information with confidence, coupled with a pleasant delivery; Broad general knowledge and interests, a curiosity and willingness to acquire new information; Ability to work as a team member; Ability to work under stress for long periods

1984; Lambert 1992b) shows that the basic assumptions about the characteristics of a suitable candidate have remained substantially the same (see Table 1).

The synopsis in Table 1 includes many abilities and personal qualities that are self-explanatory. Yet, some do require clarification. The item "Ability to comprehend", listed in the 1965 AIIC Colloque, was further specified as follows (Keiser 1965a: 6):

> Pour comprendre il faut être intelligent et connaître le monde. Cette faculté de compréhension est constituée de deux éléments: de dispositions innées (la vivacité intellectuelle, l'esprit de finesse, une bonne mémoire, qui peut se cultiver d'ailleurs, le goût des équivalences sémantiques, l'aptitude à la transposition rapide d'une langue dans une autre) et de capacités acquises.

As to the acquired skills mentioned by Keiser (1965a), these include a broad knowledge of general culture and keeping abreast of current affairs. In particular, candidates must know how to think logically, and this is provided by a university background. The concept of "ability to comprehend" also covers the otherwise elusive concept of "interpreting attitude". To sum up, an aptitude test should reveal whether a candidate

> a la faculté de compréhension nécessaire, notamment celle d'analyser une information rapidement, d'eviter les contresens et les incohérences; dispose d'un esprit rapide lui permettant le passage d'une langue à l'autre et d'un sujet à l'autre, lui permettant de ne pas perdre ses moyens s'il est tout à coup confronté avec un domaine qui lui est étranger; a, en outre, la discipline linguistique nécessaire pour assurer sans faille la dissociation linguistique (éviter les "faux amis", le mélange de plusieurs langues). (Keiser 1965b: 35)

Interestingly, Keiser cites the candidate's ability to avoid countersense and incoherence as an important component of the ability to comprehend (la faculté de compréhension nécessaire). This ability is hardly mentioned in the literature on aptitude (with the exception of Gerver et al. 1984). And yet, Russo and Pippa (2004) have found that the loss of coherence in producing speech during an on-line paraphrasing test is a strong predictor of failure in interpreter training (see § 4.2).

Even though "assertiveness" is not included in the list of prerequisites of a suitable candidate in Table 1, it is a quality which professionals consider extremely important (Lambert 1992b). By the same token, although the synoptic table does include physical and psychological stamina among the prerequisites (reported by Keiser 1978), this factor is still generally neglected in aptitude testing, as the literature on the topic shows. This is not the case for another aptitude component included in Table 1, namely memory (Herbert 1952; Longley 1989), which is included in several test batteries, either directly (Wechsler memory scale, Gerver et al. 1984) or indirectly (production of summaries, Timarová & Ungoed-Thomas

2008). In particular, strong short-term memory capacity should be tested because it highlights the candidate's ability to capitalise on intertextuality (Alexieva 1993).

2.2 Other contributions to the debate

The above-mentioned general framework of abilities and personal qualities based on the intuition of interpreting trainers, scholars and professionals (Table 1) was further enriched by different approaches to aptitude and aptitude testing, particularly the contribution of an educational psychologist, Carroll (1978), who laid the foundations for much of the research. Coming from the field of foreign language acquisition and testing, Carroll studied admission testing procedures at Georgetown University: on the one hand, he was appalled by interpreting candidates' poor knowledge of their foreign languages and, on the other, he became more aware of the main features required of an interpreting performance. Based on findings in psychometrics and in foreign language acquisition, particularly Thurstone's analysis of cognitive abilities into a series of primary factors (Thurstone 1938, quoted in Carroll 1978), he recommended testing the following cognitive abilities:

- verbal intelligence (the V factor)[8]
- general culture
- word fluency factor (ideational fluency, i.e. a rapid and coherent flow of ideas on a topic; expressive fluency, i.e. rephrasing a sentence in as many ways as possible; associative fluency, i.e. production of synonyms and antonyms).
- naming facility factor
- shadowing.

In psychology, verbal intelligence and word fluency factors were measured by means of written tests, but Carroll (1978) suggested that oral tests be introduced instead. Gerver et al. (1984) proposed the use of associational fluency tests while Lambert (1992b) opted for shadowing, extensively used in psychology for its

8. "(…) the factor usually identified with 'V', or 'verbal intelligence', is probably the most relevant to perform as an interpreter or translator. (…) (For foreign or native language), V factor involves not only the individual's knowledge of advanced vocabulary, but also his sensibility to established word usages, to nuances of idiomatic phrases, and even his ability to predict the transitional probabilities of words in phrases, as in test called Phrase Completion. (…) The verbal factor is one of the best and most easily established factors of intelligence. It is involved not only in tests of vocabulary knowledge, but also in tests for reading comprehension, ability and facility (speed) in detecting semantic and syntactic ambiguities, and ability in writing effective, highly rated themes (…). For practical testing purposes, the V factor is best measured by many vocabulary tests, with emphasis on the exact meanings of more difficult and rarer words of a language." (Carroll 1978: 123–124)

ability to assess memory structure, notwithstanding the controversy surrounding its suitability for interpreting (see footnote 4). Some items related to verbal fluency were subsequently tested by several authors: completion/detection tests (Gerver et al. 1984), lexical knowledge (Skaaden 1999) and production of synonyms (Russo & Pippa 2004).

A more recent overall description of the qualities and abilities expected of a prospective conference interpreter ("What kind of personal traits do I need to be a conference interpreter?") is provided by AIIC. The recommendations include:

- a polished command of their own native language over a range of registers and domains
- a complete mastery of their non-native languages
- a familiarity with the cultures in the countries where their working languages are spoken
- *a commitment to helping others communicate*
- an interest in and understanding of current affairs, plus an insatiable curiosity
- world experience away from home and school and a broad general education
- good training (and usually at least an undergraduate university degree)
- the ability to concentrate and focus as a discussion unfolds
- a pleasant speaking voice
- *a friendly, collegial attitude*
- *calm nerves, tact, judgment and a sense of humor*
- *a willingness to adhere to rules of conduct* (e.g. confidentiality).

(AIIC Training Committee 2010; emphasis added)

Among other things, the focus is placed on the social dimension of interpreting (helping to communicate, friendliness, sense of humor, professional ethics etc.) which is virtually neglected in the earlier contributions.

Finally, a comparative view of expected knowledge and aptitudes for spoken-language and signed-language interpreters is offered by López Gómez et al. (2007), reproduced here as Table 2.

While many of the reported items are shared by both kinds of would-be interpreters, others appear to be more relevant for spoken-language interpreters.

Another interesting aptitude highlighted by Shaw and Hughes (2006) was "learning ability" — namely the mental and behavioural ability to adjust in the face of criticism and advice. Recently, attention is being paid — or at least this is what the interpreting schools claim — to other "soft features" (Timarová & Ungoed-Thomas 2009) of the interpreter profile, but hardly any test is currently being administered to this effect (Timarová & Ungoed-Thomas 2008), even though qualities such as motivation were assessed in the past (using the interview method, in addition to sight translation; Arjona-Tseng 1994).

Table 2. Knowledge and aptitudes mentioned in the literature (López Gómez et al. 2007: 77)

Knowledge		Aptitudes	
LINGUISTIC	High command of A & B languages ↔♥ Specific vocabulary ↔	Efficient input segmentation ↔ Attentional division ↔ Use of language-pair specific strategies ↔ Predictable properties of language ↔ Change translation strategies ↔ Verbal fluency ↔♥ Processing speed ↔♥ Good long and short-term working memory ↔♥ Powers of concentration ↔♥ Capacity to sign and talk simultaneously ♥	COGNITIVE
CULTURAL	Knowledge of A & B cultures ↔♥ Cultural background ↔♥		
ACADEMIC	University degree or equivalent ↔	Adapt without delay to different speakers/ signers, situations and subjects ↔♥ Pleasant voice and public-speaking skills ↔♥ Stress resistance & self-control ↔♥ Team work ↔ Professional distance ↔♥ Likes to be well-informed ↔ Diplomacy ↔ Good self-concept ↔	PERSONALITY

↔ Aptitude mentioned for spoken-language interpreters
♥ Aptitude mentioned for signed-language interpreters

The general overview of the literature on interpreters' aptitude provided in this section highlights some basic distinctions among expected skills: (a) language knowledge and mental skills (Keiser 1978); (b) acquired skills and interpreting-related trainable skills (Lederer 1978); (c) hard skills and soft skills (Timarová & Ungoed-Thomas 2008). Are all these skills equally predictive?

3. *How* to measure interpreter candidates' skills?

Over the years, ways to test interpreting candidates' abilities have changed only slightly. All tests have always been targeted either to individual skills (for instance, memory capacity) or multiple skills (for instance, interlinguistic recall which

implies both foreign language control and memory). The first tests were inspired by intuition and common sense and presented what Clifford (2005) called "face validity", that is, they resembled tasks that interpreters would perform (for instance, listening and speaking at the same time or sight translation) or skills they displayed (general culture tests). Gradually, trainers and scholars in the field recognized the need to administer validated and reliable testing procedures. Some authors suggested the application of paradigms used in other fields such as the expert-novice paradigm (Moser-Mercer 1997, 2000), which may be valid for several aptitude traits, but not for all of them. A case in point is working memory performance, where the validity of this paradigm was recently disproved (Köpke & Nespoulous 2006).

The first author to voice the necessity for validated and reliable testing procedures was the former Director of the Interpreting Program of the Polytechnic of Central London (PCL), Patricia Longley, who set up a multidisciplinary team including psychologists (Gerver et al. 1984; Longley 1989). Ever since, all papers on aptitude testing are concerned with scientifically sound screening procedures and look for correlations with external measures such as final exam pass rates, interpreting exam average scores or the number of sessions needed to complete the training successfully. The most recent examples are Pippa and Russo (2002), Russo and Pippa (2004) and Timarová and Ungoed-Thomas (2009).

3.1 Types of tests

3.1.1 *Early aptitude test batteries*

The very first publication on the subject, the already mentioned 1965 AIIC Colloque sur l'enseignment de l'interprétation, contained a report by Walter Keiser, "Admission dans les Ecoles d'interprétation", which not only summarised the debate on skill testing, but also gave a detailed account of the selection procedures implemented by the various interpreting schools, among them ETI (Geneva), Übersetzer- und Dolmetscher-Institut (Heidelberg), ESIT (Paris), ISIT (Paris), Ecole d'Interpretation (Antwerp) and the Business School HEC (Paris). The types of tests reported at that conference varied, depending on the respective schools' curricula and the candidate's level of training prior to admission. The common features were an interview and a written or oral translation. Some schools included a written essay in the mother tongue on a topic of general culture. This was the eliminatory stage. The recommendations that emerged from the 1965 Colloque were the following: all examiners should be conference interpreters and, in the oral part of the exam, all tests should be truly oral (with the exception of sight translation from a written text). The test should measure

– les connaissances linguistiques
– la culture générale
– le niveau intellectuel, la faculté de compréhension et la rapidité d'esprit
– la facilité d'expression et la diction
– l'aptitude physique et psychique. (Keiser 1965b: 31)

A comparative view of the tests implemented by the different schools is shown in Table 3. Information on the number of students is not available.

Table 3. Tests administered by the schools participating in the 1965 Paris Colloque

School	Test type	Duration	Nature	Correlation with final exam
ETI, Geneva	1. Interview 2. Recall from A into A 3. Interpretation from B into A	20–30 minutes	Voluntary and not binding	No
Heidelberg	1. Consecutive from A into A 2. Sight translation from B into A 3. Interview in language A	20–30 minutes	Compulsory	No
H.E.C., Paris (a business school)	1. Translation B/C >A; A>B 2. Essay in language A 3. Interview A/B/C 4. Sight translation B/C>A 5. Reading aloud 6. Question on the news of the day	20–30 minutes	Compulsory	Yes
ESIT, Paris	1. Interview in A/B/C 2. Explanation in A of a passage in B/C 3. Recall in A from B/C 4. Sight translation from B/C>A	Not indicated	Not indicated	Not indicated
ISIT, Paris	1. Sight translation or elementary interpretation B>A 2. Sight translation or elementary interpretation A>B 3. Sight translation C>A 4. Interview in A/B/C	Not indicated	Not indicated	Not indicated
Ecole d'interprétation, Antwerp	1. Consecutive from B/C into A 2. Interview in A/B	Not indicated	Binding (with possible *repéchage*)	Not indicated

3.1.2 *More recent test battery descriptions*

Other detailed examples of admission tests used by different schools are provided by Lambert (1992b), along with the methods used by the University of Ottawa since 1984 to select candidates for the graduate diploma programme. The selection

exams included tests for shadowing (phonemic and phrase shadowing), cloze, sight translation/sight interpretation, memory test, and an interview. These tests are based on cognitive theories. The methods seem to have enabled examiners to identify the characteristics considered most important (general knowledge, command of A and B languages, ability to transfer meaning, specific interpreting-related skills, personality traits). The author stated that the array of methods used by the University of Ottawa[9] can identify those candidates with little chance of success in the field of interpreting. Lambert hoped to be able to determine whether a further significant correlation exists between the test results and those scored later during the diploma programme. Regrettably, the author specified neither the administration procedures nor the assessment criteria, so that no replication is possible. Furthermore, given the modification of the tests over time, since 1984, candidates were given different tests, and therefore correlation with the final exam results could not be checked.

Later on, Moser-Mercer (1994) reported both on the types of tests and when they are administered. Schools differed considerably in this regard, depending on the course design and content. For instance, abilities that are known to improve through practice are often not tested as strictly as others. Furthermore, duration of training affects testing: for example, a 4-year degree has different testing priorities than a 6-month postgraduate programme. Several other authors describe aptitude testing criteria in relation to course duration (among them, Lambert & Meyer 1988; Bowen & Bowen 1989).

A more recent overview of selection practices is provided by Tryuk (2002) with special reference to interpreting schools in Eastern Europe. In general, all contain items belonging to the concept of "aptitude to perform an acquired skill" (e.g. the consecutive test) together with those belonging to the concept of "aptitude to perform interpreter-related skills" (e.g. recall).

Another test battery was developed by the European Master in Conference Interpreting (EMCI), which was set up in 2001. This includes traditional tests, namely, recall, interview and sight translation. The test starts from the following premise:

> Our aptitude test is typically a proficiency test looking to a future situation of language use, i.e. the test performances are regarded as indicators of how the candidate would perform on related tasks during training and, ultimately, in the interpreting booth. (Sunnari 2002: 25)

9. Harris (1992) also describes the admission procedure as follows: written translation (filter test), sight translation, shadowing and short-term memory test.

Those who assess candidates and those who will train them need to answer the following questions: "Does the applicant meet our criteria? Is the applicant trainable? Will we be able to work together?" As Sunnari (2002) states, the evaluation criteria are necessarily subjective and heavily rater-dependent. The paper also offers the rationale and the description of the tests, but no indication of test efficiency (correlation with final exam pass rate) or the number of candidates tested is provided. To select students expected to work into their B language in the simultaneous mode, jurors assess candidates' recall tests on the basis of the following criteria: resourcefulness, robustness of B language, motivation, quality of A language and teachability (i.e. ability to react to advice and criticism) (EMCI 2002: 61).

The most recent test overview was published by Timarová and Ungoed-Thomas (2008), who surveyed 18 schools. The authors grouped the components of the various tests and skills into several major categories which still reflect the 1965 AIIC recommendations, except for the physical-psychological aptitudes which are hardly ever tested. According to their survey, schools test the five abilities shown in Table 4.

Table 4. Abilities tested in a sample of 18 schools (Timarová & Ungoed-Thomas 2008: 39)

Category	Type of test	No. of schools
Language	short consecutive	14
	short speech by candidate	9
	interview	7
	summary	7
	translation	7
Communication	short speech by candidate	11
	short consecutive	10
	summary	6
	interview	5
Comprehension	summary	8
	translation	8
	short consecutive	5
Analysis	summary	11
	short consecutive	6
	translation	5
General Knowledge	interview	8
	written test	4
	short speech by candidate	3

A further interesting insight into aptitude testing provided by the authors' survey concerns the number of tests best suited to assess the candidate's skills and subskills: the more tests, the better? The authors did not find any significant correlation between the number of tests in a battery and the final exam pass rate. This result is a further encouragement to intensify research into devising or selecting the most appropriate tests.

As to the testing methodology, Angelelli (2007), writing on the medical interpreting profession (dialogue interpreting), states that language and interpreting skills should be tested separately. Furthermore, as to the testing material, she writes: "Undoubtedly, the use of real discourse interpreted by real interpreters on the job […] is a stronger indicator of content and construct validity than the assessment of isolated skills and traits." (2007: 68). Finally, with respect to the type of tests, she criticises the use of sight translation, a widely used test: "Sight translation is not part of IR (Interpreter Readiness) since, as evident in the literature, it is an acquired skill and therefore needs to be taught" (Angelelli 2007: 74).

4. Test validity and reliability

"The almost total lack of data relating pass/fail results in final exams, leaves the evaluation of aptitude for interpreting wide open to criticism" wrote Barbara Moser-Mercer in 1994 (p. 65). Today data on correlations are available, but more specific studies making use of statistical models to prove their predictive significance are still scant (Angelelli & Jacobson 2009). The issue of unreliability and excessive subjectivity in jurors' assessments has been voiced repeatedly (Gerver et al. 1984; Gringiani 1990; Dodds 1990; Pearl 1995; Kalina 2000; Sunnari 2002; Pippa & Russo 2002; AIIC 2006, Timarová & Ungoed-Thomas 2008), and has impelled some schools to adopt a different approach. This is the case of the PCL under Patricia Longley (1984, 1989) which introduced an aptitude testing battery claimed to be objective (§ 4.2) and of the SSLMIT at the University of Trieste, which revised its non-binding admission procedure[10] as reported in a study by Gringiani (1990), who found that the test's power of prediction was very poor: the error percentage for the group that passed and eventually withdrew was usually higher than for the group that failed but eventually completed the course: 45% vs. 36% and 37% vs. 18%, respectively. These results spurred fruitful reflections within the school and a debate over the kind of general knowledge that is realistic

10. The test battery included: written summary B>A language, sight translation A>B, improvised speech in A and/or B, interview in A and B on current affairs and general knowledge, questions on grammatical aspects of the B language.

to include in aptitude tests, considering the fact that candidates are 20-year-old students (in particular, Gran and Taylor considered the distinction between variable and non-variable information useless) and the fact that tests failed to consider extremely important factors such as motivation and the candidates' attitude towards the test (Dodds 1990). Gringiani's results also opened up other research lines into aptitude testing at the SSLMIT of Trieste, namely, the study of text processing strategies analysed by means of on-line paraphrasing (Russo 1989, 1991, 1993, 1995; Pippa & Russo 2002; Russo & Pippa 2004).

4.1 Proposals for aptitude test designs

Several authors have suggested aptitude testing designs based on tests with some empirical support, but no observational studies have been carried out to verify their predictive power. We refer to contributions by Moser-Mercer (2000), Bernstein and Barbier (2000), Clifford (2005) and Schweda Nicholson (2005).

Moser-Mercer (2000) presented the Geneva Aptitude Project which draws upon the expert-novice paradigm (see § 3) and includes shadowing, delayed auditory feedback and verbal fluency. Results on correlations with interpreting exams, however, do not seem to have been published.

Bernstein and Barbier (2000) suggested the development of a rapid automatic screening test for prospective simultaneous interpreters based on the PhonePass-SET-10 test methodology, a fully automatic system for testing spoken language performance, designed by Bernstein. To develop this complex but promising project, however, a large interpreting population (between professionals and trainees) and considerable funding are required, which renders it impracticable for individual training institutions.

Clifford (2005) has suggested a psychometric aptitude test, which measures mental operations on the basis of exercises and questionnaires targeted at single operations (e.g. comprehension).

Schweda Nicholson (2005) interestingly suggested an exploration of the relation between personality traits and aptitude by means of a widely used standardized test, the Myers-Briggs Type Indicator (MBTI). A proposal to study personality traits, in addition to motivation and interest in becoming an interpreter or translator, had already been put forward by Szuki (1988), who suggested the development of an ad hoc test. Schweda Nicholson studied 68 interpreting trainees and found that her student population was equally distributed between Extroverts/Introverts, Sensing/Intuitives, Judging/Perceiving with a prevalence of Thinking over Feeling types. It would be interesting to replicate the study on a much larger population and correlate the results with school or professional performance.

4.2 Predictive aptitude tests

Over the years, efforts have been made to single out individual skills relevant to interpreting by means of validated and reliable aptitude tests. In order to verify the test validity, aptitude test scores were correlated with proficiency at interpreting exams and at final exams; to verify the test reliability, aptitude test results were scored by means of criteria that were claimed to be objective and therefore replicable. A preliminary consideration is that all these studies[11] do not generally have large sample sizes by commonly accepted scientific standards: 19–29 subjects (Gerver et al. 1984), 28 subjects (López Gómez et al. 2007), 83 subjects, reduced to 57 (Moser-Mercer 1985), 46 subjects (Russo & Pippa 2004). However, they have been able to detect some statistically significant predictors. Let us now consider the main features of these studies and their main results, starting from the contribution on aptitude for sign language interpreting (López Gómez et al. 2007).

The authors' hypothesis was that perceptual-motor and cognitive skills, personality factors and academic background were possible predictors of success. They correlated candidates' aptitude testing scores with their final examinations, and found perceptual-motor and cognitive abilities to play a more significant role than personality traits in predicting proficiency in learning a signed language and developing signed-language interpreting abilities. On the other hand, academic background does not appear to be predictive. The best predictive test was the one based on the use of pseudosigns.

As far as aptitude for spoken-language interpreting is concerned, the first most comprehensive and methodologically sound aptitude test was developed by Gerver et al. (1984, 1989) and was implemented at the Polytechnic of Central London. The authors concentrated on the second of the several qualities expected in a candidate (see Table 1, column 4: "Ability to grasp rapidly and to convey the essential meaning of what is being said") and suggested twelve tests of three different types: (1) text-based tests, which assume the processing of connected discourse to be a crucial feature of the interpreter's task; (2) subskill-based tests aimed at assessing verbal ability; (3) stress-based tests aimed at assessing performance under time pressure. The authors obtained the following main results:

i. Differences between students who passed/failed exams were significant for memory for text, logical memory (1 and 2), error detection, cloze (1 and 2) and synonyms.

11. For the sake of completion, Skaaden's study (1999) on the impact of written lexical knowledge on interpreting aptitude involving only 6 subjects should also be mentioned in this section because she carried out correlations with final exams, but no statistical significance is provided.

ii. Simultaneous/Consecutive differentiation: completion/detection tests correlated more with ratings of the simultaneous examination, while recall tests correlated more with ratings of the consecutive examination.
iii. Of the subskills and speed stress tests, only the synonyms test reflected significant differences between candidates who passed and those who failed.
iv. Given the different nature of the tests, the results suggest that the processing of connected discourse constitutes a crucial feature of the interpreter's task and should be included in selection tests.

From their study, synonym production seemed a particularly reliable predictor. The ability to produce words semantically related appears to be especially relevant in aptitude tests for interpreting (see also Russo & Pippa 2004). Gerver et al. (1984) concluded:

> In a general way, therefore, good performance on the tests was associated with a good showing in the interpreting examinations. This suggests that the abilities required by interpreters — to the extent that they are reflected in the examination ratings — are also reflected by the tests. (Gerver et al. 1984: 27)

The success prediction rate for those students taking the test battery was 39% (17 out of 29 students passed their final examinations), which is better than that of Trieste (36%) as reported by Gringiani (1990). By admission of the authors (Gerver et al. 1984: 20), however, it was not possible to fully assess the validity and reliability of the suggested test because this would have required the independent assessment of at least two raters for a large enough number of candidates. The sample comprised 29 students, but there were not more than 19 across the 12 different tests administered. Unfortunately David Gerver's untimely death prevented the follow-up of the test efficacy in a consistent way, though an indication of the predictive efficacy of the procedure was indirectly provided by Longley (1989). A recent noteworthy effort has been made by colleagues at Lessius University College to replicate this aptitude test battery, but conclusive results have yet to be published.

At present, the then PCL and now University of Westminster implements the following simplified admission procedure described in the FAQs of the MA course "Conference Interpreting Techniques":

> **Do I have to take an entrance test?**
> If your application is accepted, you will be invited to an entrance test. The test is as follows:
> A/CC(C) combination: Written translation from each passive language into mother tongue. You spend an hour on each language and do **two** translations per language. This takes a total of 2 hours, 3 if you offer 3 C languages.

> A/A combination: Written translation from A_1 into A_2 and A_2 into A_1
> A/B combination: Written translation from into A or B and B into A. Again, one hour per language and **two** test pieces per language.
> If you offer an A/B/C combination, you will translate A-B, B-A and C-A.
> Translation tests will be comparable to quality Broadsheet articles
>
> You will then take a one hour **General Knowledge test**.
>
> (University of Westminster 2010; emphasis in original)

Another aptitude test associated with an observational study was carried out by Moser-Mercer (1985). The author sampled 83 students (reduced to 57) on shadowing, dual-task training, memory exercises, paraphrasing and number processing after a 10-week course and correlated their scores (recommendation, qualified recommendation, no recommendation) with exam results. Her hypothesis was that the students who received a recommendation were more likely to pass the qualifying examination at the end of their first year of study and the final professional examination at the end of their second year than students who received only a qualified recommendation or none at all. A significant relationship was found to exist between the type of recommendation issued and examination results. Indeed, 80% of those receiving a recommendation (32 out of 40), 33.3% of those receiving a qualified recommendation (3 out of 9) and only 12.5% of those receiving no recommendation (1 out of 8) passed the exam.

A more recent attempt at developing an aptitude test associated with an observational study was carried out by Russo (1989, 1991, 1993, 1995), Pippa and Russo (2002) and Russo and Pippa (2004). The test entails an on-line oral paraphrasing exercise of about 4 minutes in Italian administered to Italian mother tongue interpreting candidates after a short training session (10–15 minutes) designed to familiarise them with the task of listening and speaking at the same time. The goal of the test, which hinges on linguistic and cognitive strategies, was to predict successful program completion by students wishing to attend a conference interpreting course. The test design was inspired by the following methodological prerequisites. A test is reliable if a subject repeating the same or equivalent test achieves the same results, and it is valid only if it actually measures what it sets out to measure. Furthermore, determination of its validity requires independent criteria. In our case, test scores were correlated with the average mark on interpreting exams (*proficiency*) and with the number of sessions needed to pass the interpreting exams (*time*). The candidates' reformulations were evaluated at each unit level, starting from syntax, followed by semantics and finally pragmatics. For each level of analysis, relevant categories were developed and applied. They were specifically and univocally defined. The evaluation criteria of the candidates' performance evolved from Russo's early work, in an effort to devise a valid and

Table 5. Correlations between single operations, number of sessions and average mark at interpreting exams (* = $p < .05$)

Operation	Correlation with sessions	Correlation with average mark
Loss of coherence	.33*	−.30*
A. Synonymic substitution with pragmatic loss (4 items)	−.30*	.32*
B. Synonymic substitution without pragmatic loss (3 items)	−.31*	.34*

reliable test to reduce the bias of an overly subjective assessment (for the rationale of the paraphrasing test and evaluation criteria see Pippa & Russo 2002). The test was administered experimentally, that is, not as an eliminatory entrance test, to 46 students about to begin their interpreter training course at the SSLMIT of the University of Trieste in the academic years 1988–89, 1990–91 and 1991–92.

The preliminary results (Russo & Pippa 2004) showed that of all the possible textual operations, the most powerful predictors, which significantly correlated both with the average mark and the number of sessions, were only two: loss of coherence (especially at sentence level), and synonym substitution (with or without pragmatic loss), as can be seen in Table 5.

Both of these operations involve deep semantic processing and memory capacity, required to perceive a speech while simultaneously paraphrasing it, as a single conceptual framework, either at sentence level or at inter-sentence level, and to retrieve related lexical items (associational fluency; Carroll 1978). These interpreting-related cognitive abilities are consistent with results obtained by Gerver et al. (1984) on synonym production and the processing of connected discourse as being highly relevant for an interpreter's performance (see above). It would seem that these abilities support the notion that the best performers at paraphrasing complete the course at a faster pace and receive better evaluations by their interpreting teachers in both their language combinations and throughout their school careers.

The other three operations which involve morpho-syntactic reformulations without any impact on semantics, namely, expansion (addition of verbal material without added information), permutation (mere shifting linguistic segments) and syntatic transformation are indicative of expressional fluency. This is certainly relevant to an interpreting task (Carroll 1978) and seems to have an impact on performance evaluation as well, since the average mark on interpreting exams correlates most significantly with the test scores. However, there is no statistically significant correlation between the time needed to complete the course (number of sessions) and the individual above-mentioned operations, indicating that these do not seem to play a meaningful role (Table 6).

Table 6. Correlations between single operations, number of sessions and average mark at interpreting exams (** = $p < .01$; *** = $p < .001$)

Operations	Correlation with sessions	Correlation with average mark
C. a. Operation with Ø semantic value without pragmatic gain- Expansions# (7 items)	(–.28)	.45**
b. Operation with Ø semantic value with/without pragmatic gain- Permutations# (3 items)	(–.18)	.42**
c. Operation Ø semantic value with pragmatic gain- Syntactic Transformation# (3 items)	(–.23)	.49***

This analytical study (Russo & Pippa 2004) has enabled the authors to single out the two strongest predictive operations during paraphrasing, on the basis of which to evaluate a future candidate's performance: loss of coherence and use of synonyms.

Based on these results, paraphrasing has been included in the admission testing procedure together with an oral cloze and a recall for each B language of a candidate at the SSLMIT of the University of Bologna at Forlì when the Conference Interpreting MA Program started in October 2004. We have assessed the most effective and least time-consuming admission testing procedure in the first three cohorts of students (academic years 2004–05, 2005–06 and 2006–07). Over these years we have collected data on students' achievements (average mark in first and second years and final interpreting exams and the number of sessions needed to pass them) with a view to comparing them with admission scores. The results from admission to final exams of these cohorts are reported in Russo (in this volume).

5. Conclusions

The assumption that only gifted individuals may become conference interpreters has gradually given way to a more articulated concept of interpreting aptitude which ranges from holistic assumptions (e.g. interpreting attitude) to scientifically sound measurements (e.g. aptitude predictors such as synonym production).

The overview of the literature on aptitude testing provided in this paper has shown that the ideal interpreter profile has not changed dramatically over time: this broad consistency in professionals', trainers' and researchers' views of what

makes a good interpreter is encouraging in that it provides a fairly homogeneous framework of reference for *what* to look for in an interpreting candidate. Similarly, the prevailing types of tests have changed very little over these last forty years, but unfortunately very few tests have been "put to the test" for scientific soundness. In the meantime, some interesting new aptitude testing projects such as the automatic screening of potential candidates (Bernstein & Barbier 2000) have been put forward, but unfortunately they are not always practicable.

Yet, over the last twenty-five years research efforts have been made to provide evidence of aptitude testing reliability and validity. This research has produced several statistically significant predictors of interpreting aptitude, which appear to show that interpreting-related cognitive skills and verbal fluency may be measured and may be predictive both for spoken-language and for signed-language interpreting (Gerver et al. 1984; Moser-Mercer 1985; Russo & Pippa 2004; López Gómez et al. 2007). Furthermore, available results indicate that some tests are more predictive for simultaneous than for consecutive interpretation (Gerver et al. 1984; Rejsková 1999) and that administering more tests at once may be redundant and counter-productive (Gerver el al. 1984; Timarová & Ungoed-Thomas 2008).

Test validity is determined by external measures. In interpreting studies these are usually interpreting exam marks and final exam pass-fail rate. Yet time (i.e. time needed to successfully complete the training course) is crucial in interpreting students' careers but, unfortunately, it is hardly ever taken into consideration in the correlations with aptitude testing scores. Equally neglected in the entrance examinations of most schools are those "soft skills" like motivation or personality traits (Timarová & Ungoed-Thomas 2008) whose role in determining or jeopardising a student's academic career seems to be of paramount importance, as has recently been reported by Bontempo and Napier (2009). Thus, in predicting the profile of the ideal candidate there is still ample room for improvement.

To conclude, the relation between instructional outcome and aptitude score is linked to the training specificities of each interpreting institution, hence the need for exchanging experiences in aptitude testing. In so doing, we will foster a mutually enriching learning process to better select present and future interpreting students.

References

AIIC Training Committee (2010). http://www.aiic.net/ViewPage.cfm/page56.htm (accessed 26 October 2010).

Alexieva, B. (1993). Aptitude tests and intertextuality in simultaneous interpreting. *The Interpreters' Newsletter* 5, 8–12.

Angelelli, C. (2007). Assessing medical interpreters: The Language and Interpreting Testing Project. *The Translator* 13 (1), 63–82.

Angelelli, C. V. & Jacobson, H. E. (2009). Introduction. Testing and assessment in translation and interpreting studies: A call for dialogue between research and practice. In C. V. Angelelli & H. E. Jacobson (Eds.), *Testing and assessment in translation and interpreting studies: A call for dialogue between research and practice*. Amsterdam/Philadelphia: John Benjamins, 1–10.

Arjona-Tseng, E. (1994). Psychometric selection tests. In S. Lambert & B. Moser-Mercer (Eds.), *Bridging the gap: Empirical research in simultaneous interpretation*. Amsterdam/Philadelphia: John Benjamins, 69–86.

Baigorri Jalón, J. (2000). *La interpretación de conferencias: el nacimiento de una profesión. De París a Nuremberg*. Granada: Comares.

Bernstein, J. & Barbier, I. (2000). Design and development parameters for a rapid automatic screening test for prospective simultaneous interpreters. *Interpreting* 5 (2), 221–238.

Bontempo, K. & Napier, J. (2009). Getting it right from the start: Program admission testing of signed language interpreters. In C. V. Angelelli & H. E. Jacobson (Eds.), *Testing and assessment in translation and interpreting studies: A call for dialogue between research and practice*. Amsterdam/Philadelphia: John Benjamins, 247–295.

Bowen, D. & Bowen, M. (1989). Aptitude for interpreting. In L. Gran & J. Dodds (Eds.), *The theoretical and practical aspects of teaching conference interpretation*. Udine: Campanotto Editore, 109–125.

Carroll, J. B. (1962). The prediction of success in intensive foreign language training. In R. Glaser (Ed.), *Training research and education*. Pittsburgh: University of Pittsburgh Press, 87–136.

Carroll, J. B. (1978). Linguistic abilities in translators and interpreters. In D. Gerver & H. W. Sinaiko (Eds.), *Language interpretation and communication*. New York: Plenum Press, 119–130.

Clifford, A. (2005). Putting the exam to the test: Psychometric validation and interpreter certification. *Interpreting* 7 (1), 97–131.

Cronbach, L. J. & Snow, R. E. (1977). *Aptitudes and instructional methods*. New York/London/Toronto/Sydney: Irvington Publishers.

Darò, V. (1995). Attentional, auditory and memory indexes as prerequisites for simultaneous interpreting. In J. Tommola (Ed.), *Topics in interpreting research*. Turku: University of Turku, Centre for Translation and Interpreting, 3–10.

Déjean Le Féal, K. (1998). Non nova, sed nove. *The Interpreters' Newsletter* 8, 41–49.

Delisle, J. & Woodsworth, J. (Eds.) (1995). *Translators through history*. Amsterdam/Philadelphia: John Benjamins.

Dodds, J. M. (1990). On the aptitude of aptitude testing. *The Interpreters' Newsletter* 3, 17–22.

Donovan, C. (2003). Entrance exam testing for conference interpretation courses: How important is it? *Forum* 1 (2), 17–45.

EMCI (2002). Preliminary conclusions. In *Teaching simultaneous interpretation into a "B" Language. Proceedings of EMCI Workshop*, Paris 20–21 September 2002, 60–62. Also available at: http://www.emcinterpreting.org/repository/pdf/EMCI-TeachingSimultaneousIntoB-vol1.pdf (accessed 26 April 2009).

Gerver, D. & Sinaiko, H. W. (Eds.) (1978). *Language interpretation and communication*. New York: Plenum Press.

Gerver, D., Longley, P., Long, J. & Lambert, S. (1984). Selecting trainee conference interpreters. *Journal of Occupational Psychology* 57, 17–31.

Gerver, D., Longley, P., Long, J. & Lambert, S. (1989). Selection tests for trainee conference interpreters. *Meta* 34 (4), 724–735.

Gran, L. & Dodds, J. (Eds.) (1989). *The theoretical and practical aspects of teaching conference interpretation*. Udine: Campanotto Editore.

Gringiani, A. (1990). Reliability of aptitude testing: A preliminary study. In L. Gran & C. Taylor (Eds.), *Aspects of applied and experimental research in conference interpretation*. Udine: Campanotto Editore, 42–53.

Harris, B. (1992). Teaching interpreting: A Canadian experience. In C. Dollerup & A. Loddegaard (Eds.), *Teaching translation and interpreting: Training, talent and experience*. Amsterdam/Philadelphia: John Benjamins, 259–268.

Kalina, S. (2000). Interpreting competences as a basis and a goal for teaching. *The Interpreters' Newsletter* 10, 3–32.

Keiser, W. (1965a). Admission dans les Ecoles d'interprétation. In AIIC, *Colloque sur l'enseignment de l'interprétation*, Paris 18–19 Décembre, 3–8.

Keiser, W. (1965b). Definitions d'aptitude. In AIIC, *Colloque sur l'enseignment de l'interprétation*, Paris 18–19 Décembre, 28–37.

Keiser, W. (1978). Selection and training of conference interpreters. In D. Gerver & H. W. Sinaiko (Eds.), *Language interpretation and communication*. New York: Plenum Press, 11–24.

Köpke, B. & Nespoulous, J. (2006). Working memory performance in expert and novice interpreters. *Interpreting* 8 (1), 1–23.

Lambert, S. (1992a). Shadowing. *The Interpreters' Newsletter* 4, 15–24.

Lambert, S. (1992b). Aptitude testing for simultaneous interpretation at the University of Ottawa. *The Interpreters' Newsletter* 4, 25–32.

Lambert, S. & Meyer, I. (1988). Selection examinations for student interpreters at the University of Ottawa. *The Canadian Modern Language Review* 44 (2), 274–284.

Lederer, M. (1975). Le colloque sur les examens d'entrée aux Ecoles d'Interprètes. *AIIC Bulletin* 3 (2), 25–28.

Longley, P. (1989). The use of aptitude testing in the selection of students for conference interpretation training. In L. Gran & J. Dodds (Eds.), *The theoretical and practical aspects of teaching conference interpretation*. Udine: Campanotto Editore, 105–108.

López Gómez, M. J., Bajo Molina, T., Padilla Benítez, P. & Santiago de Torres, J. (2007). Predicting proficiency in signed language interpreting. *Interpreting* 9 (1), 71–93.

Mackintosh, J. (1999). Interpreters are made not born. *Interpreting* 4 (1), 67–80.

Moser-Mercer, B. (1985). Screening potential interpreters. *Meta* 30 (1), 97–100.

Moser-Mercer, B. (1994). Aptitude Testing for Conference Interpreting: Why, When and How. In S. Lambert & B. Moser-Mercer (Eds.), *Bridging the gap: Empirical research in simultaneous interpretation*. Amsterdam/Philadelphia: John Benjamins, 57–68.

Moser-Mercer, B. (1997). Methodological issues in interpreting research: An introduction to the Ascona workshops. *Interpreting* 2 (1/2), 1–11.

Moser-Mercer, B. (2000). Searching to define expertise in interpreting. In B. Englund Dimitrova & K. Hyltenstam (Eds.), *Language processing and simultaneous interpreting: Interdisciplinary perspectives*. Amsterdam/Philadelphia: John Benjamins, 107–131.

Pearl, S. (1995). Lacuna, myth and shibboleth in the teaching of simultaneous interpreting. *Perspectives* 3 (2), 161–190.

Pippa, S. & Russo, M. (2002). Aptitude and simultaneous interpretation: A proposal for a testing methodology based on paraphrase. In G. Garzone & M. Viezzi (Eds.), *Interpreting in*

the 21st Century: Challenges and opportunities. Amsterdam/Philadelphia: John Benjamins, 245–256.

Rejsková, J. (1999). Establishing a correlation between performance in consecutive interpreting and potentially good performance in simultaneous interpreting. *Folia Translatologica* 6, 41–60.

Roland, R. A. (1999). *Interpreters as diplomats: A diplomatic history of the role of interpreters in world politics*. Ottawa: University of Ottawa Press.

Russo, M. (1989). Text processing strategies: A hypothesis to assess students' aptitude for simultaneous interpreting. *The Interpreters' Newsletter* 2, 57–64.

Russo, M. (1991). Ipotesi valutativa dell'attitudine all'interpretazione simultanea: risultati preliminari. *Zeta* 12 (14/15/16), 388–399.

Russo, M. (1993). Testing aptitude for simultaneous interpretation: Evaluation of the first trial and preliminary results. *The Interpreters' Newsletter* 5, 68–71.

Russo, M. (1995). L'interpretazione simultanea ed i test attitudinali. *Il Traduttore nuovo* 44 (1), 37–46.

Russo, M. & Pippa, S. (2004). Aptitude to interpreting: Preliminary results of a testing methodology based on paraphrase. *Meta* 49 (2), 409–432.

Schweda Nicholson, N. (2005). Personality characteristics of interpreter trainees: The Myers-Briggs Type Indicator (MBTI). *The Interpreters' Newsletter* 13, 109–142.

Skaaden, H. (1999). Lexical knowledge and interpreter aptitude. *International Journal of Applied Linguistics* 9 (1), 77–97.

Shaw, S. & Hughes, G. (2006). Essential characteristics of sign language interpreting students. *Interpreting* 8 (2), 195–221.

Sofr, O. (1976). Test of aptitude for interpreting. *L'interprète* 31 (3/4), 3–8.

Sunnari, M. (2002). Aptitude tests and selection criteria for interpreting students. In *Teaching simultaneous interpretation into a "B" language: Proceedings of EMCI Workshop*, Paris, 23–26. Also available at: http://www.emcinterpreting.org/repository/pdf/EMCI-TeachingSimultaneousIntoB-vol1.pdf (accessed 26 April 2009).

Szuki, A. (1988). Aptitudes of translators and interpreters. *Meta* 33 (1), 109–116.

Thiery, C. (1990). The sense of situation in conference interpreting. In D. Bowen & M. Bowen (Eds.), *Interpreting — Yesterday, Today, and Tomorrow*. Binghamton, NY: SUNY, 40–43.

Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: The University of Chicago Press.

Timarová, Š. & Ungoed-Thomas, H. (2008). Admission testing for interpreting courses. *The Interpreter and Translator Trainer* 2 (1), 29–46.

Timarová, Š. & Ungoed-Thomas, H. (2009). The predictive validity of admission tests for conference interpreting courses in Europa: A case study. In C. V. Angelelli & H. E. Jacobson (Eds.), *Testing and assessment in translation and interpreting studies*. Amsterdam/Philadelphia: John Benjamins, 225–246.

Tryuk, M. (2002). Les tests d'aptitude et leur role dans la formation des intérprètes de conference. In B. Lewandowska-Tomaszczyk & M. Thelen (Eds.), *Translation and Meaning*, Part 6. Maastricht: UPM, 421–434.

University of Westminster (2010). http://www.westminster.ac.uk/schools/humanities/postgraduate/modern-and-applied-languages/ma-conferencing-interpreting (accessed 26 October 2010).

Learning styles, motivation and cognitive flexibility in interpreter training

Self-selection and aptitude

Šárka Timarová and Heidi Salaets
Lessius University College*

Admission testing for conference interpreter training programmes traditionally focuses on skills directly related to the interpreting skills, and while soft skills, such as motivation, are recognised as important, they are not systematically tested or researched. The present study attempts to address this gap by exploring three traits and abilities, namely learning styles, motivation and cognitive flexibility, and to relate them to students' self-selection for interpreting and to their success on final exams. Three tests were used to compare a group of self-selected interpreting students and applicants (n = 32) and a subgroup of conference interpreting students (n = 14) to a control group of undergraduate students (n = 104), from among whom the majority of Lessius University College interpreting students are recruited: the Inventory of Learning Styles (Vermunt & Rijswijk 1987), the Achievement Motivation Test (Hermans 1968/2004) and the Wisconsin Card Sorting Test (Grant & Berg 1948). The results show that self-selected interpreting students are cognitively more flexible and are less negatively affected by anxiety. Compared to the control group, successful conference interpreting students, but not unsuccessful students, are cognitively more flexible and benefit more from some level of anxiety. Moreover, all conference interpreting students are less affected by stress than the control group and seem to have more clearly developed learning preferences.

Introduction

When asked whether anyone can be trained as an interpreter, Helge Niska, a veteran interpreter trainer at Stockholm University, replied: "In principle I would say yes, but in practice time constraints and limitations on financial resources make

* At the time the study reported here was carried out, Lessius University College had not yet been fully integrated into KU Leuven. The former designation is therefore used throughout this paper.

it advisable to select the people who need least training" (Niska 2002). While the "born or made" debate has definitely shifted towards "made" (Mackintosh 1999; Kalina 2000), not everyone remains as optimistic as Niska, certainly not in the spoken language conference interpreting field. What everyone would agree with is the need for selection. Admission testing for interpreter training programmes is motivated as much by the belief that a certain aptitude for interpreting is needed in order to become an interpreter, as by practical considerations, such as time and financial constraints or demand far exceeding the number of available places. All these factors call for as efficient an admission procedure as possible, but in reality little is still known about how successful student selection is. Most admission tests are intuitive (Lambert 1991; Moser-Mercer 1994; Russo & Pippa 2004), there is currently no reliable aptitude test (AIIC Training Committee 2006), and a number of authors have expressed scepticism about the very possibility of using an admission test to measure aptitude (Déjean Le Féal 1998; Taylor 1997; Alexieva 1993). Furthermore, there is a lack of systematic research into aptitude and admission testing (Moser-Mercer 1994) and the little research available suggests that current tests do not predict reliably whether a student will succeed or fail (Gringiani 1990; Tapalova 1990, cited in Sawyer 2004; Timarová & Ungoed-Thomas 2008, 2009). The issues involved were succinctly summed up by Kalina (2000: 13):

> There is as yet no clear evidence that the skills deemed necessary as prerequisites for successful training efforts are really the ones tested, and that it is those skills students must have before they go into interpreter training. Moreover, the tests do not as yet seem to be sufficiently objective and transparent to be used for decisions that affect the future (chances) of many students.

In effect, Kalina outlines three major tasks for research and training practice: defining the skills that need to be tested; finding a test that would appropriately assess them; and designing tests that provide reliable results. Most interpreter training admission tests focus on the so called hard skills, i.e. those directly related to language processing and interpreting, such as memory tests, summarization, short consecutives etc. (Timarová & Ungoed-Thomas 2008). Soft skills, such as steep learning curves (Moser-Mercer 1994), motivation (Gringiani 1990) and stress tolerance (Moser-Mercer 1984) have been suggested as desirable traits in students, but are rarely tested (Timarová & Ungoed-Thomas 2008). While these skills are probably not going to be determining factors, they may contribute to the skill acquisition process (Shaw et al. 2008) and may influence successful completion of a training programme (see also Moser-Mercer et al. (2000) for the distinction between the interpreting skill and interpreting skill acquisition). Indeed, López Gómez et al. (2007) report on a study in which soft skills were found to be weaker predictors than hard skills but did help to predict completion of training.

It seems relevant therefore to explore the role of soft skills and their possible inclusion in admission tests. We selected three such traits and abilities: learning styles, motivation and anxiety, and cognitive flexibility. Since our main interest lies in using the tests for purposes of aptitude and admission testing, we focused on two groups: interpreting programme students and interpreting programme applicants. In the first group, our main research question was whether tests of the selected traits predict successful completion of interpreter training. In the second group, our main goal was related to self-selection; that is, to the profile of the applicants.

Psychological traits: Learning styles, motivation and cognitive flexibility

Learning styles

Learning style is a complex concept encompassing students' content processing, regulation of learning, motivation for and goals of learning and a set of beliefs about how learning works, the roles of students and teachers, etc. (Vermunt & Vermetten 2004). A general ability to learn and acquire new skills was suggested as a factor in interpreter training (Moser-Mercer 1994) and was addressed from the trainer's perspective in terms of creating a suitable learning environment (Moser-Mercer 2008). Shaw et al. (2004) found that novice interpreting students transitioning from sign language learning to sign language interpreting stressed the importance of an open attitude to learning and of teacher support.

In a more general context of educational psychology, extensive studies of various student populations resulted in the specification of several basic learning styles and of their relation to a variety of factors. Vermunt (see for example Vermunt & Vermetten 2004) distinguishes four dimensions of a learning style: cognitive processing, metacognitive regulation of learning, overall learning conceptions and learning orientations. Cognitive processing refers to the more traditional view of learning, whereby a distinction is made between deeper and shallower processing of the content to be learned. Regulation of learning is concerned with a student's approach to learning in terms of control of learning activities using internal or external support. Learning conceptions refer to students' general beliefs about learning, how it works, how the learning environment functions, what the demands are etc. Finally, learning orientations refer to students' global attitude and motivation (Vermunt & Vermetten 2004). Elements from each dimension combine into an overall learning style. At an even higher level, Meyer et al. (1990) suggested the term *orchestrations* for complex individual or group patterns of learning styles.

A number of studies amongst student populations in higher education supported this classification and provided evidence that enhanced our understanding

of the learning styles. Personal learning orientations and strategies were found to be stable, although some adjustments were possible in response to specific course demands (Vermetten et al. 1999), perceived change in learning environment, such as during a period of study abroad (Wierstra et al. 2003) or following the acquisition of specific learning strategies (Lonka et al. 1997 cited in Vermunt & Vermetten 2004). An important finding that emerged from such studies was the concept of *dissonance*, which refers to a conflict between individual learning orientations, the learning strategies they apply and perceptions and/or demands of the learning environment. Instances of such dissonance were shown to be related to lower study success (Lindblom-Ylänne & Lonka 1999).

Learning styles and individual approaches to learning seem to be an important aspect of student experience. A general perception and conception of the learning process may play a contributing role in the context of an intensive one-year conference interpreting training programme, which, unlike a full university degree programme lasting three or four years, offers little room for finding one's way around. Some approaches to learning, which may work well in a traditional higher education institution at an undergraduate level, may not be suited to interpreter training. These would include reliance on extensive external support, such as textbooks, various study aids and detailed teacher guidance, or shallower forms of content processing, such as memorising or sequential progress through a piece of study text.

Motivation

Motivation is part of the highest level of performance control (Matthews et al. 2000) and together with capacity and opportunity forms the three most important determinants of work performance (Blumberg & Pringle 1982; see Moser-Mercer 2008 for a more detailed description of the interaction). Like learning styles, motivation too is psychologically a very complex construct which is difficult to measure objectively. Atkinson and Reitman (1958) suggested that motivation to perform a specific task is a product of two components: achievement motive and expectancy of goal attainment. Achievement motive is a general personal disposition towards applying oneself in order to achieve certain goals. Expectancy of goal attainment is related to the belief that a certain act is needed in order to achieve a goal. Motivation is the resulting state when the two components have been sufficiently aroused, and it follows that motivation differs in different situations. Hermans (1970) concluded that motivation and achievement motives are very complex and therefore it may be more appropriate to speak of measuring *an* achievement motive rather than *the* achievement motive. Nevertheless, some findings seem to be fairly robust.

Motivation, understood in more general terms as arousal, is closely related to stress and anxiety (Matthews et al. 2000:171). Hermans (1967, 1970) developed an instrument based on Atkinson's expectancy theory, in which he incorporated three scales: achievement motive and two scales representing two separate constructs of positive (facilitating) and negative (debilitating) anxiety. All three scales seem relevant in the context of interpreter training. Motivation as a general construct and in interaction with stress appears to play an important role in successful completion of training (Gringiani 1990; Moser-Mercer 2008). Trainers consider it an important trait in applicants, although doubts exist about the possibility of reliably measuring it in admission tests (Timarová & Ungoed-Thomas 2008). Shaw et al. (2004) found in interviews that interpreting students identified overcoming anxiety and taking risks as important factors in training.

Cognitive flexibility

Cognitive (mental) flexibility refers to readiness to change cognitive content and its attributes (Scott 1962) or the ability to switch to a different action as required by the situation and to avoid stereotypical behaviour (Hill 2004). The opposite, cognitive rigidity, refers to maintaining fixed images and their attributes and relations (Scott 1962). Cognitive flexibility is part of cognitive executive control and under different names, such as open-mindedness or the ability to learn quickly, it has been identified as desirable in interpreting students (Timarová & Ungoed-Thomas 2008; Moser-Mercer 1994). As such, it is an essential component of the ability to seek innovative solutions to problems and of adaptive expertise, both of which are considered important in interpreting (Moser-Mercer 2008).

The three constructs all seem to be highly relevant for interpreter training and empirical evidence in support of their role may provide further clues as to their potential for inclusion into admission testing procedures. Interpreter training programmes are less structured than traditional university programmes with their extensive pool of supportive learning materials and more tangible outcomes, such as correct answers to factual questions, and such lack of structure may provide an additional source of anxiety and require flexibility and self-reliance in learning. An effective approach to learning, stress tolerance and anxiety control are thus all desirable traits in interpreting students and interpreters (Timarová & Ungoed-Thomas 2008; Moser-Mercer 1994).

Methodology of aptitude and admission research

A methodological issue in interpreting aptitude research is related to population restriction, which is twofold. The first restriction lies in self-selection: it is quite possible that many of those who do not apply for an interpreter-training programme do possess the necessary aptitude, whereas not all of those who do apply have the aptitude to become interpreters (or at least not within the time period afforded by the training programme), which accounts for the need for admission tests. In effect, only those with an interest in interpreting (self-determined selection) and assumed aptitude (selection on the basis of admission tests) are admitted.

Aptitude and admission research typically looks at students at the beginning of their training and compares their performance on aptitude tests with their final exam performance. However, after the double selection described above, researchers have access to a very restricted population and evaluating aptitude becomes very difficult. In a nutshell, they are left with studying interpreting *students* but wish to come up with conclusions that will generalise to the population of interpreting *applicants*. These two populations, however, may be quite different and there is a danger of score range restriction. Selecting and studying only high scorers leads to lower correlations with the relevant criterion (Howell 2001: 283), in this case the interpreting course performance. Sackett et al. (2007) demonstrated through a simulation in a known population that a true correlation between a predictor and a criterion of $r = .5$ was reduced to as low as $r = .05$ when the range was restricted.

A solution to this problem is not feasible in practice. Ideally, a group of people would be subjected to interpreter training without being preselected in any way, and their performance on aptitude tests and interpreting tests could then be compared. Other, more pragmatic solutions are therefore required, such as analysing data from training programmes where admission tests are not legally allowed (removing the second selection based on admission tests), as is the case in Italy (Gringiani 1990). Another valuable source of data comes from programmes where there is a compulsory interpreting component for all students in early stages of their university education and where specialisation in two or more study tracks only takes place at a later stage. Such a programme exists for example in the Czech Republic (Rejšková 1999).

In the study reported here, we have tried to address both types of selection, using another approximation method. At Lessius University College in Belgium, most conference interpreting applicants and students are graduates of the Lessius undergraduate programme. By studying the group of third-year students towards the end of their studies, we are assuming that we will capture the unselected group

as a control group. By comparing conference interpreting applicants (self-selection) and students (selection on the basis of an admission test), we expected to gain more insight into the profiles of people who tend to self-select for interpreting and into the traits of successful and unsuccessful students.

Lessius study programmes and student groups description

Lessius University College in Belgium offers a three-year bachelors programme in applied language studies. The programme includes courses in foreign languages and/or Flemish sign language, mother tongue (Dutch), history and culture, discourse and communication studies etc., and provides all-round practical and academic preparation for advanced professional language users. After completing the bachelors programme, students choose among four one-year masters programmes: multilingual business communication, journalism, translation and (liaison) interpreting. Furthermore, Lessius offers a one-year postgraduate (post-masters) course in conference interpreting, where training is offered in consecutive and simultaneous interpreting in line with the AIIC criteria (AIIC Training Committee 2006). The training programme is designed for speakers of Dutch as a mother tongue and a knowledge of two foreign languages, with one out of English, French or German and one out of English, French, German, Spanish, Italian, Russian or Hungarian. Before the programme itself gets under way, a two-week preparatory course is offered in which applicants are taught the basic notions of interpreting, an introduction to note-taking and preparatory memory and consecutive exercises, and make a first attempt at simultaneous interpreting. The preparatory course serves both as a teaser for the applicants and as an opportunity for the trainers to meet the applicants and assess their interpreting potential over a period of two weeks. Next, applicants take an admission test that includes various aptitude exercises, a general knowledge test and a short consecutive test. Eligibility for the training programme is contingent on successful performance on this test. An alternative route exists for Lessius Master of (Liaison) Interpreting graduates, who are exempt from the preparatory course and need only take the admission test.

The majority of the conference interpreting programme applicants are Lessius graduates. Other applicants completed their master's degree elsewhere or participated in the Lessius transfer programme (an accelerated version of parts of the bachelors programme to help non-Lessius graduates meet the enrolment criteria for a given specialisation programme). Non-Lessius graduates typically completed a language-oriented programme (philology, translation) or another humanities degree. Only rarely do applicants have a substantially different background.

The third-year bachelors students therefore constitute a pool of future applicants and a population as close as possible to the selection stages (self-selection for the liaison interpreting masters programme in the fourth year and the conference interpreting postgraduate course in the fifth year). By studying the groups of third year students, it is possible to gain a better understanding of the population from which most of the interpreting students will be recruited.

Analysis 1: Self-selection

Method

Participants

Three groups of Lessius University College students participated in the study, which took place in the academic year 2007–2008. The first group were the applicants of the postgraduate conference interpreting programme (conference interpreters). The second group were students from the liaison interpreting programme (liaison interpreters). The third group consisted of students from the third year (the last year before specialisation) of an applied language studies undergraduate programme (control group). The demographic information may be found in Table 1. All participants, with the exception of one exchange student in the control group, were native speakers of Dutch. Several students reported previous training in interpreting, typically consisting of introductory interpreting classes taken during their student exchange study period abroad. The study was conducted during regular class time slots, but all participants were given the option of not participating. They received no financial remuneration for their participation.

Table 1. Participants' demographic information (2007–2008 cohort)

Group	n	Sex		Mean age (standard deviation)	Students reporting previous training in interpreting (% of total)	Completed higher education degree (bachelor, master)
		Male	Female (% of total)			
Conference interpreters	9	2	7 (78%)	23.2 (3.49)	0	Yes, all
Liaison interpreters	23	4	19 (83%)	21.4 (0.95)	6 (26%)	Yes, all
Control group	104	29	75 (72%)	21.3 (1.95)	6 (5.8%)	No

Materials

Learning styles

The Inventory of Learning Styles (ILS; Vermunt & Rijswijk 1987) is a self-report instrument comprising 120 statements with Likert-type responses. The responses are converted into scores on 16 scales in four domains: processing strategies, regulation strategies, learning orientation and conceptions of learning. Following findings and methodology in Vermunt & Vermetten (2004) and Wierstra et al. (2003), two main learning styles were considered: the *meaning-directed learning style* and the *reproduction-directed learning style*. The former is characterised by the learner's deep processing of content matter (relating it to other knowledge, drawing one's own conclusions, abstracting from details to a whole), self-regulation of learning (one's own planning of tasks, monitoring of progress, reflecting on one's own work) and taking responsibility for one's own learning. It was calculated as the sum of scores on the following scales: deep processing, self-regulation and construction of knowledge (31 items). The reproduction-directed learning style may be seen as an opposite style, where the learner shows preference for processing the content in a more stepwise manner (going through the material step-by-step, studying each element in detail and in isolation from the whole, memorising facts), for external regulation (learning on the basis of external support, such as study questions, assignments, teacher guidelines etc.) and for viewing learning as imposed externally. The reproduction-directed learning style was calculated as the sum of scores on the following scales: stepwise processing, external regulation and intake of knowledge (31 items). The remaining scales contribute to two other learning styles, *undirected* and *application-directed*, which are more frequently found in specific contexts (Vermunt & Vermetten 2004) and were not considered separately in the present analysis. Further, students were classified according to one of four orchestrations (complex patterns) of learning styles (Lindblom-Ylänne & Lonka 1999) as *both meaning high and reproduction high*, *both meaning low and reproduction low*, *high meaning/low reproduction* and *low meaning/high reproduction*, based on their above- or below-mean score for a given learning style. Selected elements of the instrument structure and sample items appear in Table 2. The description is based on the English version of ILS (Vermunt 1994). The instrument itself was administered in Dutch, the participants' native language, and the original version of the instrument.

Achievement Motivation Test

The Achievement Motivation Test (AMT; Hermans 1968/2004) is a self-report instrument. It contains 90 statements, for each of which the participants choose the most favored response (two to four options per statement). The responses are

Table 2. Selected scales of the Inventory of Learning Styles and sample items (Vermunt & Vermetten 2004; Vermunt 1994)

Domains (in italics) and scales	Domain description (in italics) and sample items	Number of items
Processing strategies	*Mental activities directly related to processing content, resulting in knowledge and understanding*	
Deep processing	I try to relate new subject matter to knowledge I already have about the topic concerned.	11
Stepwise processing	I memorise the meaning of every concept that is unfamiliar to me.	11
Regulation strategies	*Source of direction for the learning process*	
Self-regulation	I add something to the subject matter from other sources.	11
External regulation	I study according to the instructions given in the study materials or provided by the teacher.	11
Conceptions of learning	*Internal system of beliefs and knowledge about learning*	
Construction of knowledge	I should try to think up examples with the study materials of my own accord.	9
Intake of knowledge	I like to be given precise instructions as to how to go about solving a task or doing an assignment.	9

then converted into scores measuring three scales: achievement motive, facilitating anxiety and debilitating anxiety. Individuals who score high on the achievement motive scale are considered to attach great importance to performing to the best of their ability. Low scores, on the other hand, indicate that the individual is not achievement-driven, and is typically happy with a "good-enough" performance. High scores on the facilitating anxiety scale indicate that those individuals perform better when under slight stress. Some stress such as that associated with ill-defined and unstructured tasks may have a stimulating effect. Low scores indicate that the individual does not benefit from stress and unstructured tasks. Finally, the debilitating anxiety scale measures an individual's ability to withstand stress and stressful situations, such as exams. A high score indicates sensitivity to stress, which acts as an inhibitor and has a negative effect on performance. Low scores, on the other hand, indicate that such an individual's performance level is fairly stable and remains unaffected by stressful situations or a lack of structure in the task. The structure of the instrument and sample items appear in Table 3.

Cognitive flexibility

The Wisconsin Card Sorting Test (WCST; Grant & Berg 1948) was used as a measure of cognitive flexibility. The test consists of sorting a deck of 64 cards, one

Table 3. Scales of the Achievement Motivation Test and sample items (Hermans 1968/2004, 1970)

Scales	Sample items	Number of items
Achievement motive	I usually do a. much more than I resolved to do b. a bit more than I resolved to do c. as much as I resolved to do d. less than I resolved to do	44
Facilitating anxiety	I find that mild exam anxiety a. improves my performance b. has hardly any influence c. affects my performance in a negative way	19
Debilitating anxiety	If I imagine myself in a job interview then I see myself as a. calm b. slightly tense c. quite nervous	26

by one, into one of four slots. Upon placing each card in a slot, the participant receives feedback as to the correctness/incorrectness of the categorisation, but no feedback is offered as to the reasons why the particular categorisation was incorrect. Each card may be categorised according to at least three different rules, but participants are not told which rule is being applied. Moreover, the categorisation rule changes during the test, and participants must adjust to the new rule (after finding out for themselves which rule is being applied). This creates a task with a fairly 'fuzzy' structure, a high incidence of errors and the necessity of integrating fairly vague feedback.

Three measures were taken: the total number of errors, the total number of perseverative errors, and a learning-to-learn index. Perseverative errors are responses that are not correct, given the present rule, but would have been correct according to the previously applied rule. For the learning-to-learn index, the percentage of errors within each segment of the test (each application of a rule constitutes a new segment) was calculated, and then compared to the previous segment. The index was high if the percentage of errors consistently decreased and low when the participant made more and more errors.

Procedure

Participants were tested in groups of up to 30 (with the exception of the Wisconsin Card Sorting Test, which was administered individually). The tests (with the exception of WCST) were administered together with a battery of language processing ability tests (to be reported separately) either in two one-hour sessions

(the conference interpreting group) or in a single two-hour session (the liaison interpreting group, the control group). Individual appointments were made for the WCST. Not all students were available for individual testing and the sample size for WCST is thus lower. The two interpreting student groups were tested at the beginning of their interpreter training and the control group was tested at the beginning of the second semester in their third year of study (one semester before the end of their bachelors studies and their selection of master's specialisation). At the time of testing, participants had received no training in interpreting at Lessius. For each of the tests (ILS, AMT), participants received a booklet with test instructions, tests and space for providing responses. Before each test, the researcher went through the task instructions with the participants and gave them an opportunity to ask questions. The tests were administered in the same order for all groups. The order of tasks was the following: AMT, ILS, individual appointment for WCST. The testing session was paced by the researchers.

Results

For analyses concerned with self-selection, the two groups of interpreting students (conference interpreters, liaison interpreters) were considered as one group of students who made a decision to opt for interpreter training. The remaining group of bachelors students served as a control group from which the majority of interpreting students are recruited. Tables 4 and 5 provide descriptive statistics for each instrument and for the two groups: bachelors students as the control group, and interpreting students, consisting of conference interpreting programme applicants and liaision interpreting students. To maximise the use of available data, cases with missing values were excluded, analysis by analysis, which led to variable sample size per test. *T*-tests were carried out to compare the interpreting group to the control group. The results are shown in Table 4. On the WCST, there was a significant difference between the two groups on the number of perseverative errors. The mean score was 3.9 perseverative errors (*SD* 3.05) and 6.8 perseverative errors (*SD* 5.39) for the interpreting group and the control group, respectively, with a significant difference between the groups ($t(61) = -2.70$, $p < .01$, two-tailed). For the total number of errors, the interpreting group had a mean score of 15.8 errors (*SD* 7.50) and the control group a mean score of 20.1 (*SD* 10.39) with a marginally significant difference between the groups ($t(61) = -1.91$, $p = .06$, two-tailed). The learning-to-learn index did not show a significant difference. On the ILS the interpreting group and the control group scored comparably on both the meaning-directed and the reproduction-directed learning style measures. On the AMT, the mean score on the achievement motive scale was 23.1 (*SD* 8.43) for the interpreting group and 20.6 (*SD* 6.87) for the control group. The difference between the

Table 4. Means and standard deviations by group and test, and differences between the groups

	Student group										
	Interpreting			Control			Total				
	n	Mean	*SD*	*n*	Mean	*SD*	*N*	Mean	*SD*	*t*	*p*
Wisconsin Card Sorting Test											
Errors	28	15.8	7.50	35	20.1	10.39	63	18.2	9.40	−1.91	.06
Perseverative Errors	28	3.9	3.05	35	6.8	5.39	63	5.5	4.70	−2.70	<.01
Learning-to-Learn Index	26	−5.4	11.23	28	−1.0	12.89	54	−3.2	12.21	−1.33	.19
Inventory of Learning Styles											
Meaning-directed	32	96.0	13.37	99	92.9	16.02	131	93.7	15.42	1.00	.32
Reproduction-directed	32	100.8	13.59	99	102.8	13.15	131	102.3	13.23	−.73	.47
Achievement Motivation Test											
Achievement Motive	32	23.1	8.43	103	20.6	6.87	135	21.2	7.31	1.69	.09
Debilitating Anxiety	32	12.8	5.33	103	15.7	5.94	135	15.0	5.91	−2.44	.02
Facilitating Anxiety	32	11.5	4.94	103	10.6	4.38	135	10.8	4.51	.91	.38

Table 5. Frequencies of learning style orchestrations by group

	Student group		
	Interpreting	Control	Total
	n	*n*	*N*
Orchestrations			
Low meaning/reproduction	6	32	38
High meaning/low reproduction	11	21	32
Low meaning/high reproduction	8	15	23
High meaning/reproduction	7	31	38
Total	32	99	131

two groups was marginally significant with $t(133) = 1.69$, $p = .09$, two-tailed). The mean score on the debilitating anxiety scale was 12.8 (*SD* 5.33) for the interpreting group and 15.7 (*SD* 5.94) for the control group, with a significant difference between the two groups ($t(133) = -2.44$, $p = .02$, two-tailed). There was no significant difference between the two groups on the facilitating anxiety scale. Next, the learning style orchestrations of the two groups were compared. Frequencies of membership in the four categories appear in Table 5. A chi-square test showed no significant differences in the distribution between the two groups.

Analysis 2: Prediction of successful completion of interpreter training

Of the 32 participants who self-selected for conference or liaison interpreting, 14 continued to study in the conference interpreting training programme and sat their final exams. This cumulative sample comprised seven students from the 2007–8 conference interpreting cohort and seven students who attended the liaison interpreting programme in 2007–8 and progressed into the conference interpreting programme in 2008–9. In this next section, we will explore the same tests and their ability to predict successful completion of conference interpreting training.

Method

Participants

The participants were 14 conference interpreter training programme students: 11 (79%) females and three males; the mean age was 23.1 years (*SD* 4.29). The participants were divided into two groups, pass and fail, depending on their overall final interpreting exam result.

Materials and procedure

Participants' scores from the same three tests (ILS, AMT, WCST) were included. Additionally, a measure of their interpreting performance was taken, consisting of a simple pass/fail grade for their overall final interpreting exam result.

The final exam is taken at the end of the one-year conference interpreter training programme. Students are tested in both consecutive and simultaneous interpreting for each of their two languages. The tests are administered individually by a jury of 10 to 15 members composed of (a) Lessius interpreting trainers, (b) external examiners. Jury members are professional interpreters with working languages corresponding to those of the student and Dutch language teachers (who assess the quality of the target language). Each student receives a different text, which is delivered by an invited native speaker (often an interpreter as well). Consecutive interpreting with notes is approximately six minutes long and simultaneous interpreting approximately 12 minutes long. Students are judged on the quality of content and form and are awarded a pass or fail mark for each of the four tests. The performance is reviewed by a jury, which then determines the mark. To graduate, the student must receive a pass mark on all four tests. Failed students may resit the exam once after the summer break.

For the purposes of this study, we considered only the final result, i.e. an overall pass or fail. The fail mark does not distinguish between students who failed all tests and those who failed only one. Students who had failed at the first attempt

but passed on a resit (which is still considered a part of the regular academic year) were included as passing.

Results

The descriptive statistics for the two groups of students are summarised in Table 6. A total of eight students passed their final interpreting exam, and six students failed. In order to maximise the use of data, missing cases were excluded, analysis by analysis, which led to variable sample sizes for individual tests. A two-tailed *t*-test was carried out to test for differences between the pass and fail groups; none of the tests reached significance. Next, we compared the two groups separately to the control group of bachelors students on each test (Table 6) and on learning style orchestrations (Figure 1). A two-tailed *t*-test demonstrated that passing students made significantly fewer errors than control students on the WCST (means 13.0 (*SD* 5.00) and 20.1 (*SD* 10.39) respectively, $t(40) = 2.75$, $p = .01$) and also fewer perseverative errors (means 2.4 (*SD* 1.62) and 6.8 (*SD* 5.39) respectively, $t(40) = 3.96$, $p < .001$). Passing students also reported marginally significantly lower scores on the debilitating anxiety scale of AMT (means 12.6 (*SD* 4.41) and 15.7 (*SD* 5.94) respectively, $t(109) = 1.86$, $p < .1$) and significantly higher scores on the facilitating

Table 6. Means and standard deviations by group and test, and differences between each group and the control group

	Student group									
	Pass					**Fail**				
	n	**Mean**	***SD***	**Difference from control**		***n***	**Mean**	***SD***	**Difference from control**	
				t	***p***				***t***	***p***
Wisconsin Card Sorting Test										
Errors	7	13.0	5.00	2.75	.01	6	19.7	9.71	.09	.93
Perseverative Errors	7	2.4	1.62	3.96	<.001	6	5.0	3.41	.78	.44
Learning-to-Learn Index	7	−3.5	13.44	.45	.65	5	−10.6	16.18	1.47	.15
Inventory of Learning Styles										
Meaning-directed	8	95.5	10.03	−.47	.64	6	100.8	9.17	−1.22	.23
Reproduction-directed	8	97.3	13.87	1.14	.26	6	92.8	14.29	1.79	.08
Achievement Motivation Test										
Achievement Motive	8	23.7	4.20	−1.26	.21	6	21.0	8.29	−.13	.90
Debilitating Anxiety	8	12.6	4.41	1.86	<.1	6	9.2	4.96	2.64	.01
Facilitating Anxiety	8	13.9	3.91	−2.05	.04	6	12.5	5.79	−1.02	.31

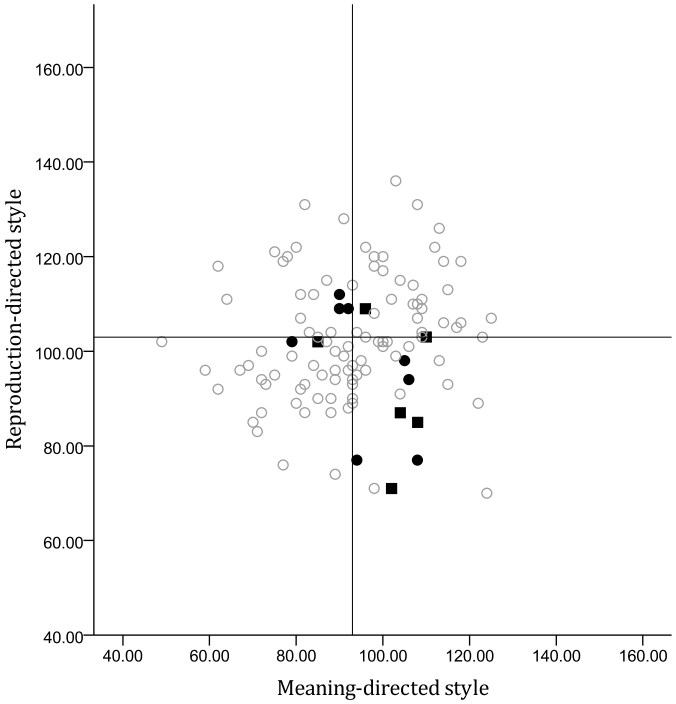

Figure 1. Learning style orchestrations in passing ● and failing ■ students, compared to the control group of bachelors students ○

anxiety scale of AMT (means 13.9 (*SD* 3.91) and 10.6 (*SD* 4.38) respectively, $t(109) = -2.05$, $p = .04$). No other test yielded significant differences between the passing and the control group. The failing students reported a marginally significantly lower score on the reproduction-directed learning style than the control group (means 92.8 (*SD* 14.29) and 102.8 (*SD* 13.15) respectively, $t(103) = 1.79$, $p = .08$) and a significantly lower score on the debilitating anxiety scale of AMT (means 9.2 (*SD* 4.96) and 15.7 (*SD* 5.94), $t(107) = 2.64$, $p = .01$). Due to the very small sample size of the passing and failing students, tests were not carried out on the orchestrations of learning styles and only visual representation of the results is presented in Figure 1, which shows a scatterplot of the meaning-directed and reproduction-directed learning styles plotted against each other. The vertical and horizontal lines are drawn at the means for the meaning-directed and reproduction-directed learning styles in the control group of bachelors students, dividing the space into quadrants representing the four orchestrations described in the method section: low meaning/low reproduction, high meaning/low reproduction, high meaning/high reproduction and low meaning/high reproduction. Passing students are represented in three quadrants (low meaning/low reproduction, high meaning/low reproduction and low meaning/high reproduction). Failing students

are also represented in three quadrants (low meaning/low reproduction, high meaning/low reproduction and high meaning/high reproduction).

Discussion

The aim of this paper was to explore the role of the so-called soft skills, such as motivation and learning, in student interpreters' success in completing interpreter training programmes. These skills were previously hypothesised as being contributing factors (Timarová & Ungoed-Thomas 2008; Shaw et al. 2008), but so far not systematically researched. This study focused on three such skills, namely learning styles, cognitive flexibility and motivation and anxiety, and explored them in two ways. First, we looked at their possible role in students' self-selection for interpreter training, and second, at the ability of several tests of the selected skills to predict successful completion of training.

As for self-selection, individuals opting for an interpreter training programme were found to differ from a control group of third-year applied language studies students on two tests. Self-selected interpreting students were found to show higher cognitive flexibility, as measured by a lower number of errors and perseverative errors on the Wisconsin Card Sorting Test (WCST), and to report higher achievement motive and lower debilitating anxiety on the Achievement Motivation Test (AMT). Higher cognitive flexibility would indicate that those individuals are, in general terms, more ready to cope with situations requiring innovative solutions to problems, problem solving and adaptability and to perform better in ill-defined domains with few if any fixed rules and structure. The WCST, more specifically, offers such a situation in that it requires the participant to respond to stimuli without being offered any instructions as to what constitutes a correct action. Moreover, the rules that guide decisions change throughout the task without the participant's knowledge. The errors measure reflects participants' ability to successfully navigate in such a 'fuzzy' situation. The perseverative errors measure reflects the speed at which participants are able to adjust to the new rules once they find out that the old rules no longer apply. Interpreter training can be seen as an ill-defined domain, requiring students to make global (process-based) adjustments to their approach to interpreting on the basis of often very local (product-based) feedback. The two measures were found to be related to students' success in completing the training programme. Specifically, students who successfully completed conference interpreter training were found to make fewer errors and fewer perseverative errors than the control group of bachelors students. Also the failing students made fewer errors and perseverative errors than the control group, but their mean scores did not significantly differ from those of the controls.

As for motivation and anxiety, the administered test measured self-reported scores on three scales: achievement motive (general preference for achievement of goals and task completion) and two types of anxiety: debilitating and facilitating. Both are related to people's response to situations which may induce some level of stress. Debilitating anxiety has a negative effects on performance and facilitating anxiety benefits performance. Students who self-selected for interpreting had reported marginally higher scores on the achievement motive scale and a lower score on the debilitating anxiety scale. The achievement motive suggests a certain level of drive to perform and, taken together with higher cognitive flexibility, may indicate individuals who actively seek novel and challenging situations. The lower score on the debilitating anxiety scale, on the other hand, suggests greater tolerance of stress, again a trait highly desirable in interpreters (Moser-Mercer 1994, Jiménez & Pinazo 2001).

Turning to the AMT as a predictor of student success, both passing and failing conference interpreting students had a higher mean score on the achievement motive scale (with passing students scoring higher than failing students), but this difference did not reach significance. The debilitating anxiety scale, on the other hand, yielded a similar result in that both passing and failing students had lower scores than the control group, although the passing students only marginally so. Interestingly, the failing students were the group with the lowest scores. This finding goes against the more intuitive expectation that failing students would be more affected by debilitating anxiety than passing students and provides some intriguing food for thought. It is possible that passing students (who also scored higher on facilitating anxiety, to be discussed next) develop better coping techniques to counteract the negative effects of anxiety, or that other factors which come into play attenuate it. This explanation is consistent with the findings reported by Jiménez and Pinazo (2001, 2002), who did not find a negative effect of anxiety on student consecutive interpreting performance even though students reported medium-high to high levels of anxiety. Passing students were also the only group that differed significantly from the control group in terms of facilitating anxiety. This type of anxiety can be thought of, in layman's terms, as excitement, arousal, perhaps even adrenaline. Interpreting is stressful and so is interpreter training. The finding that successful students are those who seem to benefit from a certain level of anxiety is very much in keeping with many trainers' expectations.

Finally, we explored students' learning preferences using the Inventory of Learning Styles (ILS) test. We calculated three measures: two for learning styles and one for more complex patterns of learning. In the analysis of self-selection, we did not find any difference between students who self-select for interpreting and the control group. In the group of conference interpreting students, the failing students scored marginally lower on their preference for the reproduction-directed

learning style. This would suggest they are less dependent on a structured learning environment and external teacher support. In this respect the higher scores of the successful students came as a surprise and appeared to suggest the need for further research. Taking a look at the learning style orchestrations shown in Figure 1, we note with interest that conference interpreting students are not equally represented in all quadrants. With the exception of the high meaning/low reproduction quadrant, conference interpreting students seem to cluster around the means (the intersecting axes) of the control group. This also applies to the four cases of students in the two least represented quadrants — low meaning/low representation and high meaning/high representation. Unfortunately our sample was too small to allow statistical analyses, but based on the visual inspection, we propose the following interpretation: The instrument allows for a calculation of four different styles, only two of which were calculated, as these are most representative of regular university students. It is possible that the control group students scoring low on both meaning-directed and reproduction-directed learning styles have a different dominant learning style than the two considered here. From the perspective of the two selected styles, this quadrant represents individuals with a low preference for either. In that respect, it is encouraging to see that conference interpreters have clearer preferences. Similarly, the high meaning/high reproduction quadrant represents students who scored high on two ostensibly conflicting styles (cf. Lindblom-Ylänne & Lonka 1999 for a more detailed discussion of conflicting styles). This may indicate a certain confusion and uncertainty in approaching learning difficulties and learning itself. These implications were not explored in the present study, and are only being offered as a possible explanation. The remaining two quadrants, high meaning/low reproduction and low meaning/high reproduction, represent more clearly defined learning style preferences, and this is where we find most conference interpreting students. While we cannot support this claim with a more detailed analysis we would suggest that in comparison to the control group, conference interpreting students have fewer conflicting learning preferences and patterns than the control group, although there does not seem to be a clear difference between passing and failing students.

In conclusion, we were able to detect some patterns and differences in soft skills both in terms of self-selection and in terms of prediction of successful completion of a conference interpreter training programme. By comparing self-selected interpreting students to a control group of third-year bachelors students, we found that those who self-select for interpreting are cognitively more flexible, are more achievement-driven and suffer less from stress. Similarly, successful conference interpreting programme graduates were found to be cognitively more flexible, to suffer less from stress and to benefit more from positive anxiety than the control group. Unsuccessful conference interpreting students, on the other hand,

did not show any difference in cognitive flexibility and differed from the control group only in that they too suffered less from stress. Both groups of conference interpreting students have also shown a clearer preference for a single dominant learning style. Given the very small size of our samples, these patterns and differences should be viewed as tendencies rather than facts and need to be confirmed and elaborated upon by further research. That said, the findings do seem to provide preliminary evidence that soft skills may indeed be important, if complementary, contributing factors of interpreting aptitude.

References

AIIC Training Committee (2006). Conference interpreting training programmes: Best practice. http://www.aiic.net/ViewPage.cfm?page_id=60 (accessed 29 June 2009).

Atkinson, J. W. & Reitman, W. (1958). Performance as a function of motive strength and expectancy of goal attainment. In J. W. Atkinson (Ed.), *Motives in fantasy, action and society*. Princeton: Van Nostrand, 278–287.

Alexieva, B. (1993). Aptitude tests and intertextuality in simultaneous interpreting. *The Interpreters' Newsletter* 5, 8–12.

Blumberg, M. & Pringle, C. (1982). The missing opportunity in organizational research: Some implications for a theory of work performance. *Academy of Management Review* 7, 560–569.

Carroll, J. B. (1962). The prediction of success in intensive foreign language training. In R. Glaser (Ed.), *Training research and education*. Pittsburgh: University of Pittsburgh Press, 87–136.

Déjean Le Féal, K. (1998). Non nova, sed nove. *The Interpreters' Newsletter* 8, 41–49.

Gerver, D., Longley, P. E., Long, J. & Lambert, S. (1989). Selection tests for trainee conference interpreters. *Meta* 34 (4), 724–735.

Grant, D. A. & Berg, E. A. (1948). A behavioural analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card sorting problem. *Journal of Experimental Psychology* 38, 404–411.

Gringiani, A. (1990). Reliability of aptitude testing: a preliminary study. In L. Gran & C. Taylor (Eds.), *Aspects of applied and experimental research on conference interpretation*. Udine: Campanotto Editore, 42–53.

Haston, A. G. (1974). *A study of the combined effects of intrinsic and extrinsic motivation on the performance of an auditory vigilance task*. Unpublished PhD dissertation, Texas Tech University.

Hermans, H. (1967). *Motivatie en prestatie*. [Motivation and achievement]. Amsterdam: Swets en Zeitlinger.

Hermans, H. J. M. (1968/2004). *Prestatie Motivatie Test* [Achievement Motivation Test]. Lisse: Harcourt Test Publishers.

Hermans, H. J. M. (1970) A questionnaire measure of achievement motivation. *Journal of Applied Psychology* 54 (4), 353–363.

Hill, E. L. (2004). Executive dysfunctions in autism. *Trends in Cognitive Sciences* 8 (1), 26–32.

Howell, D. (2001). *Statistical methods for psychology.* Pacific Grove: Wadsworth.

Jiménez Ivars, A. & Pinazo Calatayud, D. (2001). "I failed because I got very nervous." Anxiety and performance in interpreter trainees: An empirical study. *The Interpreters' Newsletter* 11, 105–118.

Jiménez Ivars, A. & Pinazo Calatayud, D. (2002). Aptitudes necesarias en la formación de intérpretes: un estudio exploratorio [Competences needed in interpreting training: an exploratory study], *Quaderns. Revista de traducció* 8, 77–97. http://ddd.uab.es/pub/quaderns/11385790n8p77.pdf (accessed 29 June 2009)

Kalina, S. (2000). Interpreting competences as a basis and a goal for teaching. *The Interpreters' Newsletter* 10, 3–32.

Lambert, S. (1991). Aptitude testing for simultaneous interpretation at the University of Ottawa. *Meta* 36 (4), 586–594.

Lindblom-Ylänne, S. & Lonka, K. (1999). Individual ways of interacting with the learning environemnt — are they related to study success? *Learning and Instruction* 9, 1–18.

López Gómez, M. J., Bajo Molina, T., Padilla Benítez, P. & Santiago de Torres, J. (2007). Predicting proficiency in signed language interpreting. *Interpreting* 9 (1), 71–93.

Mackintosh, J. (1999). Interpreters are made not born. *Interpreting* 4 (1), 67–80.

Matthews, G., Davies, D. R., Westerman, S. J. & Stammers, R. J. (2000). *Human performance. Cognition, stress and individual differences*. Hove: Psychology Press.

Meyer, J. H. F., Parsons, P. G. & Dunne, T. T. (1990). Individual study orchestrations and their association with learning outcome. *Higher Education* 20, 67–89.

Moser-Mercer, B. (1984). Testing interpreting aptitude. In W. Wills & G. Thome (Eds.), *Translation theory and its implementation in the teaching of translation and interpretation*. Tübingen: Gunter Narr, 318–325.

Moser-Mercer, B. (1994). Aptitude testing for conference interpreting: Why, when and how. In S. Lambert & B. Moser-Mercer (Eds.), *Bridging the gap: Empirical research in simultaneous interpretation*. Amsterdam/Philadelphia: John Benjamins, 57–68.

Moser-Mercer, B. (2008). Skill acquisition in interpreting. *The Interpreter and Translator Trainer* 2 (1), 1–28.

Moser-Mercer, B., Frauenfelder, U. H., Casado, B. & Künzli, A. (2000). Searching to define expertise in interpreting. In B. Englund Dimitrova & K. Hyltenstam (Eds.), *Language processing and simultaneous interpreting*. Amsterdam/Philadelphia: John Benjamins, 107–131.

Niska, H. (2002). Community interpreter training. Past, present, future. In G. Garzone & M. Viezzi (Eds.), *Interpretig in the 21st century*. Amsterdam/Philadelphia: John Benjamins, 133–144.

Rejšková, J. (1999). Establishing a correlation between performance in consecutive interpreting and potentially good performance in simultaneous interpreting. *Folia Translatologica* 6, 41–60.

Russo, M. & Pippa, S. (2004). Aptitude to interpreting: Preliminary results of a testing methodology based on paraphrase. *Meta* 49 (2), 409–432.

Sackett, P. R., Lievens, F., Berry, C. M. & Landers, R. N. (2007). A cautionary note on the effects of range restriction on predictor intercorrelations. *Journal of Applied Psychology* 92 (2), 538–544.

Sawyer, D. (2004). *Fundamental aspects of interpreter education*. Amsterdam/Philadelphia: John Benjamins.

Scott, W. A. (1962). Cognitive complexity and cognitive flexibility. *Sociometry* 25 (4), 405–414

Setton, R. (1994). Experiments in the application of discourse studies to interpreter training. In C. Dollerup & A. Lindegaard (Eds.), *Teaching translation and interpreting 2. Insights, aims, visions*. Amsterdam/Philadelphia: John Benjamins, 183–198.

Shaw, S., Grbic, N. & Franklin, K. (2004). Applying language skills to interpretation. *Interpreting* 6 (1), 69–100.

Shaw, S., Timarová, Š. & Salaets, H. (2008). Measurement of cognitive and personality traits in determining aptitude of spoken and signed language interpreting students. In L. Roberson & S. Shaw (Eds.), *Proceedings of the17th National Convention of the Conference of Interpreter Trainers: Putting the pieces together: A collaborative approach to educational excellence*. CIT, 91–109.

Sunnari, M. (2002). Aptitude tests and selection criteria for interpreting students. In *Teaching simultaneous interpretation into a "B" language. Proceedings of EMCI Workshop*, 23–26.

Taylor, C. (1997). Course profile: Degree in conference interpreting/translation. *The Translator* 3 (2), 247–260.

Timarová, Š. & Ungoed-Thomas, H. (2008). Admission testing for interpreting courses. *The Interpreter and Translator Trainer* 2 (1), 29–46.

Timarová, Š. & Ungoed-Thomas, H. (2009). The predictive validity of admission tests for conference interpreting courses: A case study. In C. Angelelli & H. E. Jacobson (Eds.), *Testing and assessment in translation and interpreting studies*. Amsterdam/Philadelphia: John Benjamins, 225–245.

Vermetten, Y. J., Lodewijsk, H. G. & Vermunt, J. D. (1999). Consistency and variability of learning strategies in different university courses. *Higher Education* 37, 1–21.

Vermunt, J.D.H.M. (1994) *Inventory of Learning Styles in Higher Education*. Tilburg: Tilburg University.

Vermunt, J.D.H.M. & Van Rijswijk, F.A.W.M. (1987). Inventaris Leerstijlen [Inventory of Learning Styles]. Tilburg: K.U. Brabant/Open Universiteit.

Vermunt, J.D.H.M. & Vermetten, Y. J. (2004). Patterns in student learning: relationships between learning strategies, conceptions of learning, and learning orientations. *Educational Psychology Review* 16 (4), 359–384.

Wierstra, R. F. A., Kanselaar, G., van der Linden, J. L., Lodewijsk, H.G.L.C. & Vermunt, J. D. (2003). The impact of the university context on European students' learning approaches and learning environment preferences. *Higher Education* 45, 503–523.

A story of attitudes and aptitudes?

Investigating individual difference variables within the context of interpreting

Alexandra Rosiers, June Eyckmans and Daniel Bauwens
Ghent University*

Unlike in the early days of interpreter training, most student interpreters nowadays are still in the process of acquiring their target language(s), which raises questions as to which skills — linguistic as well as non-linguistic — may be required at the outset of interpreter training. This study focuses on individual difference variables and how these relate to interpreting students' performance. It aims to investigate the ways in which the profiles of translation and interpreting students differ by obtaining information regarding their self-perceived communication competence, self-perception of language skills, anxiety levels and integrative motivation. These profiles are then related to the students' sight translation performances, arguably a hybrid activity between translating and interpreting that is as cognitively demanding as simultaneous and consecutive interpreting. The students' performances were assessed by a 'blind judge' along two parameters: (1) overall interpreting performance and (2) fluency. The results suggest that the two groups indeed differ significantly with regard to some of the individual difference variables. However, no significant correlations between these variables and students' sight translation performance were found.

1. Introduction

In the early days of interpreter training, most student interpreters were bilinguals whose linguistic proficiency was taken for granted, and interpreter training consisted mostly in developing the technical and task-specific skills required to function as a professional interpreter (Gile 2001; Eyckmans 2007; Zannirato 2007, 2008). The focus then was on non-linguistic traits, which were believed to make

* At the time of the study reported here, the authors were affiliated with Erasmus University College Brussels, where the research was carried out.

up the profile of a gifted trainee (see below). Nowadays, however, many students enrolling in interpreter training in Belgium are still in the process of learning one (or more) of the languages they are planning to work with. This raises questions about the kinds of aptitude — linguistic or other — that renders some students more suitable for interpreter training.

Although the importance of adequate linguistic skills for both translators and interpreters is indisputable, sufficient proficiency in one's working languages does not in itself guarantee high quality translation or interpreting. Several non-linguistic competences may help distinguish a polyglot from a professional translator or interpreter. For professional translators, heuristic competence is considered essential; i.e. translators should be able to locate and consult appropriate sources of information efficiently. Computer skills and familiarity with modern translation software and translation technology are other relevant qualities, as are good editorial skills. For interpreters, powers of concentration, the optimal use of working memory and the effective use of anticipation strategies are crucial.

While it stands to reason that the divergence in translators' and interpreters' competences grows more pronounced in the course of their specific training, it is conceivable that some traits that distinguish interpreters from translators exist before training gets under way. Indeed, professional interpreters and translators themselves often claim that interpreters are more articulate and have more highly developed oral skills than translators (Seleskovitch 1978; Henderson 1980). The translator, by contrast, is typically perceived as a perfectionist with a perpetual desire to search for the perfect wording. Teachers of interpreting and translation courses often mention the distinct personality profiles of their students: those who choose to do the interpreting courses are viewed as more self-confident than those who opt for translation.

Although most interpreter training schools take non-linguistic skills such as "analytical ability," "clarity and precision," or "delivery style" into account when testing students for admission, very few seem to link these to underlying personality profiles. Since the early 2000s, researchers have taken up the issue of personality research in second language acquisition studies as well as in interpreting studies. The widely divergent results of these studies will be discussed in section two of this article.

The descriptions of admission tests on interpreter training schools' websites often fail to provide a clear notion of how the linguistic and non-linguistic skills of future interpreters should interlock. The website of the University of Turku, for example, states that "prospective students need aptitude for the interpreter's work" (Norrman 2009), but there is no clarification of what this implies and how it will be assessed on the admission test. On the website of the International Association of Conference Interpreters (AIIC), both linguistic and non-linguistic competences

are mentioned as important elements of good aptitude testing: "A well-designed diagnostic test (which may consist of a written and an oral portion) can assess linguistic competence, general knowledge, maturity, basic code-switching ability, resourcefulness, ability to cope with stress, etc. and shed some light on the level of preparedness of a candidate wishing to enter a postgraduate interpreter training programme" (AIIC Training Committee 2006). They do acknowledge, however, that no aptitude test has yet been demonstrated to generate results that are predictive of success. Gile (2001) suggests that admission testing should include not only linguistic competence and general cultural knowledge but also the candidate's personality. As far as we are aware, however, no selection committee uses standardized measures to determine specific personality traits. It seems that the preferred personality characteristics of interpreter trainees are left to the appraisal of the selection committee in question.

Although the number of empirical studies into the personality traits, attitudes or communicative competence of translators and interpreters is quite small, these concepts have been thoroughly researched in the domain of social psychology and foreign language acquisition. Over the past 30 years, researchers have identified a wide range of so-called individual difference (ID) variables, including intelligence, aptitude, attitudes, motivation, anxiety, risk taking, cognitive style, introversion/extraversion and ego permeability (openness). The "canonical" list of ID variables traditionally comprises three main categories: cognitive variables (i.e. learner styles), affective variables (i.e. motivation, attitudes and language anxiety) and personality variables (i.e. extraversion, ego permeability) (Johnson 2001). Of these, the cognitive and personality variables are at a more preliminary stage than some of the other areas of ID research.

In the present study we will compare the ID variables in students of translation and of interpreting. More specifically, we will investigate whether students of translation and students of interpreting differ in their levels of language anxiety, linguistic self-confidence and motivation. The research questions are as follows: (1) Do interpreting and translation students (at the start of their postgraduate training) differ from each other with regard to these ID variables and, if so, in what respect? (2) Are the students' ID profiles associated with their interpreting skills as measured by a sight translation task and, if so, in what way?

Two groups of Belgian third-year students following an academic Bachelor in Applied Linguistics preparing them for an MA in translation or interpreting were asked to complete standardized questionnaires targeting ID variables. All students had English as a major in their language curriculum. The students were subsequently tested on their interpreting skills, by means of a sight translation task. We then calculated whether the scores of the students on this task correlated with the profiles obtained from the various ID questionnaires.

2. Literature

2.1 ID variables in second language acquisition

It has long been observed that language learners vary considerably in their ultimate success in mastering a second language, which accounts for the fact that the study of IDs has been a major research area in second language acquisition research (Dörnyei 2005).

When it comes to the development of competence in a second or foreign language, it is widely accepted that both external (e.g. the social context) and internal factors play an important role (Clément & Gardner 2001). The internal factors include learner-specific ID variables, such as language aptitude, attitudes, motivation, language anxiety and linguistic self-confidence. Gardner and Clément (1990: 497) distinguish three categories of ID variables related to second language acquisition: (1) cognitive variables, such as language aptitude and language learning strategies; (2) attitudes and motivation, including integrativeness (see Section 3.2), attitudes towards the learning situation, and motivation (attitudes and motivation are sometimes referred to as affective variables); and (3) personality variables such as anxiety, sociability, self-confidence, cognitive style (e.g. "field dependence" versus "field independence"), extraversion and empathy.

In social psychology and in studies on first language acquisition, bilingualism and foreign language acquisition, many models have already been developed to relate attitude, motivation and personality traits such as extraversion and anxiety to language users' communicative competence and their language learning processes (for an overview, see Dörnyei 2005). Researchers agree that positive attitude, motivation and communicative competence contribute to the language learning process and are at the same time learning goals themselves. Learning a language not only leads to linguistic effects (i.e. being able to use the foreign language), but also to changes in attitude. Communication in the foreign language requires a willingness to use that language, for example. One of the crucial factors in the complex interaction between linguistic and non-linguistic variables in second language acquisition is thus the so-called "willingness to communicate" (MacIntyre et al. 1998), a construct originally used for research purposes in the US in the domain of communication studies.

While much of the literature argues that ID variables represent relatively stable traits in learners, some qualitative studies have shown that attributes of highly successful language learners seem to change throughout the language learning process over time (Leaver & Atwell 2002), which may be of particular interest to translation and interpreting trainees since they are generally considered highly successful language learners.

2.2 ID variables in translation and interpreting studies

While still in its infancy, ID research in translation and interpreting studies is gaining in importance. Paneth (1962) described a number of criteria for recognizing students with the potential for interpreting. These include the ability to concentrate, the quality of "split-mindedness" and the "right personality traits" (1962: 109). Unfortunately, these personality traits were not specified or elaborated upon. Henderson (1980) designed a study in which a group of translators and a group of interpreters were asked to list the personality traits they felt applied to themselves and the traits they had noticed in their colleagues belonging to the other group. The way the interpreters viewed themselves turned out to correspond to the way they were viewed by the translators, and vice versa. A typical translator was described as "a perfectionist, self-sufficient and fairly adaptable introvert, obviously interested in language and a range of other subjects, with limited ambition, liking routine, socially isolated and suffering from artistic frustration, who is at the same time a self-doubting, eccentric pedant" (Henderson 1980: 220). The author comments that this description is of course "the product of a monstrous oversimplification, but the judgements are those of his peers" (Henderson 1980: 220). A typical interpreter was described as "a self-reliant, articulate extrovert, quick and intelligent, a jack of all trades and something of an actor, superficial, arrogant, liking variety and at times anxious and frustrated" (Henderson 1980: 223). Other publications on this topic claimed that interpreters were "articulate", with a "knack for communicating" (Seleskovitch 1978: 78), "nerves of steel" (Henderson 1980; Keiser 1978; Seleskovitch 1978) and a high level of "self-confidence" (Henderson 1980, 1987).

Such studies, however, typically draw their conclusions from interviews in which the respondents express their views on the (ideal) personality for an interpreter, and are sometimes based entirely on introspection when the interviewees are themselves interpreters. Clearly, studies of a more quantitative nature would be welcome to confirm (or falsify) these claims about interpreters' personality traits. For example, extraversion has often been suggested as a personality trait commonly found in interpreters (Henderson 1980, 1987; Seleskovitch 1978; Szuki 1988), but a recent study (Schweda Nicholson 2005), in which the Myers-Briggs Type Indicator (MBTI) was used to determine interpreters' personality profiles, found no evidence in support of this suggestion. Also using the MBTI, Hubscher-Davidson (2009) found some correlations between a number of personality traits and the quality of the translation. The Intuitive/Sensate subscale in particular correlated significantly with the translation task; the five most successful students were all intuitors, whereas the five weakest participants were sensors.

3. ID variables examined in this study

In this study, we will confine the comparison of interpreters' and translators' traits to the variables of linguistic self-confidence, motivation and language anxiety. Let us briefly consider these three ID variables.

3.1 Linguistic self-confidence

The concept of linguistic self-confidence, introduced by Clément (1980), concerns one's perception of one's own communicative competence. Language learners' linguistic self-confidence may be influenced by the nature of their interaction with the foreign language community, and being a motivational subsystem, will affect the language learning process.

In our study, we investigate whether interpreting and translation students differ in their degree of linguistic self-confidence and whether there is a relation between this ID variable and the quality of their interpreting performance. We will distinguish between the students' linguistic self-confidence in general (regarding both oral and written skills) and their self-confidence regarding oral communication competence (oral proficiency) in particular.

A commonly used instrument for measuring linguistic self-confidence in oral communication is the Self-Perceived Communication Competence (SPCC) questionnaire (McCroskey & McCroskey 1988). Nagy (2005) found a positive correlation ($p < 0.01$) between language users' level of proficiency and their SPCC scores: language learners who perceived themselves as communicatively competent also tended to obtain higher scores on a proficiency test.

While the SPCC questionnaire measures respondents' evaluation of their *oral* skills, our study will also take students' self-perceived global linguistic competence (writing, listening and reading as well as speaking skills) into account. We have used the items targeting self-perception of language skills from Gardner's (1985a) Attitude/Motivation Test Battery, which will be illustrated further in section four.

In translation studies, Fraser (2000) investigated levels of confidence in high and low risk-takers and found correlations between self-confidence and successful translation.

3.2 Motivation

Motivation has been thoroughly researched within the field of foreign language acquisition and one of the most important authorities in the field describes it as "the combination of effort plus desire to achieve the goal of learning the language plus favourable attitude toward learning the language" (Gardner 1985b: 10). In

the literature, a distinction is often made between intrinsic and extrinsic motivation. When students are said to be intrinsically motivated, this means that they derive satisfaction from the task itself. Extrinsically motivated students, on the other hand, are motivated to perform the task because they anticipate some kind of reward (which may be immediate, e.g. a good exam mark, or longer term, e.g. job opportunities). As alternatives to the intrinsic vs. extrinsic distinction, social psychologists also use the terms integrative and instrumental motivation. Language students who are integratively motivated have a genuine interest in the language they wish to master and in the foreign language community and culture. Instrumentally motivated students study the foreign language for a practical reason (such as finding a job).

In the present study, it is students' integrative motivation that is measured. According to Gardner (1985b), this type of motivation has three components: (1) interest in foreign language and attitudes towards the foreign language community; (2) attitudes towards the learning situation, including the language teacher and the course; and (3) a desire to learn the language and attitudes toward learning this language in particular. In this study, integrative motivation is tested by means of items extracted from the Attitude/Motivation Test Battery (Gardner 1985a).

3.3 Language anxiety

Although language anxiety is generally conceived of in social psychology as a negative trait, both positive and negative relations between anxiety and performance have been identified. On the one hand, anxiety can be an incentive to act: for example, when students feel anxious about passing exams, they may study harder. On the other hand, anxiety may impair cognitive function, to the point of preventing anxious students from performing adequately; for example, a student who experiences high levels of anxiety may become too frightened to speak at an oral exam.

In the early days of foreign language acquisition research, tests of language anxiety were derived from general psychology. These tests were not geared towards the foreign language learning context and failed to show consistently significant correlations. More recent tests, by contrast, use measures of anxiety that are situation-specific, and incorporate reference to the language learning situation with sentences such as 'It embarrasses me to volunteer answers in our French class' (Gardner 1985a). Still, it stands to reason that language anxiety must be closely connected to general anxiety, fear of public speaking and stress.

Jiménez and Pinazo (2001) investigated the relation between participants' fear of public speaking and state anxiety — i.e. the temporary anxiety one experiences when faced with threatening demands or dangers — and the quality of their

interpreting performance (more specifically, a consecutive interpreting task). As expected, they found strong positive correlations between fear of public speaking and state anxiety, but, against their expectations, did not find a negative correlation between state anxiety and interpreting performance, which suggests that fear of public speaking need not impact the quality of interpreting work.

According to Crookall and Oxford (1991), however, anxiety does hinder cognitive functioning and has a negative effect on foreign language proficiency. Yan (2007) states that anxiety can also play an important role in the learning process, especially when trainees are asked to work with languages they have not yet mastered.

In our study, we will compare interpreting and translation students' language anxiety levels and see if these are associated with the quality of their interpreting performance. For this last variable we will once more use Gardner's Attitude/Motivation Test Battery, from which we have selected items focussing on language anxiety.

4. Experiment

4.1 Participants

A total of 35 undergraduate students majoring in English at the Applied Linguistics Department of Erasmus University College of Brussels, Belgium, took part in this study. Their mother tongue was Dutch. The students were beginning their third year of a four-year training programme to become translators or interpreters. In Belgium, it is only at the start of their fourth year (i.e. the start of their master's degree) that these students start training specifically as either interpreters (MA in interpreting) or translators (MA in translation). However, in their third year an optional preparatory course in interpreting is offered for those who wish to pursue interpreting as a profession. Twelve of the participants — referred to here as the interpreting students — had enrolled in this optional course. The remaining twenty-three students will be referred to as the translation students. The participants' proficiency in English at the time of the experiment was estimated to be between B2 and C1 levels of the Common European Framework of Reference for Languages (CEFR).[1]

1. The CEFR is a guideline which describes in a comprehensive manner (i) the competences necessary for communication, (ii) the related knowledge and skills, and (iii) the situations and domains of communication. CEFR defines levels of attainment in different aspects of its descriptive scheme with an illustrative descriptors scale (Council of Europe 2007).

4.2 Method

To obtain information about the selected ID variables in our test groups, we asked the participants to complete two questionnaires. The first was the above-mentioned Self-Perceived Communication Competence (SPCC) questionnaire (McCroskey & McCroskey 1988), designed to find out how competent students believe they are when expressing themselves orally. The test describes 12 situations ranging from talking to a friend to giving a presentation for a group of strangers (see Appendix 1), and requires the respondents to indicate on a scale between 0 and 100 how communicatively competent they feel in each situation. Scores are based on adding up the percentages for each communication context — public, meeting, group, individual, and for each receiver — strangers, acquaintances and friends, with the average of the sub-scores per receiver comprising the total SPCC score. A score under 59 is taken to signal a low self-evaluation of communication competence. Scores of 87 and above indicate high self-confidence.

The second questionnaire was based on the Attitude/Motivation Test Battery (AMTB), designed by Gardner (1985a), which — to our knowledge — is the only standardized test of attitudes, motivation and underlying concepts. Based on the statements related to integrative motivation, language anxiety and self-perception of global linguistic competence, we drew up a list of 23 statements, and asked the participants to evaluate these on a 5-point Likert scale (1 = absolutely not true, 5 = absolutely true; see Appendix 2). The scores for the positively weighted items were added up and the scores for the negatively weighed items were subtracted from this figure.

To gauge the participants' interpreting skills, we administered and audio-recorded their performance on an L1-to-L2 sight translation task. Sight translation is commonly viewed as an integral part of interpreter training, and considered to be as cognitively demanding as simultaneous and consecutive interpreting (Agrifoglio 2004). From the point of view of experimental research design, sight translation has the advantage of a reasonably well controlled length of elicited output. It is a hybrid activity between translating and interpreting and since it is part of the curricula of both translation and interpreting students, seemed the most "neutral" of test tasks. Since we were interested in the potential effect of language anxiety in L2, we chose the L1-to-L2 direction. The students' performances were rated by a 'blind judge' with over twenty years of experience as a professional interpreter and interpreter trainer, who was asked to rate the recorded sight translations on a scale from 0 to 20, based on two parameters: (1) interpreting performance in general, (2) fluency in English.

While our use of only one blind judge may be considered a limitation, since it does not allow for a test of rater reliability, we would like to point out that applying

the scores of multiple judges does not always contribute to inter-rater reliability either, as shown by studies using similar experimental designs. The impossibility of aggregating multiple correlation coefficients impels the researcher to use averages — which decreases the discriminative power of the scores (Stengers 2009).

4.3 Results

4.3.1 *ID variables*

Our first observation is that the interpreters' SPCC scores were markedly higher than those of the translators (see Table 1). The Mann Whitney *U* test shows that the difference in scores is statistically significant ($U = 71$, $p = .02$). Three of the 23 students in the translation group obtained a score below 59, which signals low self-evaluation of communicative competence, whereas all twelve students in the interpreter group obtained scores above 59 percent and four even obtained a score above 87 percent, indicating an exceptionally high self-evaluation of communicative competence.

Table 1. Mean scores and standard deviations (in parentheses) for the ID variables

		Translation students ($n = 23$)	**Interpreting students ($n = 12$)**
SPCC		71.62 (10.22)	80.94 (8.82)
AMTB	Language anxiety	10.22 (5.72)	2.33 (5.02)
	Motivation	8.70 (2.10)	10.00 (2.70)
	Self-perception linguistic competence	20.30 (4.23)	21.83 (4.80)

With regard to language anxiety, the interpreting and translation students also seemed to exhibit markedly different profiles. While the translation students showed high levels of language anxiety, the interpreter group reported hardly any language anxiety at all. The difference in average group scores is highly significant ($U = 41$, $p = .001$).

For the remaining ID variables — integrative motivation and self-perception of global linguistic competence — no significant difference was found ($U = 86.5$, $p = .07$ and $U = 108.5$, $p = .30$). It seems that interpreting students are very confident when it comes to performing oral communicative tasks, while translation students are not. However, when other linguistic skills (such as reading and writing) are included in the profiling, this difference levels off.

When we related the scores of the entire test population (translation students and interpreting students) for one ID variable to the other ID variables (see Table 2), we observed an inverse correlation between SPCC scores and language anxiety,

confirming our expectations: students who report high communication competence tend to experience less language anxiety. The SPCC scores show a moderate correlation with students' own perception of global linguistic competence, but no significant correlation with the scores for motivation. Neither are the latter significantly associated with self-perception of global linguistic competence. Yet, a significant negative correlation is observed between students' language anxiety and their self-evaluation of global linguistic competence. Similarly to the SPCC scores, self-perception of global linguistic competence correlates negatively with language anxiety, but is not significantly associated with the students' levels of motivation.

Table 2. Spearman rank correlations between the ID variables

	AMTB/ Language anxiety	**AMTB / Motivation**	**AMTB/ Self-perc. ling. comp.**
SPCC	−.616 ($p < .001$)	.106 (*ns*)	.378 ($p < .05$)
AMTB/ Language anxiety	–	−.130 (*ns*)	−.341 ($p < .05$)
AMTB/ Motivation	–	–	.320 (*ns*)

4.3.2 *The interpreting performance*

Only 23 of the original 35 students (17 translation students and 6 interpreting students) participated in the sight translation task. The figures in Table 3 show the assessments of the interpreting and translation students by the blind judge to be quite similar, in terms of both fluency and overall quality of the interpreting performance ($p > .05$).

Table 3. Mean scores and standard deviations (in parentheses) for the interpreting performance)

	Translation students ($n = 17$)	**Interpreting students ($n = 6$)**
Fluency	11.65 (3.14)	12.33 (2.88)
Quality of interpreting performance	14.24 (2.99)	15.33 (2.34)

The scores obtained for "fluency" and "quality of interpreting performance" yield a spearman rank correlation coefficient of .654 ($p < .01$). In other words, the participants who scored high on the quality of their interpreting performance were generally perceived as fluent and vice versa. This is hardly surprising, as fluency is one of the key requirements in interpreting.

4.3.3 *Relating interpreting performance to ID variables*

We found no significant correlation between the scores obtained for the ID variables and those awarded for either of the interpreting performance parameters

(see Table 4). More specifically, no correlation was attested between language anxiety and the way the participants performed during the sight translation task. Nor was a correlation established between the self-perception of our participants (both of oral communicative competence and of global linguistic competence) and their interpreting performance.

So, while we did find significant differences in the ID profiles of our interpreting and translation students, these differences did not help predict the students' interpreting performance in the sight translation assignment. Self-confident students did not outperform students with marked language anxiety profiles. This finding may be relevant for contexts of admission testing where non-linguistic skills or personality traits are taken into consideration.

Table 4. Spearman rank correlations between the ID variables and the interpreting performance ($n = 23$)

	Fluency	**Quality of interpreting performance**
SPCC	.172 ($p > .05$)	.046 ($p > .05$)
AMTB/Language anxiety	–.064 ($p > .05$)	–.154 ($p > .05$)
AMTB/Motivation	–.037 ($p > .05$)	.021 ($p > .05$)
AMTB/Self-perc. ling. comp.	.275 ($p > .05$)	.206 ($p > .05$)

5. Conclusions and perspectives

Overall, the ID information we obtained on the basis of the two questionnaires seems to suggest that interpreters and translators have distinct profiles. The students who opted to train as interpreters considered themselves highly communicative and fluent, while this is substantially less the case for their peers who preferred to pursue their training as translators. The interpreting students reported higher levels of self-confidence especially with regard to oral tasks. When self-reported linguistic competence in general is taken into consideration, the attested difference in student profiles is offset, as the translation students reported similar levels of self-confidence with regard to reading and writing, for example. The findings with respect to the "language anxiety" ID variable tell the same tale. The translation students reported distinctly more language anxiety than the interpreting students. This is paralleled by the SPCC scores, which is not surprising, since students who believe that they are good communicators in the foreign language are likely to experience less anxiety when performing oral tasks.

The ID variable "motivation" was not found to be associated with the other two ID variables under examination, however. Even though the translation students are less communicatively confident and experience more language anxiety, they are just as motivated as their interpreter peers to learn the foreign language and to follow through with their training. This was perhaps the most striking finding of our study.

Finally, the differences observed between the ID profiles of the interpreting and translation students were not reflected in their sight translation performances. The interpreting students did not outperform the translators (at least on the parameters we selected, i.e. fluency and overall quality of the performance) and no significant correlations were found between students' levels of language anxiety, linguistic self-confidence and motivation, on the one hand, and the scores obtained for fluency and overall quality of the performance, on the other. This does not mean, of course, that the interpreting students will not improve through training. It should be borne in mind that the tests were administered while the students were still in the early stages of training.

It would be premature to generalize from these findings, however. After all, the number of participants was small (which is not unusual in interpreting studies) and there was only one assessor. Furthermore, the findings are drawn from only one sight-translation task, and it would be useful to determine whether other types of interpreting, such as simultaneous or consecutive tasks, would yield similar findings.

Nevertheless, the results do suggest that the distinct ID profiles that have been proposed for interpreting students and translation students on the basis of previous research through interviews and introspection are borne out by the use of standardized ID questionnaires. The results clearly show that translation and interpreting students do perceive themselves in quite different ways. According to our findings, however, there is no close correspondence between those profiles and the quality of the students' interpreting performance, at least at the start of their training. This calls for some caution regarding admission testing practices that take candidates' ID profiles into account, as these do not seem to suffice as predictors of success, at least in the short term. A larger-scale follow-up study would be welcome to establish whether ID profiling in addition to linguistic competences may help to predict how well novice interpreter trainees will perform in the longer term, and to establish the stability of individual ID variables.

Further research should take more ID variables into consideration. Cognitive style variables and personality variables such as "extraversion" and "ego permeability" might provide useful additional insights into the profiles of translation and interpreting students and lead to more reliable and transparent aptitude testing.

References

Agrifoglio, M. (2004). Sight translation and interpreting: A comparative analysis of constraints and failures. *Interpreting* 6 (1), 43–67.

AIIC Training Committee (2006). *Conference interpreting training programmes: Best practice*. http://www.aiic.net/ViewPage.cfm/page60 (accessed 20 April 2009).

Clément, R. (1980). Ethnicity, contact and communicative competence in a second language. In H. Giles, W. P. Robinson & P. Smith (Eds.), *Language: Social psychological perspectives.* Oxford: Pergamon Press, 147–155.

Clément, R. & Gardner, R. (2001). Second language mastery. In W. P. Robinson & H. Giles (Eds.), *New handbook of language and social psychology.* London: Wiley, 489–504.

Crookall, D. & Oxford, R. (1991). Dealing with anxiety: Some practical activities for language learners and teacher trainees. In E. K. Horwitz & D. J. Young (Eds.), *Language anxiety: From theory and research to classroom implications.* Englewood Cliffs, NJ: Prentice-Hall, 141–150.

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition.* Mahwah, NJ: Lawrence Erlbaum Associates.

Eyckmans, J. (2007). Taking SLA research to interpreting: Does knowledge of phrases foster fluency? In F. Boers, J. Darquennes & R. Temmerman (Eds.), *Multilingualism and applied comparative linguistics. Vol. 1: Comparative considerations in second and foreign language instruction*. Newcastle upon Tyne: Cambridge Scholars Publishing, 89–105.

Fraser, J. (2000). What do real translators do? Developing the use of TAPs from professional translators. In R. Jääskeläinen & S. Tirkkonen-Condit (Eds.), *Tapping and mapping the processes of translation and interpreting.* Amsterdam: John Benjamins, 111–123.

Gardner, R. C. (1985a). *The attitude/motivation test battery: Technical report*. Ontario: Department of Psychology, University of Western Ontario.

Gardner, R. C. (1985b). *Social psychology and second language learning: The role of attitudes and motivation.* London: Arnold.

Gardner, R. C. & Clément, R. (1990). Social psychological perspectives on second language acquisition. In H. Giles & W. P. Robinson (Eds.), *Handbook of language and social psychology.* New York: John Wiley & Sons, 495–517.

Gile, D. (2001). L'évaluation de la qualité de l'interprétation en cours de formation', *Meta* 46 (2), 379–394.

Henderson, J. (1980). Siblings observed. *Babel* 26 (4), 217–225.

Henderson, J. (1987). *Personality and the linguist*, Bradford, UK: Bradford University Press.

Hubscher-Davidson, S. (2009). Personal diversity and diverse personalities in translation: A study of individual differences. *Perspectives* 17 (3), 175–192.

Jiménez Ivars, A. & Pinazo Calatayud, D. (2001). "I failed because I got very nervous". Anxiety and performance in interpreter trainees: An empirical study. *The Interpreters' Newsletter* 11, 105–118.

Johnson, K. (2001). *An introduction to foreign language learning and teaching*. London: Longman.

Keiser, W. (1978). Selection and training of conference interpreters. In D. Gerver & H. W. Sinaiko (Eds.), *Language interpretation and communication.* New York: Plenum Press, 11–24.

MacIntyre, P. D., Clément, R., Dörnyei, Z. & Noels, K. A. (1998). Conceptualising willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *Modern Language Journal* 82, 545–562.

McCroskey, J. C. & McCroskey, L. L. (1988). Self-report as an approach to measuring communication competence. *Communication Research Reports* 5 (2), 108–113.
Nagy, B. (2005). *Willingness to communicate in an L2: A study of advanced learners of English as a foreign language.* MA thesis, Rijksuniversiteit Groningen.
Norrman, E. (2009) Post-graduate training in conference interpreting. University of Turku. http://www.konftulk.utu.fi/en/ (accessed 15 June 2009).
Paneth, E. (1962). The interpreter's task and training. *The Incorporated Linguist* 1 (4), 102–109.
Schumann, J. H. (1986). Research on the acculturation model for second language acquisition. *Journal of Multilingual and Multicultural Development* 7 (5), 379–392.
Schweda Nicholson, N. (2005). Personality characteristics of interpreter trainees: The Myers-Briggs Type Indicator (MBTI). *The Interpreters' Newsletter* 13, 109–142.
Seleskovitch, D. (1978). *Interpreting for international conferences*, Washington, DC: Pen and Booth.
Stengers, H. (2009). *The idiom principle put to the test: An exercise in applied comparative linguistics.* PhD dissertation, Vrije Universiteit Brussel.
Szuki, A. (1988). Aptitudes of translators and interpreters. *Meta* 33 (1), 108–114.
Yan, J. X. (2007). An investigation of language anxiety in interpreter training classes. Unpublished paper presentation. http://iris.cityu.edu.hk/generalpublic/ProjectInfo_GP.cfm?Pno=7002067 (accessed 20 April 2009).
Zannirato, A. (2007). What can language teachers learn from interpreter trainers? (and the other way round) In F. Boers, J. Darquennes & R. Temmerman (Eds.), *Multilingualism and applied comparative linguistics. Vol. 1: Comparative considerations in second and foreign language instruction.* Newcastle upon Tyne: Cambridge Scholars Press, 69–88.
Zannirato, A. (2008). Teaching interpreting and interpreting training: A conference interpreter's overview of second language acquisition. In J. Kearns (Ed.), *Translator and interpreter training: Issues, methods and debates.* London: Continuum, 19–38.

Appendix 1: SPCC questionnaire

Imagine that you are in an English speaking country studying for one semester at a college or university and you find yourself in the 12 situations below. People's abilities to communicate vary a lot and sometimes the same person is more competent to communicate in one situation than in another.
Please indicate how competent you believe you are in each of the situations described below.
Estimate your competence and put a percentage on the dotted line next to the sentence.

0 % = completely incompetent and 100 % = competent.

Situation

1. Present a talk in English to a group of English speaking strangers. ……%
2. Talk in English with an English speaking acquaintance. ……%
3. Talk in English in a large meeting of English speaking friends. ……%
4. Talk in English in a small group of English speaking strangers. ……%
5. Talk in English with an English speaking friend. ……%
6. Talk in English in a large meeting of English speaking acquaintances. ……%

7. Talk in English with an English speaking stranger. ……%
8. Present a talk in English to a group of English speaking friends. ……%
9. Talk in English in a small group of English speaking acquaintances. ……%
10. Talk in English in a large meeting of English speaking strangers. ……%
11. Talk in English in a small group of English speaking friends. ……%
12. Present a talk in English to a group of English speaking acquaintances. ……%

Appendix 2: Questionnaire based on AMTB

Please read the statements below. Think about how true they are for you.

1= absolutely not true, 2= somewhat false, 3= in between, 4= somewhat true, 5= absolutely true

Statement

1. I speak English almost as well as a native speaker. ……
2. I need to work a lot on my English. ……
3. I am good at writing essays in English. ……
4. I am good at doing grammar tasks. ……
5. I am good at reading in English. ……
6. I am good at understanding spoken English. ……
7. I can talk about any topic in English easily. ……
8. My English has improved a lot since I became an English major. ……
9. I have forgotten some of my English since I became a student at this college. ……
10. I work consciously to improve my English language proficiency. ……
11. I try to meet as many speakers of English as possible to practice English. ……
12. I do not need to practice my English as it is good enough for my studies. ……
13. I know I should improve my English but I am too lazy to do so. ……
14. I always feel that the other students speak English better than I do. ……
15. It embarrasses me to volunteer answers in class. ……
16. I never feel quite sure of myself when I am speaking English in class. ……
17. I usually get uneasy when I have to speak in English. ……
18. I am afraid that the other students will laugh at me when I speak English. ……
19. I feel calm and confident in the company of English speaking people. ……
20. I do not find it embarrassing at all if I have to give directions in English to tourists. ……
21. I am happy to be an English major. ……
22. If I could choose now I would not become an English major. ……
23. When I have to speak English on the phone I easily become confused. ……

1= absolutely not true, 2= somewhat false, 3= in between, 4= somewhat true, 5= absolutely true

Cognitive and motivational contributors to aptitude

A study of spoken and signed language interpreting students*

Sherry Shaw
University of North Florida

This article reports the findings of a causal-comparative study with spoken language (primarily conference) and signed language (primarily public service) interpreting students carried out at four institutions in the European Union in 2008. The study was built on two previous investigations of essential characteristics, as reported by interpreting students and their professors, to measure these characteristics with standardized performance and motivation tests. It grouped participants as "entry-level" or "advanced" depending upon their prior experience in simultaneous interpreting coursework. The study documented cognitive and motivational scores of spoken language (SP) and signed language (SL) interpreting students at both levels, using a computerized neuropsychological screening test and an achievement motivation instrument. Significant differences between the SP and SL students were found in the areas of visual memory, concentration, and internality (belief that success is due to internal causes), and between the advanced and entry-level students in the areas of concentration and the eagerness to learn new concepts in the absence of external rewards.

Introduction

The challenge of describing the complex construct of aptitude for interpreting skill acquisition (and ultimately, interpreter effectiveness) potentially begins with investigating certain cognitive traits and achievement motivation dispositions that manifest during the course of university study. Undoubtedly, programs that prepare interpreters, whether spoken or signed, are interested in admitting students with the skills, traits, and aptitudes that are most likely to affect positive outcomes.

1. This project was funded by a University of North Florida faculty research grant.

Investigative efforts around the question of a student's general suitability for learning the interpreting process have been ongoing during the past 25 years in the realm of spoken language/conference interpreting (e.g. Alexieva 1993; Gerver et al. 1989; Lambert 1991; Longley 1989; Moser-Mercer 1985, 2000; Russo 1993; Russo & Pippa 2004; Timarová & Ungoed-Thomas 2007, 2009), but to a much lesser extent within signed language interpreting studies (e.g. Humphrey 1994; López Gómez et al. 2007; Monikowski 1994). Out of these studies came critical conclusions about determining student aptitude for learning the interpreting process. For example, Moser-Mercer concluded that longitudinal observation of students "not only gives the student a chance to develop according to his own learning abilities, but also allows the teacher/professional to judge the student's potential with greater accuracy" (1985:100). Additionally, López Gómez et al. determined that "perceptual-motor abilities are more important than personality traits in predicting success [for signed language interpreting students]" (2007:88). Timarová and Ungoed-Thomas (2007:43) suggested that interpreting programs focus more on "soft" skills that included motivation and open-mindedness when considering the suitability of prospective students. It is clear that more evidence is needed about actual dispositions that students bring to their programs if we are to move forward in the quest to appropriately assess aptitude and screen prospective students for program admission.

Motivation research informs us that student learning is impacted by intrinsic and extrinsic goal orientation, task value, and affect (Pintrich 1988), and it underpins the need to examine these characteristics alongside cognitive measurements, especially if the desired outcome is to identify predictors of student success in interpreting programs. The overarching issue of programs making high-stakes admission decisions based on assessment of student aptitude is still as important to interpreting programs as when Dodds (1990:17) suggested that interpreting student aptitude tests were "unreliable, subjective, and therefore unable to correctly predict the candidate's future academic and professional performance." Evidence to date of a strong correlation between any test (or battery of tests) and prospective student performance in a given program of study remains inconclusive. Additionally, Dodds challenged programs to avoid screening candidates for skills they have not yet learned to perform and raised a compelling argument to study motivation as a consideration for program admission.

The consideration of student dispositions as a viable component of aptitude aligns with Gardner's (1993) theory of multiple intelligences, brain-based learning, and individual learning styles. According to Gardner, learning is influenced by emotion, or affective responses to information, and thinking and feeling occur simultaneously. Brain-based learning is a contemporary construct in cognitive science research that branches off from the earlier research on multiple

intelligences and elaborates on neuroscience in education. Learning style, or the cognitive process that allows a student to master new skills, is said to impact the degree to which one acquires critical thinking skills as well as interpersonal, relational abilities (Lynch 2008). These discussions on neuroscience and its influences on the cognitive and educational psychology disciplines highlight links between emotion, attention, and learning and are relevant, given the influence of cognitive psychology on interpreting and translating studies (Gile 2009). Connell (2009) provides a comprehensive and general discussion of brain-based learning, multiple intelligences, and learning styles.

Within the context of these foundational studies, the observation of interpreting students in both language modalities (spoken and signed) within a single empirical study is a relatively new venture that began with Shaw et al. (2004), who attempted to elucidate similarities and differences between the two types of interpreting students. Questions remain about what skills and dispositions we should expect of students at the onset of their training programs and which of these might be acquired and refined as they progress. Observing the modalities in tandem reflects current literature on such topics as working memory and language availability (Gile 2009). Additionally, comparison between the groups often necessitates personal and institutional alliances (as in the present study) and these alliances incidentally create the momentum for institutions to recognize the value of official research partnerships on this topic of shared concern (D. Gile, personal communication). For these purposes, a causal-comparative study was designed to examine cognitive and motivational contributors from the standpoint of students already admitted to interpreting programs to measure the most highly valued characteristics identified in previous studies on signed and spoken language students by Shaw et al. (2004) and Shaw and Hughes (2006). These two studies represented a mixed-methods effort to describe interpreting student and faculty perceptions of key factors that contributed to student success. The qualitative data obtained in the first study informed the construction of a survey instrument for the second study to identify which characteristics were critical to learning to interpret. The latter study resulted in descriptive data that isolated motivation, flexibility, and confidence in the category of *Personality Characteristics*, as well as figure-ground discrimination, mind mapping, critical thinking, multitasking, and rapid auditory/visual processing in the *Information Processing* category, and self-regulated learning and broad knowledge base in the category of *Academic Habits and Skills*. The present study continued this line of investigation to actually measure some of the characteristics that were self-reported by students and faculty members as extremely important for learning to interpret. It provided a venue to make preliminary comparisons between student groups based on their modalities and their level of study. The research design used in this study was modeled after a study

by Lynch (2008) in which university students were compared across level (years of study). Considering that the above-mentioned characteristics encompass hard skills (cognitive, technical components of a task) and soft skills (non-cognitive components, personal habits, optimism, and motivation), and that these may comprise the elusive construct of *interpreting student aptitude*, the study pursued answers to the following questions:

1. What are the prevalent cognitive traits and motivational dispositions of entry-level (beginning simultaneous interpreting courses), advanced (nearing program completion), spoken language (SP) and signed language (SL) interpreting students?
2. What are the similarities and differences between the student groups?

The null hypothesis was that there would be no differences in cognitive or motivational characteristics across the four groups.

Method

Participants

Forty-seven students from SP ($n = 29$, 62%) and SL ($n = 18$, 38%) interpreting programs volunteered to participate in the study and were classified into the two categories: entry-level ($n = 18$, 38%) and advanced-level ($n = 29$, 61.7%). Coincidentally, the breakdown of participants was the same for the level (entry and advanced) and for the modality (spoken and signed) variables. The mean age was 24.23 ($SD = 3.84$ years) and the gender was primarily female ($n = 41$, 87%). The sample represented interpreting programs in the Czech Republic ($n = 11$, 23.4%), Austria ($n = 11$, 23.4%), Belgium ($n = 15$, 31.9%), and The Netherlands ($n = 10$, 21%). The sample was intentionally diverse, to allow for meaningfully comparing them across modalities and across levels.

As one of the psychological instruments used in the study was not available in Dutch or Czech, English proficiency was a criterion for participation in the project. This was not problematic because many of the students had English as a working language, had studied abroad, and/or were accustomed to using English in daily interactions. On the other hand, the computerized assessment gave students the option of selecting a test language, which would then allow for accurate calculations of reaction times (to the millisecond) in the students' native language without imposing additional processing time related to working from a B language into an A language.

Materials

The instruments used in the study were selected by an international team of reviewers — researchers and faculty members from spoken and signed language interpreting programs in the EU and the US. These reviewers rated four potential instruments selected by the researcher based on the capacity to measure student characteristics identified in previous studies (Shaw et al. 2004, 2006). The group reviewed the d2 Test of Attention (Brickenkamp & Zilmer 1998), the Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich & DeGroot 1990), CNS Vital Signs (CNS Vital Signs, LLC 2003) and the Achievement Motivation Inventory (Schuler et al. 2004). CNS Vital Signs (CNSVS) was the only computerized test among those being reviewed, and reviewers had the opportunity to take the test at no charge in order to provide useful feedback concerning its potential use in the study. Some reviewers, including one in the EU, had prior experience with the Achievement Motivation Inventory (AMI) and the other reviewers were provided with summaries of the AMI and d2 Test of Attention. The MSLQ, however appropriate for measuring learning strategies, metacognitive processing, resource management, and other motivational factors, was designed to measure these attributes within the context of a single course; thus, the reviewers did not recommend it for use in this study. Considering such features as each test's content validity (relevance to interpreting students), languages of availability, translation quality, cost, method and materials for scoring, and technical support, the CNSVS and AMI were selected for the purposes of the study.

The following institutions agreed to participate in the study and provided test venues:

1. Charles University, Prague, Czech Republic (SP program)
2. Karl-Franzens-University of Graz, Austria (SP and SL programs)
3. Lessius University College, Antwerp, Belgium (SP program)
4. University of Utrecht, The Netherlands (SL program)

These locations were selected based upon many factors, including the convenience of data collection in relation to each program's schedule. Other factors were the availability of program leaders to recruit student participants and of technical support staff to prepare the venue by downloading and installing the necessary software in a computer lab on campus.

The AMI and CNSVS addressed the study's need for data on hard and soft skills without introducing risk, such as might be associated with more robust neuropsychological evaluations. The CNSVS is a neurocognitive test battery, used primarily in medical settings for clinical screening in five domains of cognitive functions: *verbal and visual memory*, *psychomotor speed*, *reaction time*, *complex*

Table 1. Test scale descriptions

AMI Scale	Description (adapted from AMI manual)
Compensatory Effort	willingness to expend extra effort to avoid failing at a task, even if effort results in over-preparation (constructive reaction to possibility of failure)
Competitiveness	motivation derived from competing with others
Confidence in Success	confident even when obstacles to overcome are present
Dominance	tendency to exercise power and influence over others
Eagernesss to learn	desire and willingness to spend time enlarging one's knowledge base
Engagement	desire to be regularly engaged in activity, usually work related; high priority on work and prefers to have something to do
Fearlessness	lack of fear of failing at difficult tasks
Flexibility	willingness to accept changes; finding enjoyment in challenging new tasks
Flow	ability to concentrate on something for a long time without distraction by situational influences
Goal Setting	tendency to set goals and make long term plans to achieve them
Independence	tendency to take responsibility for one's own actions
Internality	belief that success is due to internal causes rather than to situational variables
Persistence	willingness to exert large amounts of effort over long periods to reach a goal
Preference for Difficult Tasks	tendency to seek challenging rather than easy tasks; desire to seek greater challenges once one has already been completed
Pride in Productivity	sense of accomplishment from doing one's best; self esteem dependent upon achievement
Self-Control	ability to delay gratification and organize oneself and work; concentrate on work with great deal of self-discipline
Status Orientation	desire to attain high status in one's personal life and progress professionally

attention, and *cognitive flexibility*. The CNSVS publisher provided technical and advisory support to the researcher throughout the test selection process, software installation, and data analysis due to the company's interest in extending the use of the test to include educational research. The instrument requires no previous computer experience and uses only the space bar and arrow keys. Directions for each sub-test, warm-up lessons, and stimulus vocabulary were available to each participant in his or her native language so as to reduce the language processing time used during timed psychomotor sub-tests. Additionally, the test includes

Table 1. (*continued*)

CNSVS Scale	Description
Verbal Memory	ability to remember 15 words in field of 15 distractors; correct hits refers to number of target words recognized; correct pass refers to avoiding words not included in target set; test repeated at end of battery for measurement of delayed recall
Visual Memory	ability to remember 15 geometric figures in field of 15 distractors; correct hits refers to number of target figures recognized; correct pass refers to avoiding figures not included in target set; test repeated at end of battery for measurement of delayed recall
Stroop Test	test of attentional vitality and flexibility; word is displayed in a color different from the color it actually names; measures processing speed, cognitive flexibility, and inhibition/disinhibition in simple and complex reaction times

practice sessions in the language of choice that introduce the tasks to be performed. Participants could choose among the following languages in which to take the CNSVS: Czech (11), Dutch (24), English (1), and German (11).

The AMI is a self-report, psychological instrument that measures traits relevant to interpreting skill acquisition such as *compensatory effort* (willingness to expend extra effort in order to avoid failing), *competitiveness*, *confidence in success*, *dominance*, *eagerness to learn*, *engagement* (desire to be regularly busy with school-related activities), *fearlessness*, *flexibility*, *flow* (ability to concentrate on something for a long time without distraction and situational influences), *goal-setting*, *independence*, *internality* (belief that personal successes are due to internal causes regardless of situation), *persistence*, *preference for difficult tasks*, *pride in productivity*, *self-control*, and *status orientation*. The inventory consists of 170 items on a 7-point Likert scale and requires approximately 45 minutes to complete; however, participants were not restricted in the amount of time they needed and were permitted to ask clarification questions of the administrator concerning meanings of words or colloquial phrases. (See Table 1 for more detailed descriptions of each test scale on both instruments.)

Procedure

The study was approved by the Institutional Review Board for research involving human subjects at the researcher's home institution prior to data collection. Each participating university's program leader coordinated participant recruitment for the two-hour testing session, arranged a testing location with individual computer access, provided technical support personnel to install the software needed to run the CNSVS, and assisted the test administrator with local accommodations.

Table 2. AMI sample items

Item	Scale
When I see that others are more competent than I am, it is an incentive for me to try harder.	Competitiveness
I have often noticed that everything else becomes unimportant when I totally concentrate on my work.	Flow
So that I will not be subject to criticism, I prefer to double my effort.	Compensatory Effort
I don't feel comfortable when I have nothing to do.	Engagement
I find it difficult to concentrate for a long time without becoming tired.	Persistence
Even when faced with a difficult task, I always expect to achieve my goal.	Confidence in Success
When I see or hear something new, I try to retain as much as possible.	Eagerness to Learn
When I work with other people, I usually take the initiative.	Dominance
My success depends mainly on my behavior	Internality
I have not undertaken some activities for fear of not being successful.	Fearlessness
I am open to everything new.	Flexibility
I know exactly what professional position I would like to hold in five years.	Goal Setting
I like being solely responsible for what I do.	Independence
I frequently put off until tomorrow things that I should do today.	Self Control
Tasks that I am unsure of being able to perform are particularly appealing to me.	Preference for Difficult Tasks
I have a sense of satisfaction when I improve on my performance.	Pride in Productivity
I want to be admired for my achievements.	Status Orientation

All participants signed informed consent forms and were reminded that they were free to leave the testing environment at any time and withdraw from the study with no risk of penalty. Testing occurred in one or two sessions at each participating site, depending on the number of participants and lab size, over a period of four weeks in summer, 2008. Students completed a pencil-paper version of the AMI in which they identified characteristics, attitudes, emotional responses, approaches to learning, and behaviors that they felt represented themselves. For example, students rated the applicability of sentence items that described their responses to fear, anxiety, social pressure, difficult tasks, complicated concepts, opposition, risk, and challenge. (For specific sample items, see Table 2.) The AMI was followed by CNSVS testing on the computer, after which participants were given printouts of their neurocognitive profile with the publisher's explanations of each subset and domain measured.

Results

In a general sense, the CNSVS measures the ability to (a) remember words or geometric figures and recognize them in a field of distractors, (b) perform executive control tasks that require adjusting responses to randomly changing rules, (c) sustain attention, (d) coordinate psychomotor and visual-motor responses, and (e) rapidly react to complex directions. The latter is measured in milliseconds and is of particular relevance to the purpose of this study. The AMI measures personal characteristics around the three themes of self-assurance, ambition, and self-control. Independent sample *t*-tests were conducted on all subscales of the CNSVS and AMI. The next section divides the data report of group differences and similarities by type and level of student participants. All of the statistical tests used 95% confidence intervals.

Type: Signed language (SL) and spoken language (SP) group comparisons

The SL and SP groups were nearly identical on several characteristics measured in the AMI, including *compensatory effort*, *confidence in success*, *goal setting*, *preference for difficult tasks*, and *pride in productivity*. On the CNSVS, the groups were most closely aligned on *verbal memory-correct passes* (avoiding the selection of words that were not included in the test set) and *Stroop* (attention flexibility) correct responses. Table 3 provides a summary of nearly-identical results. To test the null hypothesis, independent sample *t*-tests were conducted on the group's cognitive scores (CNSVS) and motivation scores (AMI). Significant differences were noted in CNSVS items that measured a student's visual memory (the ability to correctly select geometric shapes immediately upon presentation, $t(45) = 2.6$, $p = .011$, and after a delay, $t(45) = 2.9$, $p = .005$). Likewise, there were slightly significant differences on AMI items of *flow* (concentration), $t(46) = 2.1$, $p = .036$, and *internality* (attributing success to internal causes), $t(46) = 2.081$, $p = .043$. Based on this evidence, the null hypothesis that there is no difference between the signed language and spoken language interpreting student groups was rejected. The results suggest that (a) SL students have more highly developed visual memory skills in the presence of distracting information and retain visual information longer than SP students, and (b) SP students are more vulnerable to distractions and more likely than SL students to believe that their successes are due to their own efforts.

Although not statistically significant, it should also be noted that SL students were less likely than the SP students to exert large amounts of effort over a long period of time to achieve a goal (*perseverance*) or to value attaining a high status in their personal or professional lives (*status orientation*). Additionally, general visual motor reaction times were somewhat shorter for the SL students than for the SP

Table 3. Nearly-identical characteristics of groups

	Signed		**Spoken**	
Item	**Mean**	***SD***	**Mean**	***SD***
CNSVS				
Verbal Memory (correct passes)	14.56	1.0	14.52	.91
Stroop (correct responses)	11.94	.236	11.97	.186
AMI				
Compensatory Effort	48.7	5.6	48.9	7.7
Confidence in Success	48.1	7.7	48.2	6.7
Goal Setting	44.6	8.1	44.7	7.8
Preference for Difficult Tasks	42.7	5.2	42.5	4.1
Pride in Productivity	57.5	6.0	57.9	5.2
	Advanced		**Entry**	
CNSVS				
Verbal Memory (immediate correct passes)	14.52	.94	14.56	.98
Verbal Memory (correct hits)	14.17	1.1	14.11	1.1
Stroop (complex commission errors)	11.97	.18	11.94	.23
Stroop (correct responses)	23.9	.25	23.9	.23
AMI				
Internality	37.1	7.1	37.2	5.4
Confidence in Success	48.2	7.7	48.1	5.9
Self-Control	37.7	4.6	38.0	5.2
Verbal Memory (delayed correct passes)	14.48	.82	14.50	.98
Compensatory Effort	49.06	6.5	48.55	7.8
Independence	44.80	5.0	45.50	5.0

students, and there were no noticeable differences in the ability of the two groups to shift attention rapidly.

Level: Entry-level (EL) and advanced-level (AL) group comparisons

The EL and AL groups were nearly identical on the following AMI items: *internality*, *confidence*, *self-control*, *compensatory effort*, and *independence,* and on CNSVS items: *verbal memory-correct passes*, *Stroop correct responses*, *verbal*

Table 4. Significant differences between groups

	Signed		Spoken	
Item	**Mean**	***SD***	**Mean**	***SD***
CNSVS				
Visual Memory (immediate correct hits)	12.5	1.38	8.28	6.68
Visual Memory (delayed correct hits)	12.44	1.97	8.38	5.55
AMI				
Flow	51.4	6.6	46.9	6.9
Internality	34.5	5.6	38.54	6.6
	Entry		**Advanced**	
CNSVS				
Visual Memory (delayed reaction time)	897.54	270.654	647.01	308.2
AMI				
Eagerness to Learn	40.6	6.00	45.7	5.45
Flow	45.3	7.45	50.47	6.21

memory-delayed correct passes, and *verbal memory correct responses*. Independent sample *t*-tests on the means revealed one significant difference from the CNSVS data on the *visual memory* ability to correctly select geometric shapes after a delay, $t(45) = 2.8$, $p = .007$. The AMI data revealed significant differences on *eagerness to learn*, $t(46) = 3.03$, $p = .004$, and on *flow*, $t(46) = 2.57$, $p = .014$. Based on these findings, the null hypothesis that there is no difference between entry-level and advanced-level students is rejected. The results suggest that over the course of their studies, students develop faster reaction times to non-linguistic stimuli after other information has been presented, spend increasingly more time expanding their knowledge bases in the absence of external rewards, and develop the ability to concentrate for long periods of time without being distracted by situations around them. A summary of significant differences is presented in Table 4.

Group comparisons outside of statistical testing indicated that advanced students are likelier than entry-level students to be engaged in activities related to their "work" as students for long periods of time (*engagement*) and experience more fear of failing and of being judged by others (*fearlessness*). Entry-level students also demonstrated a more energetic approach to achieving their goals (*persistence*).

Discussion

The fact that CNSVS was developed to gather clinical baseline data on a person's cognitive functioning causes some subscales to present a ceiling effect that limits sensitivity to distinctions between the groups studied here. This may account for the fact that the *Neurocognition Index*, an average of the standard scores of the five domains, did not reveal significant group differences either between the levels or between types of students. It is important to keep this ceiling effect in mind when reviewing the presentation of the nearly identical CNSVS subscale means in Table 3 and to focus on the items that measured reaction times to the millisecond. CNSVS reveals a not-too-surprising difference between the visual memory skills of students who study and interpret a visual language and those who work in a spoken language. Their ability to hold visual (non-linguistic) images for long periods of time was expected to distinguish them from the other student group; to wit signed language interpreting programs may wish to use this skill as part of a screening tool, given the finding that students require less time to perform these tasks as they progress through the program. In other words, the realization that certain skills are teachable or acquirable throughout a course of study is helpful in the selection of items used for screening.

The AMI's construct of *flow*, reported to be significantly more prevalent in signed language students, is at first thought to be a positive trait, but could be construed as detrimental if it causes a student to "over-focus" attention, thus hindering the ability to shift from one task to another with sufficient ease. While the finding should be interpreted with caution, given that students do not necessarily self-report this characteristic, it certainly points us to further study in the area of dispositional factors and their effects on attention (i.e. Dale & Amell 2010). The fact that signed language students were less likely (although not significantly different from spoken language students) to exhibit *status orientation* could reflect the comparatively low-profile, one-on-one nature of public service interpreting with consumers who are deaf, hard of hearing, or deaf-blind. Status was understandably more of a consideration for the spoken language/conference interpreting students who may eventually work in the public eye. Another non-significant finding that stands out is the fear of failing, which was more common in advanced students. This may speak to the reality that final exit examinations loom more closely for these students than for entry-level students, who have yet to internalize the possibility of not completing the program. Possibly for the same reason, entry-level students reported being more energetic toward the prospect of achieving their goals than advanced students. In other words, the gravity with which final examinations impact a student's future appears to correspond directly to the fear factor measured in the AMI. This conclusion parallels the assertions of Lynch (2008)

concerning test anxiety in upper-level students as course difficulty increases and demands on students to exit successfully becomes greater, especially if program completion is contingent upon a final examination outcome.

Limitations and future research

The limitations of this study include the inherent barriers to generalization of a self-report instrument, which, if possible, "should be substantiated with other sources of evidence" (Suter 2006). However, the AMI was developed and standardized in Europe (under the name LMI) and has undergone extensive validity and reliability testing since its development in 2001 (Schuler et al. 2004). Whereas direct measurement may be preferable to self-reporting, correlating AMI results with actual student performance on CNSVS tasks is a place to start observing interactions between motivational factors on task-based performance within an interpreting program. Although small by some standards and limited in its inferential power, the sample is of respectable size for focusing future research. The results do not allow generalization beyond the groups studied, but may serve to guide further study that should use the preferable form of trait measurement (cognitive or dispositional) by collecting data from a group over time. This recommendation may or not be feasible, given the difficulties of assembling larger samples of students and monitoring them throughout their course of studies. Another potential difficulty concerns the need for researchers, who might not be directly employed by the institutions, to gain access to test scores and other data amid privacy restrictions imposed by institutions and external controls.

The concurrent measurement procedure used in this study was the most practical means of gathering data for preliminary group comparisons on the modality variable and noteworthy first comparisons between entry-level and advanced groups. By uncovering relations between characteristics and levels/types of students, this contribution to our current body of literature should lead to further study using empirical methods that have the potential to triangulate the data and produce more rigorous evidence. Of particular interest will be research that concentrates on complex attention, visual memory, and response-time improvement of students as they move through their interpreting programs. The distinctive approach of this investigation rests on its attempt to compare two levels of signed and spoken language interpreting students on test items that directly relate to the interpreting process and to acquiring the skills needed to become effective interpreters. It is hoped that other researchers will recognize the value of conducting studies that expand upon the methodology, sample, and instruments used here in order to continue illuminating the aptitude construct for interpreting skill acquisition.

References

Alexieva, B. (1993). Aptitude tests and intertextuality in simultaneous interpreting. *The Interpreters' Newsletter* 5, 8–12.

Brickenkamp, R. & Zilmer, E. (1998). *d2 Test of Attention*. Cambridge, MA: Hogrefe & Huber.

CNS Vital Signs, LLC (2003). *CNS Vital Signs*. Chapel Hill, NC: Author.

Connell, J. D. (2009). The global aspects of brain-based learning. *Educational Horizons* 88 (1), 28–39.

Dale, G. & Amell, K. (2010). Individual differences in dispositional factors of attention predict attentional blink magnitude. *Attention, Perception, & Psychophysics* 72 (3), 602–606.

Dodds, J. (1990). On the aptitude of aptitude testing. *The Interpreters' Newsletter* 3, 17–22.

Gardner, H. (1993). *Frames of mind: The theory of multiple intelligences*. 2nd edn. New York, NY: Basic Books.

Gerver, D., Longley, P., Long, J. & Lambert, S. (1989). Selection tests for trainee conference interpreters. *Meta* 34 (4), 724–735.

Gile, D. (2009). *Basic models and concepts for translator and interpreter training*. 2nd edn. Amsterdam/Philadelphia: John Benjamins.

Humphrey, J. (1994). Effective screening procedures for entering students in interpreter education. In E. A. Winston (Ed.), *Mapping our course: A collaborative venture. Proceedings of the Tenth National Convention of the Conference of Interpreter Trainers*. Charlotte, NC: Conference of Interpreter Trainers, 144–173.

Lambert, S. (1991). Aptitude testing for simultaneous interpretation at the University of Ottawa. *Meta* 36 (4), 586–594.

Longley, P. (1989). The use of aptitude testing in the selection of students for conference interpretation training. In L. Gran & J. Dodds (Eds.), *The theoretical and practical aspects of teaching conference interpretation*. Udine: Campanotto, 105–108.

López Gómez, M. J., Bajo Molina, T., Padilla Benítez, P. & Santiago de Torres, J. (2007). Predicting proficiency in signed language interpreting. *Interpreting* 9 (1), 71–93.

Lynch, D. J. (2008). Confronting challenges: Motivational beliefs and learning strategies in difficult college courses. *College Student Journal* 42 (2), 416–421.

Monikowski, C. (1994). Required backgrounds, proficiencies, and aptitudes for students entering an interpreter education program. In E. A. Winston (Ed.), *Mapping our course: A collaborative venture. Proceedings of the Tenth National Convention of the Conference of Interpreter Trainers*. Charlotte, NC: Conference of Interpreter Trainers, 31–34.

Moser-Mercer, B. (1985). Screening potential interpreters. *Meta* 30 (1), 97–100.

Moser-Mercer, B. (2000). Simultaneous interpreting: Cognitive potential and limitations. *Interpreting* 5 (2), 83–94.

Pintrich, R. R. (1988). A process-oriented view of student motivation and cognition. In J. S. Stark & L. A. Mets (Eds.), *Improving teaching and learning through research*. San Francisco, CA: Jossey Bass.

Pintrich, R. R. & DeGroot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance, *Journal of Educational Psychology* 82, 33–40.

Russo, M. (1993). Testing aptitude for simultaneous interpretation: Evaluation of the first trial and preliminary results. *The Interpreters' Newsletter* 5, 68–71.

Russo, M. & Pippa, S. (2004). Aptitude to interpreting: Preliminary results of a testing methodology based on paraphrase. *Meta* 49 (2), 409–432.

Schuler, H., Thornton, III, G. C., Frintrup, A. & Mueller-Hanson, R. (2004). *Achievement Motivation Inventory technical and user's manual.* Cambridge, MA: Hogrefe & Huber.

Shaw, S., Grbić, N. & Franklin, K. (2004). Applying language skills to interpretation: Student perspectives from signed and spoken language programs. *Interpreting* 6 (1), 69–100.

Shaw, S. & Hughes, G. (2006). Essential characteristics of sign language interpretation students: Perspectives of students and faculty. *Interpreting* 8 (2), 195–221.

Suter, N. (2006). *Introduction to educational research: A critical thinking approach.* Thousand Oaks, CA: Sage.

Timarová, Š. & Ungoed-Thomas, H. (2007). Admission testing for interpreting courses. *The Interpreter and Translator Trainer* 2 (1), 29–46.

Timarová, Š. & Ungoed-Thomas, H. (2009). The predictive validity of admission tests for interpreting courses in Europe: A case study. In C. V. Angelelli & H. E. Jacobson (Eds.), *Testing and assessment in translation and interpreting studies*. Amsterdam/Philadelphia: John Benjamins, 225–245.

Evaluating emotional stability as a predictor of interpreter competence and aptitude for interpreting*

Karen Bontempo and Jemina Napier
Macquarie University / Heriot-Watt University

This paper proposes that variance in interpreter performance is dependent on factors of both general cognitive ability and personality. Whilst there is no doubt of the interplay between individual personality traits and job performance across many occupations, the greatest interest lies in determining which traits play the most important role; and to what extent these variables impact on learning and achievement. The paper reports on a study of 110 accredited signed language interpreters in Australia. Psychological constructs of self-efficacy, goal orientation and negative affectivity were measured, as were interpreter ratings of self-perceived competence as practitioners. The most significant finding revealed the dimension of emotional stability (represented on the negative end of the continuum by traits of anxiety and neuroticism, and measured in this study by the negative affectivity scale) as a predictor of interpreter's self-perceived competence. Based on these findings, recommendations for admission testing and interpreter education curricula are discussed.

Introduction

Despite a recent "mini-explosion" of research on signed language interpreting, which crosses a range of disciplines (Napier, 2011), very little work since earlier psycholinguistic studies of the simultaneous sign language interpreting process (e.g. Isham & Lane 1993) has drawn on psychological constructs to evaluate aspects of the interpreting process, or of signed language interpreters themselves. By drawing on an established body of psychological research that informs our

* The Auslan interpreters who willingly gave of their time to participate in this study are gratefully acknowledged for their contribution to our increased understanding of dispositional factors that may be predictive of interpreter performance.

understanding of occupations and the relationship between personality and work, and on established testing protocols, we hope to enhance our understanding of personality and the work of signed language interpreters.

This paper reports on a study conducted on 110 accredited signed language interpreters in Australia.[1] Psychological constructs of self-efficacy, goal orientation and negative affectivity were measured, as were interpreter ratings of self-perceived competence as a practitioner. Before presenting an overview of the methodology and results of the study, we review the relevant literature on psychological and signed language interpreting research.

Literature review

In recent years it has been more widely accepted that both personality factors and general cognitive ability influence performance — in the classroom and in the workplace. Such individual differences are further influenced by environmental and social factors, such as learning conditions. It needs to be stated that the primacy of general cognitive ability as a predictor of occupational performance remains largely undisputed, despite gains in the field of personality studies (Ree & Earles 1992; Ree et al. 1994; Schmidt & Hunter 1998). Essentially this means there is typically a complementary fit between a person's intellectual resources and the cognitive demands of their chosen occupation.

However, accounting for at least some of the variance in occupational performance amongst individuals, the research published in recent years demonstrates a convincing relationship between personality and performance as well (Barrick & Mount 2005; Barrick et al. 2001). Although personality may be only a small part of the bigger picture, Ones et al. (2007) suggest that personality constructs may account for specific attitudes, behaviors and performance in an occupational context. Notably, the role of personality in successful completion of courses of study (Phillips et al. 2003); skill acquisition (Oakes et al. 2001); job performance (Judge et al. 1999); and career success (Bozionelos 2004) should not be underestimated. An individual's preferences and desires evidently influence his/her person-vocation fit (Reeve & Heggestad 2004), and "adaptability, positive relationships, openness to experiences, and social and psychological capital" (Fouad 2007:556) impact on career exploration. Ultimately, individual personality differences do relate to outcomes at work.

1. This study was designed and conducted by Karen Bontempo, PhD candidate in the Department of Linguistics, Macquarie University, under the supervision of Dr Jemina Napier.

Furthermore, it is now understood that the actual relationship between general cognitive ability and personality is negligible. There may be a very bright individual who is lazy, anxious, insecure and unmotivated, which potentially impacts on job performance and productivity. Similarly, there may be a very responsible, goal oriented, secure and careful individual who may lack the general cognitive ability required for a particular job, even if apparently suited in temperament. It appears personality is a poor predictor of general cognitive ability, and that tests of cognitive ability should be interpreted separately from personality tests administered to individuals, with limited inferences made between the constructs. This reinforces the view that attention needs to be paid to both general cognitive ability and personality factors in any recruitment and selection process in an occupational context (Judge et al. 1999).

Barrick and Mount (2005) describe the cognitive factors affecting job performance as the *"can do"* factors — an individual needs a requisite level of general cognitive ability, to perform the tasks inherent in the job. Personality however may be described as the *"will do"* factor — will the person be dependable, motivated, confident and goal-oriented enough to apply their individual capabilities towards effectively performing the tasks of the job? In essence, and quite logically, performance is predicated on both ability/capacity and motivation/willingness factors.

To concentrate research efforts on the cognitive factors that may predict performance in a profession is neglectful of the range of affective factors that have implications for the psychology of work. Without belittling the obvious importance of general cognitive ability in interpreters (including linguistic proficiency, intelligence, working memory, processing speed and so on), the present study examines more specifically those personality factors that may be predictive of performance and central to understanding individual differences in this field of work.

Personality and job performance

Empirical evidence indicates the most significant broad personality construct that consistently predicts occupational performance is conscientiousness (Mount & Barrick 1998; Salgado 1997; Bozionelos 2004). This appears logical, with conscientiousness as a construct that includes traits where a person strives for achievement; is self-disciplined, dutiful, dependable, persistent, goal-oriented, organized, responsible, efficient, thorough, deliberate, hard-working, and a perfectionist (Judge et al. 1999). It may be surmised that these are characteristics that employers and educators would probably like to see in their employees and students.

The other construct that has revealed the most interesting findings is that of emotional stability, which is positively related to on-the-job success and work performance (Salgado 1997; Mount & Barrick 1998; Judge et al. 1999; Barrick et al.

2001; O'Brien et al. 2008). Emotional stability can be seen as a continuum, with one end of the scale being emotional stability and the other being emotional instability, often also described as neuroticism, which is closely linked to trait anxiety. Sub-traits of neuroticism include insecurity, fearfulness, worry, tendency towards depression/negative moods, feelings of guilt, high emotional reactivity and irritability. Bozionelos (2004) adds that individuals who score high on this dimension often have low self-confidence and are more vulnerable to stress. At the other end of the continuum, individuals who score low in neuroticism tend to be more emotionally stable and calmer, experience fewer persistent negative emotions, and react less severely to stress. Trait anxiety and neuroticism has been collectively described and measured on some personality tests as "negative affectivity" — an enduring tendency to experience negative mood and emotion.

Given the confirmed significance of emotional stability in predicting work performance in the wider literature, it is a construct of interest in researching aptitude for interpreting. The research on foreign language learning achievement points to anxiety having a negative impact on language learning and success in second language acquisition (Gregersen & Horwitz 2002; Dörnyei 2005). Theories suggest second language learning poses a significant personal threat to self-concept and self-confidence in students (Onwuegbuzie et al. 2000), and traits of perfectionism and self-consciousness can be highly detrimental to language learners (Gregersen & Horwitz 2002). Additionally, a students' self-efficacy (a belief in one's capacity to undertake a task) and self-confidence may be diminished in a second language learning classroom due to the nature of the tasks undertaken and the extent of teacher correction that is often required (Dörnyei 2005). Such findings could easily be applied to interpreting students, and indeed to the practice of interpreting.

Research on personality and interpreters

To date however, there appears to be little empirical research on the emotional stability of interpreters. A handful of pioneering studies with very low numbers of signed language interpreter participants were conducted to investigate "interpreter personality" (Schein 1974 and Frishberg & Enders 1974, both cited in Frishberg 1990; Rudser & Strong 1986; Doerfert & Wilcox 1986), but these studies had few points of agreement.

More recent international studies, at times including larger samples of participants and occasionally including both spoken and signed language interpreters, have been somewhat more successful in trying to pin-point factors predictive of interpreter performance. In defining a wider range of cognitive and personality measurements, the following researchers have started to identify some common themes of interest with regard to predicting proficiency and outlining

characteristics that may contribute to success in interpreting courses and in the profession of signed language interpreting.

López Gómez et al. (2007) administered a battery of tests to 28 signed language interpreting students in Spain, examining perceptual-motor coordination, cognitive skills, personality factors and academic background. Perceptual-motor coordination was found to be the most significant predictor of proficiency in learning a signed language. Cognitive and personality factors were also found to be influential, although personality factors to a lesser degree than cognitive skills. The personality factor of dominance was found to be of particular interest — high scores on this factor indicated a person was assertive, resourceful, confident, task-oriented, responsible, and stress-resistant. Low scores point to low self-confidence, rigidity in problem solving and unreliability. López Gómez et al. found this trait of dominance to be relevant as a predictor of signed language interpreting abilities, which supports Rudser and Strong's (1986) earlier work. In addition, cognitive abilities such as abstract reasoning and memory skills were identified as important, supporting the findings of Seal (2004). Academic background was not found to have a significant bearing on results.

After surveying 1357 signed language interpreter education program participants and faculty in North America and Europe, Shaw and Hughes (2006) identified characteristics thought to be most important for the successful completion of interpreting coursework and those characteristics perceived as most in need of development in students. Their research highlighted three areas of interest — academic habits and skills, information processing skills and personality characteristics. In many cases, students and faculty held different opinions. On the personality characteristics scale, however, both groups firmly agreed that self-confidence was vitally important — but also in great need of further development in students.

Shaw and Hughes (2006) note that academic habits and personal qualities and traits may seem to influence success in courses and in the profession, but "until causal factors for student success are established, caution should be used in developing admission screening devices that do not account for the students ability to learn, develop and enhance critical personal and cognitive characteristics" (p. 218).

Shaw, Grbic and Franklin (2004) held focus groups with spoken and signed language interpreting students to compare their perceptions of factors that contribute to, or inhibit, readiness to apply language skills to interpretation studies. Essentially they found that students experience a period of transition and quickly realize that the task of interpreting is more complex than being bilingual. Participants specifically identified confidence and risk-taking as primary personality assets that contributed to successful adaptation and performance while studying interpreting, in line with the earlier work of Rudser and Strong (1986).

In analysing 1379 signed and spoken language interpreting students' responses to a personality questionnaire, Stauffer and Shaw (2006) did not identify any significant predictive characteristics. Stauffer and Shaw did not administer a recognised psychometric instrument to study participants; instead they analysed responses regarding 17 personality characteristics considered relevant for interpreters. The characteristics were derived from the focus group discussions with practitioners in the earlier study by Shaw et al. (2004) mentioned above. Although their scale did not produce significant findings regarding predictive dispositional traits, Stauffer and Shaw recommended that researchers undertake further work to clearly identify the specific personality elements that may predict success in interpreter education programs and later in the field, and that such elements then be introduced into student selection processes for entry into interpreter education programs.

Findings regarding the psychological make up of the signed language interpreter point to the potential role of personality in successful occupational performance. However the research conducted to date has delivered inconsistent results, and in the vast majority of cases, has been conducted on interpreting students rather than practitioners. The studies that have been conducted with interpreting practitioners have often included very small samples of practitioners, resulting in sampling errors and an inability to generalize findings.

No study to date has obtained a psychological profile from a viable number of qualified interpreters with a measurement of emotional stability, to determine whether emotional stability is a salient predictor of performance in the study of interpreting, and/or for the occupation of interpreting. This is despite evidence that suggests the capacity to control stress should be a prerequisite for interpreting (Moser-Mercer 1985; Brisau et al. 1994; Kurz 2003); and the intuitive understanding that higher levels of neuroticism, anxiety and reactivity to stress in an individual are likely to impede the acquisition and performance of interpreting skills.

Self-efficacy, goal orientation and negative affectivity

A number of factors that could be considered promising non-cognitive predictors of occupational performance have never been measured in signed language interpreters. These aspects of personality include factors of self-efficacy, goal orientation and negative affectivity, all constructs that can be measured with existing valid and reliable psychometric tools, and considered predictive of performance in a variety of contexts.

Self-efficacy is essentially a level of expectancy of succeeding at a task, resulting from belief in one's overall performance competence (Chen et al. 2001). It is widely considered a powerful predictor of effective learning, work attitudes and job performance. Furthermore, self-efficacy is considered to be associated with

the psychological constructs of self-concept, self-esteem and assertiveness (Choi et al. 2001). Self-efficacy is also positively related to conscientiousness (Chen et al. 2001), perseverance, feedback seeking, learning and achievement in training and employment contexts (Bell & Kozlowski 2002).

Achievement goal theory is related to motivation (Dweck & Leggett 1988). Specifically, goal orientation is a dispositional trait that leads some individuals to seek challenging tasks and to thrive under difficult conditions. When faced with failure, such individuals consider the experience grounds for useful feedback, and take action towards sustaining or improving future performance (Button et al. 1996).

Like self-efficacy, goal orientation is considered a relatively stable personality trait. In this paper we are specifically concerned with learning goal orientation (rather than performance goal orientation, a related but different construct), given its potential link to aptitude for interpreting. Learning goal orientation is sometimes described in the literature as "mastery orientation" or "action orientation," and is a desire to develop oneself, to acquire new skills, to improve one's competence or to master a new situation or context (Bell & Kozlowski 2002).

It is suggested by Dweck and Leggett (1988) that the disposition trait of goal orientation at least partially determines locus of control. Locus of control is an important concept in personality psychology and refers to the extent to which an individual believes they have personal control over outcomes or events that occur in their lives. In essence, the difference between believing "fate" or chance has dealt them their lot in life, that is, having a high *external* locus of control, versus believing one is responsible for one's own actions and behaviours, that is, having a high *internal* locus of control. People with a high internal locus of control tend to assume they will be successful (within reason) at whatever they try, and that they can problem-solve to influence events and situations in their own lives. People with a high external locus of control exhibit a type of "learned helplessness," tend to be less resilient, and more prone to stress, anxiety and depression due to their low level of perceived personal control (Skinner & Greene 2008). Goal orientation is also positively linked to self-esteem (Button et al. 1996) and to self-efficacy — people with high levels of learning orientation seem to be "buffered" by the negative effects of failure (Bell & Kozlowski 2002). Indeed, people who are learning-goal oriented "are not threatened by failure; to them, failure represents an opportunity to extend one's competence through enhanced effort" (Button et al. 1996: 31).

A final dimension of interest is negative affectivity (Watson et al. 1988). As noted earlier, negative affectivity (NA) is a specific dispositional construct closely linked to neuroticism and trait anxiety. It can be considered either a trait (enduring temperament) or a state (temporary mood), but for the purposes of this study we are measuring it as a trait. The literature suggests people with high levels

of negative affectivity respond poorly in stressful situations, have less resourceful coping strategies, are more emotionally reactive, become anxious more rapidly, have a negative perception of themselves, and tend to always see the negative side of a situation or a problem. People with low levels of negative affectivity on the other hand display a more positive outlook on life, have better coping strategies for dealing with stress, are more emotionally stable, and so on. Essentially, high NA individuals respond differently to stressors compared to low NA individuals (Watson & Clark 1984).

A study by O'Brien et al. (2008) found that when presented with high demand tasks with low behavioral controls, people with high NA were more likely to use negative, and often emotion-based, coping strategies such as self-blame, and later reported lower levels of task satisfaction/job satisfaction. O'Brien et al. also found people high in NA to be particularly responsive to the amount of personal control they have in their work environment and very reactive to negative features of the environment. Furthermore, high NA individuals "may not respond to, or make use of, positive resources within the environment" (O'Brien et al. 2008: 76).

The constructs of self-efficacy, goal orientation and negative affectivity appear well-suited for further investigation in signed language interpreters, given that interest in some of these constructs, or related dimensions, have been touched on in the literature previously regarding the psychology of interpreters (Rudser & Strong 1986; Shaw et al. 2004). Furthermore, the findings drawn from studies on second language acquisition (Onwuegbuzie et al. 2000; Gregersen & Horwitz 2002) indicate that some of these factors may be worth exploring in interpreters. In addition, these constructs are acknowledged to have predictive value in the wider psychology literature on evaluating occupational performance and personality links (Bell & Kozlowski 2002).

Researchers like Skinner and Greene (2008) note that individual differences and perceived control are significant predictors of motivation, coping and success or failure in a wide variety of lifespan activities. Given the clear overlap between efficacy, perception of control/orientation, emotionality and the interplay of these constructs with personal competence, we should expect to see interesting patterns on these dimensions when measured in accredited signed language interpreters. The results of the investigation might then lead to the collection of these measures into a simple personality test battery to test interpreter aptitude (to be used in conjunction with other measures of cognitive ability, see for example Macnamara 2009). As predictive validity for these scales has already been established, the goal would be to use a simple test battery for screening purposes, with a view to determining aptitude for the interpreting skill, based on disposition factors.

Research hypotheses

To examine the unexplored psychological constructs in signed language interpreters, the following hypotheses were formulated:

1. Goal orientation will be positively related to perceived interpreter competence;
2. Self-efficacy will be positively related to perceived interpreter competence;
3. Negative affectivity will be inversely related to perceived interpreter competence;
4. Goal orientation, self-efficacy and negative affectivity will be salient predictors of interpreter competence.

These hypotheses were formulated on the basis of the literature and of the researchers' extensive expertise as working practitioners and educators of interpreters.

Methodology

To test the above hypotheses, a mail questionnaire was designed for distribution in Australia, drawing on literature from organizational psychology and personality psychology, interpreting and translation studies, and applied linguistics. The survey was designed to determine the incidence and distribution of, and interrelations among, sociological and psychological variables. We compared respondents' personal parameters (such as their route to qualification, level of qualification, years of interpreting experience, etc.) with their opinions and attitudes about general linguistic skill, other knowledge and abilities, overall competence and some personality measures. Demographic information was obtained in order to develop a profile of the participants and the profession, and to allow for examination of the relations among these variables as well as the overall interpreting competence rating reported by respondents.[2] In addition, participants were asked about their perception of the effectiveness of interpreter education programs for Auslan (Australian Sign Language)/English interpreters.[3]

2. Findings with regard to interpreter perceptions of competence and reported skills gaps are discussed in detail in Bontempo & Napier (2007).

3. A discussion of the perceived efficacy of interpreter education programs and program admission testing for Auslan interpreters can be found in Bontempo & Napier (2009).

Participants

Accreditation by the National Accreditation Authority for Translators and Interpreters (NAATI) as an Auslan/English interpreter (described hereafter as "Auslan" interpreter) was an essential criterion for participation in the study to ensure that only practitioners who met minimum benchmarks for work in the field could respond to the study. Survey respondents had passed an interpreting examination at a prior point in time (either via a NAATI-approved course of study or by direct NAATI testing) either at Paraprofessional or Professional Interpreter level.[4] NAATI accreditation is the only recognized licence to practice as an interpreter in Australia (in both spoken and signed languages), and federal legislation such as the Disability Discrimination Act, and state government language policies protect the rights of deaf people to demand a NAATI-accredited Auslan interpreter when accessing services in the wider community.[5]

As only NAATI-accredited Auslan interpreters could participate, potential subjects were identified and sourced via a number of avenues. Information regarding the study was distributed nationally using direct mailing lists and through snowball sampling. A flyer regarding the study was posted or e-mailed to 500 accredited Auslan interpreters on the NAATI mailing list at that time. All accredited Auslan interpreters were eligible and could self-select to participate in the study. Information was also distributed by the large employers of signed language interpreters in Australia at the time, and by the Australian Sign Language Interpreters' Association (ASLIA). Employers and ASLIA would have had access to most of the same population contacted directly by NAATI, with information estimated to have reached approximately 200 working interpreters via employers and approximately 300 members of ASLIA at that time, with some overlap.[6] It is estimated that approximately 500 accredited Auslan interpreters received information about the study via one or more sources.

Interested parties contacted the researchers and were then sent the questionnaire[7] by e-mail or regular mail, along with introductory information, and

4. Refer to NAATI website http://www.naati.com.au/at-accreditation.html for more details. Only Paraprofessional Interpreter and Professional Interpreter levels are available to Auslan interpreters.

5. For more details regarding the Australian context for signed language interpreting refer to Bontempo & Levitzke-Gray (2009) or Napier et al. (2010).

6. ASLIA allows student membership, and "inactive" interpreters can also retain membership. Membership of ASLIA is not compulsory for practice in Australia, so ASLIA membership numbers are not directly reflective of the number of accredited and active Auslan interpreters in Australia.

7. Approved by the ethics committee of Macquarie University and subject to standard requirements for data collection.

a stamped addressed envelope if requested. A total of 82 Auslan interpreters requested a copy of the questionnaire. Surveys were not coded in any way, so it is not possible to determine if all of the requesting parties returned a completed questionnaire.

A total of 110 completed questionnaires were received from interpreter respondents, which is more than the number of requests received for the survey. It is assumed that some respondents may have passed a copy of the questionnaire on to colleagues, or that some employers forwarded copies of the survey directly to interpreter employees.

Whilst 722 interpreters had been accredited by NAATI between 1982 and the release of the survey in early 2005, a report commissioned by the Federal Government Department of Family and Community Services noted that only 257 accredited interpreters were working in the field at that time (Orima 2004). This was a little more than the figure reported by employers (approximately 200 active interpreters were sent the flyer by employers) and a little less than the national ASLIA membership at the time (approximately 300 members) and therefore appears to be an accurate reflection of the number of working practitioners in early 2005.

Thus an estimated response rate of 42% (110 respondents / an assumed pool of 257 working practitioners) was considered more than adequate. This is considered a higher than average return rate in survey methodology, whereby a return rate of 20–30% is deemed to be acceptable (Jackson 2003).

The instrument

The survey instrument was a 10-page questionnaire, including a carefully planned set of questions based on the literature. A total of 22 questions were presented, arranged in related subsets of five main sections — demographic information, skills gap information, perceptions of performance, interpreter education programs/training options, and personality rating scales, which were drawn from existing psychometric measures and are well documented in the literature.

In summary, a combination of open-ended, close-ended, partially open ended, and various Likert rating scales (with five options, to obtain interval data) were to be completed by participants. The rating scales pertained to overall competency as an interpreter, a detailed skills gap analysis, and various self-reporting personality measures of self-efficacy, negative affectivity, and goal orientation.

The first ten questions collected sociological data, to develop a profile of the profession in Australia. These included closed questions and partially open-ended questions regarding work status, accreditation level, year of accreditation, work setting, age group, gender, state or territory of residence, first language, secondary

schooling, post-secondary schooling and extent of formal interpreter education completed.

Question 11 listed fifty defined skills and areas of knowledge that may be relevant to signed language interpreters, as drawn from the literature. Participants were asked to rate the importance of each skill, knowledge or ability, and correspondingly, offer their assessment of their own competence in that particular skill or knowledge domain. This information provided the researchers with quantifiable skills gap data.

Questions 12 to 17 were open-ended, asking participants to express their thoughts on additional skills, knowledge or abilities of an Auslan interpreter not listed in question 11: the effectiveness of interpreter training; reasoning for decisions in regard to work selection; and perceptions of performance. These questions were designed to provide qualitative data, which could then be cross-referenced with the quantitative data collected.

Respondents were then asked to rate themselves on a scale, assessing their overall competence as an interpreter on question 18. This information would provide a key variable in relation to interpreters' responses on other sociological and psychological variables.

Questions 19, 20 and 21 were scales with an established history of use in the fields of organizational and personality psychology, assessing the constructs of goal orientation, negative affectivity and self-efficacy respectively.

The goal orientation scale developed by Button et al. (1996) formed question 19. This scale is used to measure achievement motivation in individuals, by tapping into learning goal orientation. Of the range of popular goal orientation scales available, it is considered the most reliable and valid, and Button et al. (1996) provide evidence of its construct validity. Other researchers have also found it to be particularly reliable (Bell & Kozlowski 2002). An 8-item set of questions for this dispositional variable was also included. Respondents' answers ranged from *strongly disagree* (1) to *strongly agree* (5).

Question 20 was the Positive and Negative Affect Schedule (PANAS) developed by Watson, Clark and Tellegen (1988), a tool that demonstrates high levels of internal consistency and stability, and is widely considered a reliable, valid and efficient means of measuring dimensions of positive and negative affectivity. It correlates highly with other tools measuring similar constructs (for example, the NA items on the PANAS correlates well with the Beck Depression Inventory). The 10-item version (for each — PA and NA) of the PANAS was administered in the survey, and respondents scored their answers on a Likert scale of from *very slightly/not at all* (1) to *extremely* (5).

The last scale at question 21 was a self-efficacy measurement tool, using the New General Self-Efficacy scale (NGSE) developed by Chen et al. (2001). This

scale is internally consistent and stable, has high content and predictive validity, and consists of only 8 items, making it more attractive than previous iterations of general self-efficacy scales. Respondents answered questions on a response scale of *strongly disagree* (1) to *strongly agree* (5).

Finally, at the open-ended question 22, participants were encouraged to add comments and anything further if they so wished.

The first draft of the questionnaire was piloted with two Paraprofessional Interpreters and one Professional Interpreter to obtain feedback regarding the comprehensibility of the material, and suitability of the line of questioning. Following the review and feedback by colleagues, some minor amendments were made to the preliminary version before it was released to participants in the study.

Procedure

Participants in the study completed the survey instrument after receiving it in the mail, or via e-mail. Questionnaires were estimated to take up to 40 minutes to complete, and respondents completed the survey in English (handwritten or typed responses were possible) at their leisure and in their own preferred environment. Participants had access to information about the study and potential possession of the questionnaire for up to 8 weeks, and returned their questionnaires by post or e-mail upon completion. Analysis of the completed questionnaires centered on any areas of significance, using descriptive as well as parametric and non-parametric inferential statistical analysis.

Results

A total of 110 NAATI-accredited signed language interpreters returned completed questionnaires. 67.3% of these were accredited at Paraprofessional level and 32.7% at Professional Interpreter level. The age range of participating interpreters varied from 20 years to over 60 years, but the largest group was aged between 30 and 49 years (74.5%). The gender profile of respondents was 83.5% female and 16.5% male. Tertiary education qualifications issued by either a Technical and Further Education (TAFE) college, or a university, were held by 86.4% of respondents.

Details regarding demographic findings and skills gaps for interpreters are reported in Bontempo & Napier (2007), but in summary, the sociological data collected from participants offered a profile of the "typical" Auslan interpreter respondent. She is female, between 30 and 49 years of age, works part-time as an interpreter in community interpreting assignments, is accredited at NAATI Paraprofessional level, has English as her first language, has undertaken formal

interpreter training and has a post-secondary qualification of some type (interpreting-related and/or a generic qualification).

The key concern of this paper is the findings in regard to questions 19 to 21 of the survey, and their relationship with perceived interpreter competence (addressed in question 18). The psychometric tools used elicited data on self-ratings of goal orientation, self-efficacy and negative affectivity.

Hypotheses 1–3 predicted that goal orientation (Hypothesis 1) and self-efficacy (Hypothesis 2) would be positively related to self-perceived interpreter competence, whereas negative affectivity (Hypothesis 3) would be negatively related.

As shown in Table 1, self-perceived interpreter competence is positively related to self efficacy ($r = .21$, $p < .05$) and negatively related to negative affectivity ($r = -.26$, $p < .01$). Perceived interpreter competence was not, however, significantly related to goal orientation.

In summary, hypotheses 2 and 3 were supported whilst hypothesis 1 was not.

Table 1. Means (*M*), standard deviations (*SD*) and intercorrelations

Measures	*M*	*SD*	1	2	3	4
1. Goal orientation	4.22	.43	–			
2. Negative affectivity	1.83	.62	.02	–		
3. Self-efficacy	4.16	.42	.44**	−.25*	–	
4. Interpreter competence	3.53	.65	.09	−.26**	.21*	–

* = $p < .05$; ** = $p < .01$

Hypothesis 4 predicted that goal orientation, self-efficacy and negative affectivity would be salient predictors of perceived interpreter competence. A multiple regression analysis was conducted in which interpreter competence was regressed on all of the personality constructs.

As shown in Table 2, the overall model comprising the three personality factors was significant, explaining 9% of the variance in ratings of interpreter competence, $F = 3.14$, $p < .05$. However, only negative affectivity ($\beta = -.23$, $p < .05$.) was a significant predictor of interpreter competence.

Table 2. Multiple regression analysis for interpreter competence (* = $p < .05$)

Variables	β	Overall R^2
Goal orientation	.04	
Negative affectivity	−.23*	
Self-efficacy	.13	
		.09*

The findings indicate that goal orientation was not a personality factor of major interest and was not a predictor of interpreter performance in this study. Self-efficacy was only of minor interest. Negative affectivity on the other hand was a significant predictor of perceived competence, although the results were not as compelling as anticipated. Effectively however, people who had high levels of negative affectivity reported lower levels of competence as interpreters.

Discussion and conclusions

Factors of goal orientation, self-efficacy and negative affectivity were found to account for 9% of overall variance in ratings of interpreter competence in the study. Although by no means an overwhelming figure, the result is statistically significant, and does demonstrate personality impacts on interpreter's self-perceptions of competence. This finding is supported by the reports of other researchers (e.g. Dörnyei 2005), who found that personality variables may account for up to 15% of the variance in individual difference in second-language learning achievement.

Emotional stability, as measured by the negative affectivity scale in this study, demonstrated the strongest impact in regard to perceived competence. It is known that people with a high level of negative affectivity may not make use of positive coping resources and may be more reactive to job-related stressors. Moreover, such individuals are likelier to create stressful work situations because of their behavioral tendencies (O'Brien et al. 2008). It follows that the findings in this study may have some practical applications when it comes to occupational stress management in interpreters. The emotional stability factor may add to our understanding of practitioner responses to situations of high demand and their use of coping strategies and resources to mitigate work pressures, expectations and challenges.

The findings herein complement and expand on the work of Rudser and Strong (1986), Seal (2004), Shaw and Hughes (2006), Stauffer and Shaw (2006) and López Gómez et al. (2007). They are noteworthy in that they provide further support for the emerging picture of a relationship between personality and work performance of signed language interpreters.

Further investigation of the predictive power of personality factors in relation to interpreter performance may be valuable, and may contribute to our unraveling of the predictors of interpreting skill as well as more effective assessment of aptitude for interpreting.

Implications for interpreter education

It appears that we may eventually benefit from considering the notion of personality traits upon admission into training programs. At this stage, however, since the predictive potential of the range of factors that may impact on interpreter aptitude and performance is not fully understood, it may be unwise to implement formal personality testing in program admission screening. However, the fact remains that a clear relationship between occupational performance and general cognitive ability, conscientiousness and emotional stability does figure in the scholarly literature (Schmidt & Hunter 1998; Judge et al. 1999). This body of research, combined with recent findings of interpreter studies such as this one, suggest that attention to interpreter disposition should be considered for screening purposes at some point in the future. The data available to us firmly indicates that in selecting for success, "organizations will be better off selecting individuals who are conscientious and emotionally adjusted" (Judge et al. 1999: 647) — though we have yet to determine the manner in which this can best be achieved.

Even if they cannot screen for specific personality traits and general cognitive ability at this stage, interpreter education program coordinators and lecturers may take comfort in the fact that simply being aware of the impact of trait tendencies (such as negative affectivity) allows for the opportunity to mitigate effects in the classroom or in workplaces. Incorporating trait awareness into interpreter training, and developing aspects such as self-confidence, positive coping skills, assertiveness and resilience would most certainly be useful, given the broader evidence pointing to these aspects of personality as relevant for effective occupational performance.

Teaching interpreting students and accredited practitioners to better manage anxiety and occupational stress may be conducive to improving their interpreting performance. In addition, providing practitioners with appropriate formal support structures within the workplace (e.g. debriefing, supervision, mentoring etc.) may promote emotional well-being and lead to improved competence and performance.

Although it is understood that aptitude alone cannot predict how an interpreter trainee will progress through a program of study and into the profession — which is clearly dependent on a range of learner-based characteristics as well as the interaction between learner and teacher, the instructional conditions and the learning context — testing for aptitude can nevertheless provide us with important information. With significant sums of money and time expended annually in interpreter education programs world-wide on students who fail to make the grade (Timarova & Ungoed-Thomas 2008; Bontempo & Napier 2009), it is certainly worth the effort to try and better define aptitude for learning interpreting.

Whether this leads to more stringent screening of candidates or impacts the curricula of interpreter education programs remains to be seen. What we do know is that, in the signed language interpreting field in some countries, such as Australia, the current supply of qualified interpreters does not meet the demand in the marketplace (Orima 2004). The danger lies in a situation where programs of study, and employers, may be disinclined to implement discretionary screening procedures, even if based on valid psychometric tools, simply because the industry demand for interpreters is too great to be turning people away.

The thorny notion of testing for "personality" at program admission is further exacerbated by the possibility that such testing could potentially be perceived as discriminatory according to the legislation in some countries. However, considering that clear relationships exist between individual difference variables such as general cognitive ability and personality traits and resulting job performance in a wide range of occupations (Judge et al. 1999), there is no reason why interpreting should be any different. In the knowledge that interpreting is a job with high stakes, where public trust is paramount, and the task itself is extremely complex as well as intrinsically stressful, the screening and selecting of interpreters could be organized far more effectively. It is argued that taking such rigorous steps to screen at program admission stage would ultimately lead to more competent and resilient interpreters completing courses with fewer skills gaps than is presently the case (Bontempo & Napier 2007, 2009).

In addition to the potential future application of the findings from this study in interpreter education and program admission, we also see the potential for findings regarding "interpreter personality" as being useful for the development of mentoring programs for working interpreters. Such information may provide legitimate data to support focusing on the strengths and weaknesses of individuals when matching mentors and mentees in the workplace. A further possible occupational application of findings related to the link between disposition and interpreter performance is in regard to team interpreting, where personality is intuitively understood to impact on team dynamics. It is likely that the personality construct of agreeableness comes into play in synergistic teamwork, and it would be interesting to investigate this further.

Mount and Barrick (1998: 856) put it most succinctly: "No matter what job you are selecting for, if you want employees who will turn out to be good performers, you should hire those who work smarter and work harder." Appreciating and incorporating the disposition predictors as well as the cognitive predictors of the interpreting skill in any aptitude screening process will help us move closer to this ideal. There is an indication that emotional stability is a predictor of work performance for interpreters. This provides grounds for optimism in working towards defining the traits that may influence interpreter performance, and should provide

the impetus for a research agenda[8] which will further investigate personality–performance links in interpreters.

References

Barrick, M. R. & Mount, M. K. (2005). Yes, personality matters: Moving on to more important matters. *Human Performance* 18, 359–372.

Barrick, M. R., Mount, M. K. & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *Personality and Performance* 9, 9–30.

Bell, B. S. & Kozlowski, S. W. J. (2002). Goal orientation and ability: Interactive effects on self-efficacy, performance and knowledge. *Journal of Applied Psychology* 87, 497–505.

Bontempo, K. & Levitzke-Gray, P. (2009). Interpreting Down Under: Signed language interpreter education and training in Australia. In J. Napier (Ed.), *International perspectives on sign language interpreter education*. Washington, DC: Gallaudet University Press, 149–170.

Bontempo, K. & Napier, J. (2007). Mind the gap: A skills analysis of sign language interpreters. *The Sign Language Translator and Interpreter* 1 (2), 275–299.

Bontempo, K. & Napier, J. (2009). Getting it right from the start: Program admission testing of signed language interpreters. In C. V. Angelelli & H. E. Jacobson (Eds.), *Testing and assessment in translation and interpreting studies*. Amsterdam/Philadelphia: John Benjamins, 247–295.

Bozionelos, N. (2004). The relationship between disposition and career success: A British study. *Journal of Occupational and Organisational Psychology* 77, 403–420.

Brisau, A., Godijns, R. & Meuleman, C. (1994). Towards a psycholinguistic profile of the interpreter. *Meta* 39 (1), 87–94.

Button, S. B., Mathieu, J. E. & Zajac, D. M. (1996). Goal orientation in organisational research: A conceptual and empirical foundation. *Organisational Behaviour and Human Decision Processes* 67 (1), 26–48.

Chen, G., Gully, S. M. & Eden, D. (2001). Validation of a new general self-efficacy scale. *Organizational Research Methods* 4 (1), 62–83.

Choi, N., Fuqua, D. R. & Griffin, B. W. (2001). Exploratory analysis of the structure of scores from the multidimensional scales of perceived efficacy. *Educational and Psychological Measurement* 61 (3), 475–489.

Doerfert, K. & Wilcox, S. (1986). Meeting students affective needs: Personality types and learning preferences. *Journal of Interpretation* 3, 35–43.

Dörnyei, Z. (2005) *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.

Dweck, C. S. & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review* 95, 256–273.

8. An international collaborative study to further investigate personality variables as predictors of interpreter competence has been initiated between Macquarie University, Australia (K. Bontempo and J. Napier) and Eastern Kentucky University, USA (L. Hayes and associates).

Fouad, N. A. (2007). Work and vocational psychology: Theory, research and applications. *Annual Review of Psychology 58:* 543–564

Frishberg, N. (1990). *Interpreting: An introduction*. 2nd edn. Silver Spring, MD: RID Publications.

Gregersen, T. & Horwitz, E. K. (2002). Language learning and perfectionism: Anxious and non-anxious language learners' reactions to their own oral performance. *The Modern Language Journal* 86 (iv), 562–570.

Isham, W. & Lane, H. (1993). Simultaneous interpretation and the recall of source language sentences. *Language and Cognitive Processes* 8 (3), 241–264.

Jackson, S. L. (2003). *Research methods and statistics: A critical thinking approach*. Belmont: Wadsworth Thomson.

Judge, T. A., Higgins, C. A., Thoreson, C. J. & Barrick, M. R. (1999). The Big Five personality traits, general mental ability, and career success across the life span. *Personnel Psychology* 52, 621–652.

Kurz, I. (2003). Physiological stress during simultaneous interpreting: A comparison of experts and novices. *The Interpreters' Newsletter* 12, 51–67.

López Gómez, M. J., Bajo Molina, T., Padilla Benitez, P. & Santiago de Torres, J. (2007). Predicting proficiency in signed language interpreting. *Interpreting* 9 (1), 71–93.

Macnamara, B. (2009). Interpreter cognitive aptitudes. *Journal of Interpretation* 2008–2009, 9–32.

Moser-Mercer, B. (1985). Screening potential interpreters. *Meta* 30 (1), 97–100.

Mount, M. K. & Barrick, M. R. (1998). Five reasons why the "Big Five" article has been frequently cited. *Personnel Psychology* 51, 849–857.

Napier, J. (2011). Signed language interpreting. In K. Windle & K. Malmkjær (Eds.), *The Oxford handbook of translation studies*. New York: Oxford University Press, 357–376.

Napier, J., McKee, R. & Goswell, D. (2010). *Sign language interpreting: Theory and practice in Australia and New Zealand* (2nd ed.). Sydney: Federation Press.

O'Brien, A., Terry, D. J. & Jimmieson, N. L. (2008). Negative affectivity and responses to work stressors: An experimental study. *Anxiety, Stress and Coping* 21 (1), 55–83.

Oakes, D. W., Ferris, G. R., Martocchio, J. J., Buckley, M. R. & Broach, D. (2001). Cognitive ability and personality predictors of training program skill acquisition and job performance. *Journal of Business and Psychology* 15 (4), 523–548.

Ones, D. S., Dilchert, S., Viswesvaran, C. & Judge, T. A. (2007). In support of personality assessment in organizational settings. *Personnel Psychology* 60 (4), 995–1027.

Onwuegbuzie, A. J., Bailey, P. & Daley, C. E. (2000). Cognitive, affective, personality and demographic predictors of foreign language achievement. *Journal of Educational Research* 94 (1), 3–15.

Orima (2004). *Supply and demand for Auslan interpreters across Australia*. Australian Government, D. of Families, Housing, C. Services and Indigenous Affairs. http://www.fahcsia.gov.au/sa/disability/pubs/policy/Pages/access-auslan_report.aspx (accessed 5 October 2010).

Phillips, P., Abraham, C. & Bond, R. (2003). Personality, cognition and university students' examination performance. *European Journal of Personality* 17, 435–448.

Ree, M. J. & Earles, J. A. (1992). Intelligence is the best predictor of job performance. *Current Directions in Psychological Science* 1 (3), 86–89.

Ree, M. J., Earles, J. A. & Teachout, M. S. (1994). Predicting job performance: Not much more than "g". *Journal of Applied Psychology* 79 (4), 518–524.

Reeve, C. L. & Heggestad, E. D. (2004). Differential relations between general cognitive ability and interest-vocation fit. *Journal of Occupational and Organisational Psychology* 77, 385–402.

Rudser, S. F. & Strong, M. (1986). An examination of some personal characteristics and abilities of sign language interpreters. *Sign Language Studies* 53, 315–331.

Salgado, J. F. (1997). The five-factor model of personality and job performance on the European community. *Journal of Applied Psychology* 82 (1), 30–43.

Schmidt, F. L. & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin* 124 (2), 262–274.

Seal, B. C. (2004). Psychological testing of sign language interpreters. *Journal of Deaf Studies and Deaf Education* 9 (1), 39–52.

Shaw, S. & Hughes, G. (2006). Essential characteristics of sign language interpreting students: Perspectives of students and faculty. *Interpreting* 8 (2), 195–221.

Shaw, S., Grbić, N. & Franklin, K. (2004). Applying language skills to interpretation. *Interpreting* 6 (1), 69–100.

Skinner, E. & Greene, T. (2008). Perceived control: Engagement, coping, and development. In T. L. Good (Ed.), *21st century education: A reference handbook*. Newbury Park: Sage Publications, 121–130.

Stauffer, L. & Shaw, S. (2006). Personality characteristics for success in interpreting courses: Perceptions of spoken and signed language interpretation students. *Journal of Interpretation* 2006, 11–24.

Timarová, Š. & Ungoed-Thomas, H. (2008). Admission testing for interpreting courses. *The Interpreter and Translator Trainer* 2 (1), 29–46.

Watson, D. & Clark, L. A. (1984). Negative affectivity: The disposition to experience aversive emotional states. *Psychological Bulletin* 96 (3), 465–490.

Watson, D., Clark, L. A. & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology* 54 (6), 1063–1070.

Domain-general cognitive abilities and simultaneous interpreting skill

Brooke N. Macnamara, Adam B. Moore, Judy A. Kegl* and Andrew R. A. Conway
Princeton University and University of Southern Maine*

This exploratory study examined domain-general cognitive abilities that may serve as aptitudes for interpreting skill by comparing highly skilled signed language interpreters (those considered competent in most interpreting situations) and less skilled signed language interpreters (those considered less than competent in most interpreting situations) on various measures. Specifically, the current study examined the feasibility of predicting interpreter skill level based only on a variety of cognitive abilities and personality traits. We collected data on several cognitive measures, including processing speed, psychomotor speed, cognitive control and task switching ability, fluid intelligence, working memory capacity, and mental flexibility, as well as several personality measures, including risk-taking orientation and emotion-cognition integration style, and intrinsic motivation to engage in complex cognitive tasks. Significant differences emerged between the two groups on both cognitive and personality measures suggesting that a combination of stable domain-general cognitive abilities and personality traits may be responsible for differentiating highly skilled from less skilled interpreters and may therefore be predictive of individuals' future interpreting effectiveness and skill level.

Introduction

Simultaneous interpreting (SI) is considered an extremely complex cognitive processing task (Christoffels et al. 2003; Frauenfelder & Schriefers 1997; MacWhinney 1997). Its basic components are similar to the processes engaged during normal monolingual dialogue: listening, comprehending, communication planning, and language production. However, in normal dialogue these processes take place serially, with some overlap of comprehension and utterance planning (Garrod & Pickering 2004). What makes SI so complex is that the interpreter is simultaneously

listening, comprehending, planning, and speaking, and is doing so in two different languages with little to no control over the input rate or content. In addition, the interpreter is alternately activating and suppressing the two languages, and analyzing the speaker's goals, inferences, and subtleties while deciding how to convey the meaning in a second language and culture all in real time.

Throughout the process, the interpreter attends to the incoming message and output while holding information in memory (Cowan 2000; Macnamara 2009: 19–20; Moser 1978; Shlesinger 2003), manages the process and the demands (Christoffels et al. 2003: 202; Macnamara 2009: 18–19), analyzes and reasons (Macnamara 2009: 16; Seal 2004: 49), and makes decisions based on analyses, demands, and abilities (Macnamara 2009: 22–23; Treisman 1965: 369). These information-processing demands are not limited to linguistic tasks alone; ample psychological research has demonstrated that general cognitive abilities have substantial explanatory power across multiple content domains, including working memory capacity (e.g., Daneman & Carpenter 1980; Engle et al. 1991; Kiewra & Benton 1988; Kyllonen & Stephens 1990; Ormrod & Cochran 1988), reasoning ability (e.g., Johnson-Laird 1999: 113; Ree & Earles 1992; Sternberg 1982), and psychomotor speed (e.g., Ackerman 1988; Ownby et al. 2008). These domain-general cognitive abilities are typically assumed to be fairly innate qualities since they are present early in development (e.g., Bayliss et al. 2003; Starkey 1992) and, after adulthood, are relatively stable over time (e.g., Rund 1998: 426–428; Wichertsa et al. 2004; Waters & Caplan 2003). It is therefore reasonable to assume that not only will language and other skill learning predict interpreting performance, but individual differences in these more general cognitive abilities will predict interpreting aptitude as well.

Gerver et al. (1989) studied differences in a variety of discourse processing and verbal abilities between passing and failing interpreter students. Their results indicated that passing interpreter students had better memory for texts, comprehension, and verbal generation. Discourse-processing abilities such as those measured by Gerver et al. are mediated by working memory and more domain-general cognitive skills (Gernsbacher 1990; Just & Carpenter 1992; Kintsch 1988; Kintsch & van Dijk 1978). In other words, domain-general ability can predict performance differences in these language tasks over and above what can be explained by linguistic skill and domain-specific training alone.

Despite the plethora of evidence for the predictive power of domain-general cognitive abilities in psychological research, evidence as to the relationship between these types of measures and interpreting performance has been mixed. In a study of domain-general cognitive abilities with signed language interpreter trainees, López Gómez et al. (2007) found positive correlations between interpreting skill and short-term memory, a confidence-type personality trait, and reasoning

ability. However, other studies have shown that working interpreters demonstrate higher than average reasoning abilities, but that the degree of intelligence did not reliably differentiate the interpreters based on skill (Rudser & Strong 1986; Seal 2004). Similarly, examinations of working memory capacity of interpreters and interpreting students or non-interpreters have shown mixed results. Köpke and Nespoulous (2006) found that interpreter students had larger working memory capacity than expert interpreters, suggesting that interpreting experience is detrimental to working memory capacity while Padilla et al. (1995) and Christoffels et al. (2006) found that interpreters demonstrated higher working memory capacity than non-interpreters, suggesting that interpreting experience is beneficial to working memory capacity.

The discrepancies observed in previous research may be due to several factors. First, the limited nature of the population studied may compromise the comparability of the experimental subjects in the respective studies and/or the statistical analysis may be underpowered to detect meaningful effects. Second, differences that are found between interpreters and non-interpreters in cross-sectional designs could be due either to changes in cognitive ability as a *result* of interpreting experience or self-selection of individuals with certain cognitive abilities into the interpreting field. Third, the tasks used to measure cognitive abilities of interest are not all equally valid or reliable. Fourth, some cognitive abilities/traits will make no difference in one's future interpreting skill; some may relate to interpreting skill only up to a certain point (i.e. the first few months/years of training), while others may still correlate with interpreting performance regardless of experience or training. Finally, and perhaps most importantly, performance in complex real world interaction(s) depends on the interplay of many cognitive abilities, so that even those studies that have measured multiple abilities/traits but only analyzed first-order relationships between each measure and population type or skill level will have missed the interactive contingencies that are certain to exist.

This is a long series of challenges, and no one study can meet them all. Here, we focus centrally on the last point as a first step; through the use of tasks that have excellent psychometric validity and reliability coupled with advanced statistical analysis, we seek to assess which *combination* of general cognitive abilities and emotion-cognition interaction traits predict interpreting expertise. In the United States there is a wide variety of standards for the level of skill required for graduation and professional practice. This set of circumstances surrounding American Sign Language (ASL)-English interpreters allows us to study groups of simultaneous interpreters who have similar SI experience and training, but who vary in SI skill.

We are interested in comparing cognitive abilities of highly skilled and less skilled simultaneous interpreters to determine which combination of cognitive

abilities and personality traits predicts group membership. In other words, if highly skilled interpreters have similar cognitive abilities, which are typically assumed to be stable over time regardless of training or experience, then this combination is likely to predict who will become a highly skilled interpreter. Note: While the interpreters studied in this paper are ASL-English interpreters, we assume that cognitive processes in SI are largely similar regardless of language modality. Unless specifically noted, the authors are interpreting the measurements used in this paper as applicable to all simultaneous interpreters.

Descriptive cognitive process models illustrating the complexity of SI began to appear in the 1970s (e.g., Gerver 1976). As both cognitive psychology and SI research emerged, our understanding of the specific cognitive mechanisms engaged during SI developed (cf. e.g. Moser 1978) and today we can reasonably argue for the involvement of several key cognitive processing abilities and personality traits known to impact the constituent processes of SI. This paper specifically addresses the following cognitive processing abilities: reasoning, working memory capacity, processing speed, cognitive control, psychomotor speed, and mental flexibility.

Reasoning is essential for linguistic, environmental, and affective analysis of the source message for comprehension and prediction (Cokely 1992; Colonomos 1997, 2008; Macnamara 2009; Moser 1978) as well as planning the target message output (Cokely 1992; Colonomos 1997, 2008; Macnamara 2009).

Working memory is the simultaneous storage and processing of information in the short term, often when the information being operated upon is different than that which must be stored. Working memory capacity, one's limit of information that can be stored while simultaneously carrying out a processing task, is positively correlated with language comprehension (Daneman & Carpenter 1980) and discourse processing ability (Just & Carpenter 1992; Kintsch 1988; Kintsch & van Dijk 1978; Gernsbacher 1990), essential components of SI. In addition to language processing, working memory is also critically involved in real-time problem-solving, reasoning, and planning, as well as manipulating or transforming incoming information — all of them abilities certainly needed during SI.

Simultaneous interpreting demands that information be processed rapidly. High working memory capacity and robust reasoning ability are not useful to interpreters if they cannot process the incoming information and execute decisions at a rate faster than or consistent with the incoming information. In addition, information that is quickly processed, and no longer held in the focus of attention, allows more attentional capacity (Moser 1978).

Psychomotor speed and accuracy, or perceptual-motor coordination, allows signed language interpreters to produce the target message, when interpreting into the signed language, with precision. Psychomotor skill is also involved in physical mimicry, playing a role in learning and producing manual signs (López Gómez et

al. 2007). Psychomotor skill is assumed to be applicable only to signed language interpreters as opposed to both spoken and signed language interpreters (López Gómez et al. 2007).

During SI, the interpreter rapidly switches among subcomponents of interpreting: comprehending the source message, determining meaning equivalences, planning the production, and producing the target message (Cokely 1992; Colonomos 1997; Moser 1978). Cognitive control is engaged in order to switch among tasks and to manage task execution effectively (Monsell 2003).

Interpreting is a practice profession. Practice professions (e.g. medicine, teaching, counseling, law, and investigation) require technical knowledge and skills, but perhaps more importantly, they require assessments of ever-changing situational and human interaction factors that impact how the technical knowledge and skills should be implemented in each situation (Dean & Pollard 2005). Interpreters cannot perfectly predict the incoming message and will rarely interpret the same source message more than once. Decisions made while interpreting are adjusted for constantly changing situations as a function of this indeterminism. Interpreters rely on adaptive responses when handling the incoming message and other human interaction factors. One's capacity to adaptively coordinate actions in relation to others' actions in interpreting relies on mental flexibility, the final cognitive ability measured in this paper.

Cognitive abilities do not exist in a vacuum. This paper therefore also considers specific personality traits known to impact cognitive processing: willingness to engage in complex cognitive tasks, reward sensitivity, and risk sensitivity. We hypothesize that because SI is a complex cognitive task, certain cognitive abilities are necessary to successfully perform it. However, one's willingness to engage in such a complex cognitive task will also affect the amount of effort undertaken. Willingness to employ mental resources during the task affects performance and thus interacts with other cognitive abilities (Cacioppo & Petty 1982).

In addition to their willingness to engage in cognitive tasks, individuals vary in their motivation to engage cognitive control and make decisions based on sensitivities to the potential outcomes. To this end, individuals vary in their sensitivity to reward and risk (e.g., Gray 1982, 1987; Lopes 1987; Schneider & Lopes 1986). Individuals' sensitivity to reward motivates behavior toward subjectively positive outcomes, and increases the likelihood of decisions designed to approach desirable goals (Gray 1982).

Individuals with high risk sensitivity, on the other hand, experience anxiety when presented with potential threat, non-reward, or novelty and will structure their behavior around avoiding risk of aversive outcomes, as opposed to achieving positive outcomes (Gray 1982, 1987, 1990). Individuals with high anxiety will experience reduced working memory capacity (Eysenck 1979, 1985; Leon &

Revelle 1985; Schmader & Johns 2003; Wine 1971) and are likelier to respond more quickly to stimuli and produce more errors (Leon & Revelle 1985). SI demands that interpreters process novel, unrehearsed stimuli and adjust processing time (aka ear-voice span) based on the incoming message and interpreting abilities. An interpreter's risk sensitivity may therefore affect how she or he responds to the incoming stimuli and subsequent cognitive processing.

We have described several cognitive abilities that we believe are relevant to spoken and signed language simultaneous interpreters: reasoning, working memory capacity, processing speed, cognitive control, task switching, and mental flexibility; and one cognitive ability relevant to signed language interpreters only: psychomotor speed. We have also described personality traits that interact with cognitive abilities that we believe to be relevant to both spoken and signed language simultaneous interpreters: need for cognition, reward sensitivity, and risk sensitivity. The issue addressed in this paper is not whether these cognitive abilities and emotion-cognition interaction traits are engaged during interpreting, but whether they can predict expertise in SI and, if so, what combination of these traits best differentiates skill level.

Method

Participants

Twenty-nine ASL-English interpreters from two diverse locations in the United States participated in the study. Ages ranged from 20 to 60 ($M = 42$, $SD = 12$). Interpreting experience ranged from 6 months to 35 years ($M = 11$, $SD = 11$). Participants were not recruited on the basis of age or years of professional experience. Participants had similar interpreting training.

Rating

Five raters were used to classify interpreters based on general interpreting competency. The raters were ASL-English interpreters who were familiar with the simultaneous interpreting work of the participants in a variety of settings within the prior year, and with a variety of consumers. These observations were the basis of their ratings. While participants did not undergo a standardized laboratory-based competency exam, the multiple real-life observations along with the extremely high inter-rater reliability (discussed later in this section) suggest that this method has high ecological validity. All of the raters also had previous interpreter rater/evaluation training.

Competency was defined as performing well in skill-based aspects of the interpreting process including comprehension, language production,[1] message equivalence, and the ability to perform flexibly along the interpreting-transliterating continuum depending on the appropriateness to the situation. Transliterating, retaining the word order and syntax of the source language while producing target language words, is preferred by some Deaf persons and may be understood to various degrees by others. Many signed language interpreters provide transliterating services even when interpreting service is necessary for monolingual ASL-using Deaf individuals. Therefore, some of the participants who work competently in many situations as transliterators were not placed in the highly skilled group if they did not exhibit competence to interpret into sign following ASL grammar.[2] Interpreters were not rated on ethical behavior, maturity, attitude toward consumers or colleagues, or other professional behavior.

The raters scored the interpreters on a three-point scale as (1) highly skilled, (2) somewhat skilled, and (3) less skilled. The two raters for the first location had very high inter-rater reliability, Cronbach's $\alpha = .978$. Rater 1–1 was not familiar with two participants' interpreting skill and Rater 2–1 was not familiar with two other participants' interpreting skill. Because the inter-rater reliability was highly correlated between the two raters, the rating from the familiar rater was accepted. The three raters for the second location also had very high inter-rater reliability, Cronbach's $\alpha = .958$. While the first two raters at the second location had very high inter-rater reliability, Rater 1–2 and Rater 2–2 Cronbach's $\alpha = .935$, both raters were unfamiliar with one participant's interpreting skill. A third rater was secured who was familiar with the skill level of the participant in question. The third rater demonstrated high inter-rater reliability with both the other raters, Cronbach's $\alpha = .935$ and 1.0, respectively. The third rater's rating of the participant in question was accepted.

Since there were only two cases in which a rater rated a participant as somewhat skilled (the only two in which raters differed), and since inter-rater reliability was extremely high, the rating of the other rater was accepted. For the first instance in the first location, the participant was rated by one rater as in-between the two skill groups and by the second rater as less skilled — and was placed in the less skilled group. For the second case in the second location, one rater rated the participant as in-between the two skill groups. The other two raters rated the

1. Language skills were rated as an aspect of interpreting skills. Language ability is strongly correlated with interpreting ability (e.g., Padilla et al. 1995).

2. Interestingly, not exhibiting competence to interpret into sign following ASL grammar is not necessarily a language ineptitude. Many transliterators can produce accurate ASL in conversation, but not while interpreting, suggesting a cognitive or emotion-cognition interaction trait may be responsible for this discrepancy.

participant as less skilled and the participant was placed in the less skilled group. Due to the generally dichotomous ratings, as participants were generally rated as either highly skilled or less skilled and the few instances of *somewhat skilled* ratings were the only ones not agreed upon, the middle group was removed and the two groups were termed *highly skilled* and *less skilled*. There were no instances of polar classifications (one rater placing a participant in the less skilled group and another placing the same participant in the highly skilled group).

The highly skilled group consisted of 15 interpreters and the less skilled group consisted of 14 interpreters. There were no significant differences between the two groups in age (highly skilled group $M = 42$, $SD = 13$; less skilled group $M = 42$, $SD = 10$) or years of professional interpreting experience (highly skilled group $M = 12$, $SD = 11$, less skilled group $M = 10$, $SD = 10$), both Fs < 1.

At first, the lack of correlation between years of experience and rated skill appears counter-intuitive. Expertise research assumes that experts, individuals who consistently perform superiorly to the majority of practitioners, have accumulated over ten years of domain-specific experience (e.g., Chi et al. 1988; Hoffman 1992; Simon & Chase 1973). Experience alone, however, does not necessarily make an expert (Ericsson et al. 1993) since it is not a good predictor of proficiency (Ericsson et al. 1993; McDaniel et al. 1988). This appears especially true for professions and skills that require adaptation based on human interaction factors, which have the lowest correlations between proficiency and years of experience after the first couple of years.[3] (For review see Ericsson et al. 1993.)

Measurements

Seven cognitive ability measurements and three emotion-cognition interaction measurements (specific personality dimensions) were administered to participants. (See Table 1 for a summary of the measurements.) The seven cognitive ability measurements were the Raven's Advanced Progressive Matrices (Raven 1962), Connections Tests (Salthouse et al. 2000), Letter Comparison (Salthouse & Babcock 1991), Pattern Comparison (Salthouse & Babcock 1991), Symmetry Span task (Unsworth et al. 2005) and a computerized version of the Wisconsin Card Sorting Test (Grant & Berg 1948). The emotion-cognition interaction measurements were the Behavioral Inhibition System/Behavioral Approach System (BIS/BAS) scales (Carver & White 1994) and the Need for Cognition scale (Cacioppo & Petty 1982). All measurements used in this study are psychometrically sound and are common measurements for their respective abilities/traits.

3. If the interpreter had less than two years of experience, the raters were asked to rate skill relative to experience.

Table 1. Measurements

Task	Ability/Trait Measured
	Cognitive abilities
Ravens	fluid intelligence; reasoning
Symmetry Span	working memory capacity
Letter Comparison	cognitive processing speed
Pattern Comparison	cognitive processing speed
Connections A	psychomotor speed
Connections B	psychomotor speed; cognitive control; task switching
WCST	mental flexibility
	Emotion-cognition interaction (personality) traits
Need for Cognition scale	willingness to engage in complex cognitive tasks
BAS scale	sensitivity to reward
BIS scale	sensitivity to risk

Note. Ravens = Ravens Advanced Progressive Matrices. WCST = Wisconsin Card Sorting Test. BAS = Behavioral Approach System. BIS = Behavioral Inhibition System.

Cognitive Ability Measurements

Raven's Advance Progressive Matrices

Raven's Advanced Progressive Matrices (referred to simply as "Ravens" in this paper) is a multiple-choice abstract reasoning test designed to measure general fluid intelligence and reasoning in above-average intelligence individuals. Problems consist of a 3 X 3 matrix where each element is a pattern that shares some features with adjacent elements. The lower right element is missing, and participants are asked to choose which of the available choices best completes the pattern. Participants had ten minutes to complete as many of the 18 increasingly difficult matrices as possible. (We used the odd problems only. Previous research has shown the odd-even split-half corrected reliability coefficient is .96 [Burke 1972: 253].) Individuals who score high on the Ravens are better able to educe meaning in abstract patterns, think clearly, and reason than those who score low on the test.

Symmetry Span

Symmetry Span is one of several "complex span tasks" that measure working memory capacity by measuring item recall in the face of interference. Complex span tasks are commonly used to measure working memory capacity. Participants are tasked with performing the processing components of the task as well as retaining as many of the memoranda as possible. In symmetry span, participants make judgments about the symmetry of abstract figures along the vertical axis that are

interleaved with the presentation of a colored square on a 4 X 4 grid. Participants are tasked with remembering the position of the colored squares. After the 3–7 symmetry decision and colored square presentation trials, participants recall the position of the 3–7 colored squares in sequence. Individuals who score high on symmetry span have higher working memory capacity than those who score low on the task.

Letter Comparison

The Letter Comparison task assesses cognitive processing speed by measuring perceptual and decision-making speed through comparison of orthographic patterns. Participants are presented with 8½ X 11-inch pages with pairs of non-lexical letter strings. Participants must compare the two items in the pair and mark whether they are the same or different. They complete as many pairs as possible in 30 seconds (per page). Two pages with different figures/letters are presented for each condition. Scores are calculated based on the number of correct decisions minus the number of uncorrected errors.

Pattern Comparison

The Pattern Comparison task is exactly the same as the Letter Comparison task except for the actual stimuli. Instead of non-lexical letter strings, Pattern Comparison uses abstract visual figures.

Connections Test A

Connections A measures psychomotor speed with two conditions: numbers and letters. Participants are presented with 8 ½ X 11-inch pages with numbers or letters in circles. Participants connect the numbers or letters in sequence with a pen. The numbers or letters are not presented on the page in sequence, but a sequential letter or number is always adjacent (in any direction) to the former letter or number. Participants are allowed 20 seconds per page. Scores are based on the quantity of accurate connections minus the number of uncorrected errors, and reflect psychomotor speed ability.

Connections B

Connections B measures cognitive control, task switching, and psychomotor speed. Participants are presented with materials similar to those in Connections A, but must connect numbers *and* letters in alternating sequence. Connections B has two conditions: numbers-letters in which the sequence begins with a number (i.e. 1 ⇒ A ⇒ 2 ⇒ B ⇒ 3 ⇒ C, etc.) and letters-numbers in which the sequence begins with a letter (i.e. A ⇒ 1 ⇒ B ⇒ 2 ⇒ C ⇒ 3, etc.). Scoring and time allowances for Connections A and B are the same. Individuals who score high on

Connections B exhibit more robust and faster task switching abilities, more efficacious cognitive control, and faster psychomotor speed than those who score low on Connections B.

Wisconsin Card Sorting Test

The Wisconsin Card Sorting Test (WCST) measures set-shifting and mental flexibility. Set-shifting is the ability to adapt to changes and to shift among different sets of rules. Participants view the images of four decks of cards face-up on the screen and are asked to sort new cards into one of the four piles based on either the color, the number, or the shape of the symbol(s) on the card. Participants are not informed of the correct sorting rule, but are informed after each decision whether the sort choice was correct or incorrect. They must then infer the sorting rule from the feedback provided. Without warning, the rule will be changed during this task and the participants must discard the old rule and infer a new one based on the relevant symbol dimension. Individuals who score high on the WCST (based on percentage of correct sorts) are more able to flexibly adapt to changing reinforcement than those who score low on the test.

Emotion-Cognition Interaction Measurements

BAS

The BAS scale measures reward sensitivity, drive, and fun-seeking traits. BAS reflects approach orientation, as when the goal is to move toward something desired. Individuals who score high on the BAS scale are more sensitive to positive rewards and personal enjoyment, and are more likely to actively pursue activities that yield such rewards than are those who score low on the scale. The BAS scale is an individual differences assessment. Participants respond on paper to a series of statements with a Likert-type scale, choosing whether each statement is "very true for me", "somewhat true for me", "somewhat false for me", or "very false for me."

BIS

The BIS scale measures risk-taking sensitivity and anxiety surrounding aversive stimuli and novelty. It is interleaved with the BAS scale.[4] Participants respond to BIS scale statements exactly as they respond to BAS scale statements.

Need for Cognition Scale

The Need for Cognition scale measures an individual's intrinsic motivation to engage in complex, cognitively demanding tasks. Participants respond on paper to a

4. BIS sensitivity and BAS sensitivity are orthogonal.

series of statements with a Likert-type scale, choosing whether in their view each statement is "completely true", "mostly true", "mostly false", or "completely false."

Procedure

Everyone was tested individually for approximately one hour and ten minutes and was paid for participation. Participants were administered the tasks in the following order: (1) BIS/BAS scales, (2) Need for Cognition scale, (3) Letter Comparison, (4) Pattern Comparison, (5) Connections Tests (A and B alternated, the standard procedure for administering the Connections Tests), (6) the Ravens, (7) Symmetry Span, and (8) the Wisconsin Card Sorting Test.

Results and discussion

Univariate analyses

Prior to statistical analysis, all measurement scores were examined for accuracy of data entry, missing values, and normality of distribution. Missing values occurred for two subjects for the questionnaires (BIS/BAS scales and the Need for Cognition scale) and one subject for the WCST (due to timing issues during data collection.) The values were not replaced. There were no univariate outliers and all distributions were normal. (See Table 2 for descriptive statistics.)

One-way analyses of variance (ANOVAs) were conducted for each of the measurements prior to multivariate analyses. The ANOVAs revealed that

Table 2. Descriptive statistics ($N = 29$)

Measurements	Less skilled ($n = 14$)		Highly skilled ($n = 15$)	
	M	*SD*	*M*	*SD*
BIS	23.15	3.76	20.46	2.21
BAS	13.61	1.53	13.46	1.84
Need for Cognition	66.54	11.50	69.46	8.81
Connections A	28.63	5.78	33.65	6.04
Connections B	14.02	5.53	18.09	3.36
Pattern Comparison	.62	.13	.72	.14
Letter Comparison	.51	.10	.56	.10
Ravens	8.00	2.77	8.67	3.09
Symmetry Span	.54	.18	.61	.10
WCST	68.85	12.12	76.92	9.06

Table 3. Cohen's *d* effect sizes

Measurement	Effect Size *d*	Relative size	% Standing	% of Non-overlap
Connections B	.93	> Large	82	51.6
BIS	.92	> Large	82 (LS)	51.6
Connections A	.88	> Large	82	51.6
WCST	.79	Large	79	47.4
Pattern Comparison	.77	Large	79	47.4
Letter Comparison	.52	Medium	69	33.0
Symmetry Span	.50	Medium	69	33.0
Need for Cognition	.30	> Small	62	21.3
Ravens	.24	Small	58	14.7
BAS	.09	< Small	54 (LS)	7.7

Note. "% Standing" indicates the average percentile standing of the highly skilled interpreters relative to the less skilled interpreters (highly skilled interpreters scoring higher than less skilled interpreters) unless otherwise noted as "(LS)", in which case the less skilled interpreters scored higher than the highly skilled interpreters and the "% Standing" then refers to the average percentile standing of the less skilled interpreters relative to the highly skilled interpreter. "Relative size" is based on common interpretations of the magnitude of the effect size. "Percent of non-overlap" refers to the percent of the distributions of the two groups that do not overlap and is a way to view how different the two groups are. Relative size, percentile standing, and percent of non-overlap are based on *d* rounded to the tenth decimal.

Connections A scores, $F(1, 27) = 5.170$, $p = .031$, $\eta^2 = .161$ and Connections B scores, $F(1, 27) = 5.844$, $p = .023$, $\eta^2 = .178$ were significantly different between the two groups with highly skilled interpreters scoring higher. ANOVAs also revealed that BIS scores, $F(1, 24) = 3.733$, $p = .065$, $\eta^2 = .135$; Pattern Comparison scores, $F(1, 27) = 3.642$, $p = .067$, $\eta^2 = .119$; and WCST scores, $F(1, 27) = 3.194$, $p = .085$, $\eta^2 = .106$ were marginally significant with highly skilled interpreters scoring higher for Pattern Comparison and WCST and scoring lower on the BIS scale. No other measurements were significant ($ps > .10$).

The results of the ANOVAs suggest that faster psychomotor speed (Connections A), stronger cognitive control and task switching (Connections B), increased willingness to take risks (less inhibition surrounding risk-taking) (BIS), faster cognitive processing speed (Pattern Comparison), and more mental flexibility (WCST), are important for differentiating highly skilled and less skilled interpreters. To further explore effect sizes of the ten measurements, Cohen's *d*s were calculated (see Table 3). Cohen's *d* is a ratio: it is the difference between two group means relative to the pooled standard deviation of the two groups. Thus, a Cohen's *d* of 1.0 means that one group is an entire standard deviation greater than the other.

The Cohen's *d* effect sizes corroborate the ANOVAs by demonstrating that the five predictors which the ANOVAs revealed as significant and marginally

significant have the largest effect sizes. While being partially redundant, exhibiting all the effect sizes regardless of an arbitrary $p < .05$ cut-off reveals the range of effects and can provide insight as to which measurements are the most appropriate to use for future studies and which ones may be the most fruitful in further analyses. Specifically, the five measurements with the largest effect sizes are likely to reveal significant effects in a future higher-powered study and in multivariate analyses of the current data, which we turn to next.

Multivariate analyses

ANOVAs reveal differences between groups along a single dimension without taking into account other factors. Analyzing complex data sets in this manner can cause researchers to miss effects from contributing factors working in tandem with the variables being parceled out for univariate analysis. Discriminant function analysis reveals the best linear *combination* of predictors that differentiate two groups. Discriminant function analysis is similar to binary logistic regression, but is more powerful and efficient (Tabachnick & Fidell 2007: 441) and provides more accurate classifications and hypothesis testing as long as statistical assumptions are held (Grimm & Yarnold 1995: 241).

Prior to analysis, the cases were analyzed for multivariate outliers. Mahalanobis distance indicated two (one from each skill-level group) multivariate outliers (z-scores > 3.0). The two outliers were removed. Cases with missing values were also removed when measurements with the missing values were entered into the analyses in keeping with the requirement for discriminant function analysis. With the removal of the two multivariate outliers as well as cases with missing values, and with no more predictors (number of measurements entered into the analysis) than cases in the smallest group, the statistical assumptions underlying discriminant function analysis were met, leaving 89 percent of the cases available for analysis.

Discriminant function analysis creates statistical models in order to predict group membership from a set of predictors (also known as classifiers). Various combinations of predictors were evaluated. A successful statistical model was revealed, Wilks' Lambda = .551, $\chi^2(5) = 11.616$, $p = .040$, correctly classified cases = 83.3% (see Table 4). As predicted, the measurements with the largest effect sizes created the significant discriminant function model: Connections A, BIS, Connections B, Pattern Comparison, and WCST.

Structure coefficients are the correlations of each classifier to the discriminant functions (similar to factor loadings in factor analysis). The structure coefficients are used to assign meaningful labels to the discriminant functions. The structure matrix revealed the following coefficients: Connections B = .659, Connections A = .506, WCST = .370, Pattern Comparison = .352, and BIS = –.193. Since the highly

Table 4. Discriminant Function Analysis classification results

Count			
	Predicted Group Membership		
Actual Group	**Highly skilled**	**Less skilled**	**Correct Classifications**
Highly skilled	10	3	10/13
Less skilled	1	10	10/11
Total			20/24
Percentage			
	Predicted Group Membership		
Actual Group	**Highly skilled**	**Less skilled**	**Correct Classifications**
Highly skilled	76.9	23.1	76.9
Less skilled	9.1	90.9	90.9
Total			83.3

skilled group served as the reference group, positive coefficients indicate the positive correlation of scores from the highly skilled group to the discriminant functions while negative coefficients indicate the negative correlation of scores from the highly skilled group to the discriminant functions. In other words, the higher an individual scored on Connections B, Connections A, WCST, and Pattern Comparison and the lower the score on BIS, the more the scores correlated to the discriminant functions and the likelier the model was to predict that the individual belonged in the highly skilled group. The following discriminant function labels were created from the result of the structure coefficient matrix: task switching ability (Connections B), psychomotor speed (Connections A), mental flexibility (WCST), cognitive processing speed (Pattern Comparison), and aversion to risk (BIS).

The standardized canonical discriminant function coefficients indicate the unique contribution of each classifier to the discriminant functions and are used to determine the relative importance of the classifiers in predicting group membership (similar to beta weights in multiple regression). Mental flexibility (.634) is the most important predictor relative to the other entered classifiers followed by cognitive processing speed (.612), aversion to risk (−.520), task switching ability (.513) and, substantially less important, psychomotor speed (.220). See Table 5 for a summary of the coefficients.

To further establish the validity of the classification, a permutation test was conducted. This examines the possibility that the discriminant function solution does not capture something fundamentally different between these groups, but is simply a brute-force mathematical solution for separating cases into groups,

Table 5. Structure Coefficients and Standardized Canonical Discriminant Function Coefficients

Measurement	SC	Label	Importance (SCDFC)
WCST	.370	Mental Flexibility	.634
Pattern Comparison	.352	Cognitive Processing Speed	.612
BIS	−.193	Sensitivity to Risk	−.520
Connections B	.659	Task Switching Ability	.513
Connections A	.506	Psychomotor Speed	.220

Note. SC = Structure Coefficient, Label = the structure coefficients' corresponding factor labels, SCDFC = Standardized Canonical Discriminant Function Coefficients. The SCDFCs indicate relative importance. Wilks' Lambda = .551, $\chi^2(5) = 11.616$, $p = .040$.

regardless of what those groups might be. Put differently, the permutation test assumes that discriminant function analysis will find a solution predicting group membership for any arbitrary groups. To carry out this test, group membership is randomly reassigned for all the cases and the discriminant function analysis re-conducted with the original predictors. If this produces a statistically significant classification, then the original result is undermined and the discriminant function analysis has not captured true empirical differences between these groups. However, if the discriminant function analysis is unable to produce a significant classification of the permuted data, then the original result does reflect true differences between these groups in the indicated abilities. Discriminant function analysis was performed using the same five predictors with the cases now randomly assigned to the less skilled group and the highly skilled group. The results were not significant, Wilks' Lambda = .744, $\chi^2(5) = 5.766$, $p = .330$, indicating that the original discriminant function analysis model was, in fact, tracking true and important differences between highly skilled interpreters and those of lower skill level.

Discussion

The results from the ANOVAs, effect sizes, and discriminant function analysis clearly and strongly suggest that highly skilled interpreters are more mentally flexible, have faster cognitive processing speed, are less anxious about risks, are faster and more accurate when task switching, and have faster psychomotor speed than less skilled interpreters, regardless of the fact that both groups have the same amount of professional experience. Additionally, results from the multivariate analysis indicate that mental flexibility and cognitive processing speed are the most important predictors closely followed by willingness to take risks and task switching ability and, to some extent, psychomotor speed. These five predictors

were all significant and relatively powerful individual predictors, and together were the only combination of predictors to significantly and successfully predict group membership. The convergence of all the analytical results supports the hypothesis that certain domain-general cognitive abilities and emotional-cognitive interaction traits are strongly related to simultaneous interpreting performance.

Additionally, two other measures, Letter Comparison and Symmetry Span, exhibited medium effect sizes, but were not statistically significant predictors, given the current sample size. These results, however, imply that variations of the tasks measuring the same or similar constructs may contribute significantly in a future study with higher power. (A variant of Letter Comparison, Pattern Comparison, already provided significant statistical results in the current study.)

The medium effect size demonstrated by Symmetry Span suggests that working memory capacity differences may be predictive of interpreter skill level, but further studies should explore various tools that better capture the processing and storage capacities needed during simultaneous interpreting. Oberauer (2004) defines three distinct types of working memory capacity measurements that load onto a single working memory factor. They are (1) tasks that measure storage of briefly presented material concurrent with interfering processing, (2) formation of new structures and relationships, and (3) some executive functioning, such as updating. Symmetry Span primarily taps the first factor. SI, on the other hand, engages complex and simultaneous storage and processing demands along with concurrent and continuous context-based restructuring. Therefore, future studies utilizing working memory measures of updating and restructuring ability are likely to prove more auspicious when assessing variation among interpreters or students of interpreting than the measurement used in the current study. Indeed, the current results identify several psychological constructs for which multiple measures have been developed. In future work with substantially larger sample sizes a factor-analytic or structural equation modeling approach would potentially provide much greater resolution on the issue of exactly which abilities and traits combine, and in what fashion, to predict SI performance. Furthermore, such techniques are naturally suited to address issues of causality, provided that the relevant longitudinal data are available.

General discussion and future directions

As depicted in Table 4, by using a combination of domain-general cognitive abilities and personality traits, the current study was able to correctly classify 76.9% of the highly skilled interpreters and 90.9% of the less skilled interpreters, for an overall cross-validated accuracy of 83.3%. Among the several constructs tested,

mental flexibility, cognitive processing speed, task switching ability, psychomotor speed, and aversion to risk appear to be important in differentiating interpreters with a high level of SI skill from those with a low level of SI skill.

Because this study was unbiased by years of experience or age and because the cognitive constructs measured are generally stable, the results beg the question as to whether these traits and abilities may be reliable predictors of future interpreting skill levels prior to experience or training. Specifically, our results suggest that, particularly in combination, high levels of mental flexibility, cognitive processing and psychomotor speed, task switching ability, a low level of risk sensitivity, and possibly working memory capacity increase the likelihood that one will be a highly skilled interpreter.

As mentioned above, other cognitive mechanisms are likely to be crucial during the learning process that may be less relevant to differences in levels of performance after extensive experience. For example, individuals with low fluid intelligence or little willingness to engage in complex cognitive tasks may be less likely to complete an interpreter training program and so were not included here. Fluid intelligence and willingness to engage in complex cognition may therefore be predictive constructs for successful interpreting during the learning phase. Measures predictive of learning *and* performance should therefore be administered in any future longitudinal study.

The constructs measured in this study are unlikely to be the only abilities and traits predictive of interpreting learning and skill. Further studies are needed to explore other aptitudes such as executive functions (e.g. planning, updating, and selection and inhibition of irrelevant stimuli); social interaction abilities (e.g. boundary balancing, attitude, and ethical reasoning); meta-cognitive abilities (e.g. performance monitoring, internal-state monitoring, audience monitoring, and speaker meaning, speaker goal-state, and speaker-state monitoring); and learning ability (e.g. second language learning aptitude, general knowledge learning aptitude [crystallized intelligence], and skill acquisition aptitude). This complex constellation of candidate predictors reinforces the appeal of a structural equation modeling (SEM) approach for both larger sample cross-sectional follow-up studies as well as longitudinal investigations. SEM techniques would reduce the dimensionality of the predictive problem to a few central psychological constructs that could potentially explain the trajectory of learning as well as 'ultimate' performance after experience.

The relatively high predictive success rate of the current analysis coupled with the general stability of the identified measures over time suggest that domain-general cognitive abilities may also be good predictors of interpreting skill before one has received interpreter training. Further research is needed to evaluate this possibility. However, pending the results of such a study, one can envision the

development of an aptitude test that includes the measurement of abilities and traits that takes advantage of the readily available methods of measuring the abilities and traits identified here. If there is an extension from the current post-training classification to pre-training identification of those individuals most likely to succeed as interpreters, then such an aptitude battery could potentially further enhance the predictive power of admission tests, increasing the likelihood that each accepted student later becomes a highly skilled interpreter.

References

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General* 117, 291–294.

Bayliss, D. M., Jarrold, C., Gunn, D. M. & Baddeley, A. D. (2003). The complexity of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General* 132, 71–92.

Burke, H. R. (1972). Raven's Progressive Matrices: Validity, reliability, and norms. *Journal of Psychology: Interdisciplinary and Applied* 82 (2), 253–257.

Cacioppo, J. T. & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology* 42 (1), 116–131.

Carver, C. S. & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology* 67, 319–333.

Chi, M. T. H., Glaser, R. & Farr, M. (Eds.) (1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.

Christoffels, I. K., de Groot, L. M. B. & Kroll, J. F. (2006). Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency. *Journal of Memory and Language* 54 (3), 324–345.

Christoffels, I. K., de Groot, A. M. B. & Waldorp, L. J. (2003). Basic skills in a complex task: A graphical model relating memory and lexical retrieval to simultaneous interpreting. *Bilingualism: Language and Cognition* 6 (3), 201–211.

Cokely, D. (1995). *Interpretation: A sociolinguistic model.* Burtonsville, MD: Linstok Press.

Colonomos, B. (1997). *Pedagogical model of the interpreting process.* College Park, MD: Bilingual Mediation Center.

Colonomos, B. (2008). *Revised pedagogical model of the interpreting process.* College Park, MD: Bilingual Mediation Center.

Cowan, N. (2000). Processing limits of selective attention and working memory: Potential implications for interpreting. *Interpreting* 5 (2), 117–146.

Daneman, M. & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior* 19 (4), 450–466.

Dean, R. K. & Pollard, Jr., R. Q. (2005). Consumers and service effectiveness in interpreting work: A practice profession perspective. In M. Marschark, R. Peterson & E. A. Winston (Eds.), *Sign language interpreting and interpreter education: Directions for research and practice.* New York: Oxford University Press, 259–282.

Engle, R. W., Carullo, J. J. & Collins, K. W. (1991). Individual differences in working memory for comprehension and following directions. *Journal of Educational Research* 84, 253–262.

Ericsson, K. A., Krampe, R. T. & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review* 100, 363–406.

Eysenck, M. W. (1979). Anxiety, learning and memory: A reconceptualization. *Journal of Research in Personality* 13, 363–385.

Eysenck, M. W. (1985). Anxiety and cognitive task performance. *Personality and Individual Differences* 6 (5), 579–586.

Frauenfelder, U. H. & Schriefers, H. (1997). A psycholinguistic perspective on simultaneous interpretation. *Interpreting* 2, 55–89.

Garrod, S. & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences* 8, 8–11.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gerver, D. (1976). Empirical studies of simultaneous interpretation: A review and a model. In R. W. Brislin (Ed.), *Translation: Applications and research*. New York: Gardner Press, 165–207.

Gerver, D., Longley, P. E., Long, J. & Lambert, S. (1989). Selection tests for trainee conference interpreters. *Meta* 34 (4), 724–735.

Grant, D. A. & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology* 38 (4), 404–411.

Gray, J. A. (1982). *The neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system*. New York: Oxford University Press.

Gray, J. A. (1987). *The psychology of fear and stress*. 2nd edn. Cambridge: Cambridge University Press.

Gray, J. A. (1990). Brain systems that mediate both emotion and cognition. *Cognition & Emotion* 4 (3), 269–288.

Grimm, L. G. & Yarnold, P. R. (1995). *Reading and understanding multivariate statistics.* Washington, DC: American Psychological Association.

Hoffman, R. R. (Ed.) (1992). *The psychology of expertise: Cognitive research and empirical AI*. New York: Springer.

Johnson-Laird, P. N. (1999). Deductive reasoning. *Annual Review of Psychology* 50, 109–135.

Just, M. A. & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review* 99, 122–149.

Kiewra, K. A. & Benton, S. L. (1988). The relationship between information processing ability and note-taking. *Contemporary Educational Psychology* 13, 3–44.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review* 95, 163–182.

Kintsch, W. & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review* 85, 363–394.

Köpke, B. & Nespoulous, J. (2006). Working memory performance in expert and novice interpreters. *Interpreting* 8 (1), 1–23.

Kyllonen, C. & Stephens, D. L. (1990). Cognitive abilities as the determinants of success in acquiring logic skills. *Learning and Individual Differences* 2, 129–160.

Leon, M. R. & Revelle, W. (1985). The effects of anxiety on analogical reasoning: A test of three theoretical models. *Journal of Personality and Social Psychology* 49, 1302–1315.

Lopes, L. (1987). Between hope and fear: The psychology of risk. In L. Berkowitz (Ed.), *Advances in experimental social psychology.* San Diego: Academic Press, 255–296.

López Gómez, M. J., Bajo Molina, T., Padilla Benítez, P. & Santiago de Torres, J. (2007). Predicting proficiency in signed language interpreting: A preliminary study. *Interpreting* 9 (1), 71–93.

Macnamara, B. (2009). Foundational cognitive aptitudes of sign language interpreters. *Journal of Interpretation* 2008–2009, 9–32.

MacWhinney, B. (1997). Simultaneous interpretation and the Competition Model. In J. H. Danks, G. M. Shreve, S. B. Fountain & M. K. McBeath (Eds.), *Cognitive processes in translation and interpreting.* Thousand Oaks, CA: Sage, 215–233.

McDaniel, M. A., Schmidt, F. L. & Hunter, J. E. (1988). Job experience correlates of job performance. *Journal of Applied Psychology* 73, 327–330.

Monsell, S. (2003). Task switching. *TRENDS in Cognitive Sciences* 7 (3), 134–140.

Moser, B. (1978). Simultaneous interpretation: A hypothetical model and its practical application. In D. Gerver & H. W. Sinaiko (Eds.), *Language interpretation and communication*. New York/London: Plenum Press, 353–368.

Oberauer, K. (2004). The measurement of working memory capacity. In O. Wilhelm & R. Engle (Eds.), *Handbook of understanding and measuring intelligence*. Thousand Oaks, CA: Sage Publications, 393–407.

Ormrod, J. E. & Cochran, K. F. (1988). Relationship of verbal ability and working memory to spelling achievement and learning to spell. *Reading Research and Instruction* 28, 33–43.

Ownby, R. L., Czaja, S. J., Loewenstein, D. & Rubert, M. (2008). Cognitive abilities that predict success in a computer-based training program. *Gerontologist* 48 (2), 170–180.

Padilla, P., Bajo, M. T., Cañas, J. J. & Padilla, F. (1995). Cognitive processes of memory in simultaneous interpretation. In J. Tommola (Ed.), *Topics in interpreting research*. Turku: University of Turku, Centre for Translation and Interpreting, 61–71.

Raven, J. C. (1962). *Advanced Progressive Matrices, Set II*. London: H. K. Lewis.

Ree, M. J. & Earles, J. A. (1992). Intelligence is the best predictor of job performance. *Current Directions in Psychological Science* 1 (3), 86–89.

Rudser, S. F. & Strong, M. (1986). An examination of some personal characteristics & abilities of sign language interpreters. *Sign Language Studies* 53, 315–331.

Rund, B. R. (1998). A review of longitudinal studies of cognitive function in schizophrenia patients. *Schizophrenia Bulletin* 24 (3), 425–435.

Salthouse, T. A. & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology* 27 (5), 763–776.

Salthouse, T. A., Toth, J., Daniels, K., Parks, C., Pak, R., Wolbrette, M. et al. (2000). Effects of aging on efficiency of task switching in a variant of the trail making test. *Neuropsychology* 14 (1), 102–111.

Schmader, T. & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology* 85 (3), 440–452.

Schneider, S. L. & Lopes, L. L. (1986). Reflection in preferences under risk: Who and when may suggest why. *Journal of Experimental Psychology: Human Perception and Performance* 12, 535–548.

Seal, B. C. (2004). Psychological testing of sign language interpreters. *Journal of Deaf Studies and Deaf Education* 9 (1), 39–52.

Shlesinger, M. (2003). Effects of presentation rate on working memory in simultaneous interpreting. *The Interpreters' Newsletter* 12, 37–49.

Simon, H. A. & Chase, W.G. (1973). Skill in chess. *American Scientist* 61, 394–403.

Starkey, P. (1992). The early development of numerical reasoning. *Cognition* 43 (2), 93–126.

Sternberg, R. J. (1982). Reasoning, problem-solving, and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence*. Cambridge: Cambridge University Press, 225–307.

Tabachnick, B. G. & Fidell, L. S. (2007). *Using multivariate statistics*. 5th edn. Boston: Allyn and Bacon.

Treisman, A. M. (1965). The effects of redundancy and familiarity on translating and repeating back a foreign and a native language. *British Journal of Psychology* 56 (4), 369–379.

Unsworth, N., Heitz, R. P., Schrock, J. C. & Engle, R.W. (2005). An automated version of the operation span task. *Behavior Research Methods* 37, 498–505.

Waters, G. S. & Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, and Computers* 35 (4), 550–564.

Wichertsa, J. M., Dolana, C. V., Hessena, D. J., Oostervelda, P., van Baalb, G. C. M., Boomsmab, D. I. & Spanc, M. M. (2004). Are intelligence tests measurement invariant over time? Investigating the nature of the Flynn effect. *Intelligence* 32, 509–537.

Wine, J. (1971). Test anxiety and direction of attention. *Psychological Bulletin* 76, 92–104.

Testing aptitude for interpreting

The predictive value of oral paraphrasing, with synonyms and coherence as assessment parameters*

Mariachiara Russo
University of Bologna at Forlì

This longitudinal study on the predictive value of 'simultaneous' Italian-to-Italian oral paraphrasing as an aptitude test for interpreting was conducted on 64 conference interpreting students at the University of Bologna (Forlì). All students completed their course with Italian as their 'A' language, having passed the entrance examination (which included a paraphrase test, recorded for evaluation) during the period 2004–2006. Using paraphrase as a pre-interpreting aptitude test in a smaller sample, Russo and Pippa (2004) found a significant correlation between course outcome measures (average interpreting exam mark and number of exam sessions needed to pass all interpreting exams) and two of the evaluation parameters for paraphrasing: 'synonymic substitution' (use of synonyms and equivalent expressions) and 'loss of coherence'. The present study examined whether, in a different setting, these two parameters and the students' actual admission test mark for paraphrase/recall testing correlated with the same outcome measures. Ability to use synonyms showed the highest validity in relation to the number of interpreting exam sessions, correctly classifying 48 out of 64 students (75%) as 'slow' (> 6 sessions: test sensitivity = 80%) or 'fast' (≤ 6 sessions: test specificity = 71%). Results thus indicate that an oral 'real time' paraphrasing test can help identify such prerequisites of effective interpreting as mental flexibility and expressive ability.

1. Introduction

Pre-admission aptitude testing for courses in interpreting, recommended as a best practice by the International Association of Conference Interpreters (AIIC 2010a),

* I wish to thank my husband for his assistance with statistics and the two anonymous referees for their suggestions.

was recognised as a need by leading authorities on interpreter training as far back as the 1960s (Keiser 1965a, 1965b). Far from being discriminatory, aptitude testing makes sense for a number of reasons: (i) the danger that unlimited enrolments will make the ratio of students to teaching staff unworkable; (ii) public funding cuts, not only limiting availability of essential equipment and resources but endangering the very existence of training courses and facilities; (iii) shrinking markets for qualified conference interpreters, both nationally and internationally.

There are numerous references in the literature to the 'ideal' interpreter's profile (AIIC 2010b; Carroll 1978; Gerver et al. 1984; Henderson 1987; Herbert 1952; Keiser 1965a, 1965b, 1978; Lambert 1992; Longley 1989; Paneth 1962; Seleskovitch 1978), and a variety of aptitude tests have been developed accordingly. Generally, there seems to be reasonable consensus about the abilities and personal qualities required of interpreters (Russo, in this volume) — e.g., command of the source and target languages, mental rapidity, broad general culture, good memory, expressive ability, physical and psychological stamina, and ability to work in a team. Recently, attention has also been paid to such personal traits as: learning styles, motivation and cognitive flexibility (Timarová & Salaets, in this volume); cognitive and motivational contributors (Shaw, in this volume); emotional stability (Bontempo & Napier, in this volume); linguistic self-confidence, motivation and language anxiety (Rosiers et al., in this volume); domain-general cognitive abilities (Macnamara et al., in this volume); anxiety and student-perceived self-efficacy (Zannirato 2013). Current approaches thus tend to be focused on the candidate's 'soft' skills — a trend which was reflected at the international conference entitled "Aptitude for Interpreting: Towards Reliable Admission Testing" (the first on this subject), held in Antwerp in 2009.

From the ongoing debate on admission testing for potential interpreters (see Russo, in this volume), two major schools of thought emerge. Some advocate the testing of interpreting-related skills and the capacity to learn (Angelelli 2007; Donovan 2003; Lederer 1975; Shaw, in this volume; Sofr 1976; Tryuk 2002), broadly defined by Seleskovitch and Lederer as 'teachability' (Keiser 1965a, 1965b). Other authors, by contrast, recommend assessment of already acquired skills which are particularly relevant to (or actually part of) interpreting — e.g., sight translation (Sunnari 2002) or 'short' consecutive, which is used at 14 out of 18 interpreting schools surveyed by Timarová and Ungoed-Thomas (2008).

Since 1988, my research on aptitude testing has been focused on two features of the 'teachability' usually identified with particular predisposition for interpreting: mental flexibility and expressive ability. The rationale for this interest is that essential prerequisites of effective interpreting include rapid assimilation of the incoming message, together with the linguistic sensitivity and flexibility needed to convey it fluently in the target language. Command of coherent and

cohesive linguistic output implies two forms of fluency, described by Gerver et al. (1989: 728): 'associational fluency', explained as "the ability to produce words which share a given area of meaning or some other common semantic property"; and 'expressional fluency', defined as "the ability to think rapidly of words, groups of words, or phrases as well as contrast with the production of single words by focusing on the compositional aspects of sentences and on the manipulation of syntactic constructions". As Gerver el al. (1989: 728) state: "Re-expression is basic to the interpreter's task." The text-processing abilities which this involves reflect what Carroll (1978) calls 'word fluency factors': he identifies these as 'ideational fluency' (a rapid and coherent flow of ideas on a topic), 'expressive fluency' (rephrasing a sentence in as many ways as possible) and 'associative fluency' (production of synonyms and antonyms).

There is thus good reason to think that admission test candidates able to assimilate and reproduce oral discourse under conditions as similar as possible to simultaneous interpreting will become successful interpreting students. This explains my interest in the use of 'real time' oral paraphrasing as a means of assessing aptitude for interpreting. It is not a particularly common choice of admission test (Timarová & Ungoed-Thomas 2008): out of 93 papers about interpreting aptitude published between 1950 and 2013 (for details see Russo, in this volume), nine examine the use of paraphrasing (see Longley 1989; Moser-Mercer 1985, 1994; Pippa & Russo 2002; Russo 1989, 1991, 1993, 1995; Russo & Pippa 2004).

Moser-Mercer (1985) administered a testing battery which included paraphrasing, but she did not assess its individual contribution to the success rate of her screening procedure.

As a component of the aptitude testing battery implemented at the Polytechnic of Central London (Gerver et al. 1989), Longley (1989) describes the use of paraphrasing (more precisely 'rewriting', as the speech was aurally presented and the candidate had to recast it in writing). In addition, she reports that students who passed their finals had achieved a higher score in prior assessment on a synonyms test than those who failed.

Against this background, an Italian-to-Italian 'real time' oral paraphrasing test was designed, piloted and described by Russo (1989); it was subsequently refined by Pippa and Russo (2002). The paraphrase test was systematically used on an experimental basis, between 1988 and 1994, at the Advanced School of Modern Languages for Interpreters and Translators (SSLMIT) of the University of Trieste. At that time, undergraduates enrolled for a four-year degree course, specialising in conference interpreting or in translation during the third and fourth years. The paraphrase test was used purely for internal evaluation of students beginning the two-year specialisation in interpreting. A longitudinal study focusing on the academic careers of these students was performed, involving 46 individuals (Russo

1993; Russo & Pippa 2004). The study examined to what extent the many different assessment parameters identified in the authors' paraphrase evaluation model correlated with two outcome measures: the number of exam sessions each individual student took to complete all the interpreting exams (students who failed exams could resit them, as is still the case), and the average marks obtained at interpreting exams. In the event, only two of the many parameters taken into account showed a significant correlation with the outcome measures: ability to use synonyms and equivalent expressions (the higher the score on this measure, the better), and loss of the target speech's logico-semantic coherence (the lower the score the better, the optimum being zero) (Russo & Pippa 2004).

Following on from the research described above, the present study of Italian-to-Italian paraphrasing as an aptitude test for interpreting was conducted at the University of Bologna at Forlì. The study examined whether the significant correlations identified in Russo and Pippa's 2004 study would again be present, this time in a larger sample of students.

2. Methods

2.1 Participants

Between 2004 and 2006, 114 candidates passed the entrance examination for admission to the 2-year M.A. in Conference Interpreting at what is now called the Department of Interpreting and Translation (DIT)[1] of the University of Bologna (Table 1). The entrance examination included a cloze test for initial screening, followed by more detailed assessment based on an Italian-to-Italian 'real time' oral paraphrase test and a recall test. These tests were recorded. Of the 114 successful candidates, 81 actually enrolled. All those who went on to complete the course received the same training, and were evaluated from entrance examination to final exams by the same teaching staff.

To ensure as homogenous a sample as possible, the study was limited to the 64 individuals who went on to graduate with a language combination of Italian 'A' and two 'B' languages ('ABB'). The remaining 17 students who enrolled were excluded, for the following reasons: two 'A' languages (2 students); 'A' language other than Italian (4 students); 'ABCC' language combination (6 students); studies abandoned before graduation (5 students).

1. Until 2012, the DIT was known as the Advanced School of Modern Languages for Interpreters and Translators, or SSLMIT.

Table 1. Entrance examination data for the University of Bologna's M.A. in Conference Interpreting, from 2004 to 2006

Academic Year	Number of candidates		Number of enrolments	
	Taking the exam	Passing	Total	Included in this study
2004–2005	96	38	28	22
2005–2006	196	38	25	22
2006–2007	225	38	28	20
All	517	114	81	64

Table 2. Study sample (64 individuals): breakdown by 'B' languages and gender

'B' languages	Interpreting students		
	All ($n = 64$)	Females ($n = 56$)	Males ($n = 8$)
English – French	21	16	5
English – Spanish	18	18	0
English – Russian	11	9	2
English – German	8	8	0
French – Russian	3	3	0
French – German	1	0	1
French – Spanish	1	1	0
Russian – German	1	1	0

The average age of the 64 individuals evaluated, at the time of starting the course, was 23 years ($SD = 1$; min. = 20, max. = 29). Details of gender and language combination are shown in Table 2.

2.2 Scoring

The Italian-to-Italian paraphrase component of the entrance examination required 'real time' oral reformulation of a recorded impromptu speech lasting approximately 3 minutes, delivered by a politician or journalist. Every year a different speech was submitted: a speech by a journalist taken from the radio in 2004 and 2005 (first topic: politics; second topic: terrorist attacks), and an extract from a talk on European political identity given by a politician at our School in 2006. Common features of the speech samples used were: redundant information, informal register and lack of technical lexicon. After about 15 minutes' practice of listening to a recording of a speech and speaking at the same time in order to convey its meaning, students were asked to recast the actual test speech in their own

words while listening to it for the first and only time. Their performances were recorded, for joint evaluation by an Italian language specialist and an interpreter: the former gave a holistic evaluation of content and style, while the latter assessed use of synonyms and loss of coherence through a simplified version of the evaluation grid proposed by Pippa and Russo (2002).

For the present study, the author made a thorough and independent evaluation of the candidates' recorded paraphrase test, with specific reference to synonyms and loss of coherence. She was not aware of the admission test scores actually given at the time of the tests, and applied the Pippa and Russo (2002) evaluation grid in its full original form for these two items. This evaluation of each candidate's performance took an average of about 20 minutes.

Use of synonyms and equivalent expressions was marked according to the following criteria: (i) substitution of a whole sentence = 2 points — e.g., "Mi scuso intanto di quelle che il mio vecchio professore di latino avrebbe chiamato 'le poche briciole' che io metto alla sua, alla vostra mensa e che lui, il mio professore diceva, avrebbe trasformato in qualche modo in un panino" (*I apologise for what my old Latin teacher would have called 'the few crumbs' which I put on his, on your table and which he, my teacher would say, would somehow turn into a sandwich.*) → "Mi scuso per il mio limitato contributo alla conferenza" (*I apologise for my limited contribution to the conference.*); (ii) substitution of a single lexical item with a different word = 1 point — e.g., "tema scottante" (*burning issue*) → "tema attuale" (*topical issue*); (iii) substitution of a single lexical item with one item derived from it = 0.5 points — e.g., "ammiro" (*I admire*) → "ammirazione" (*admiration*), or substitution of a verb tense = 0.5 points — e.g., "ho ammirato" (*I have admired*) → "ammirai" (*I admired*).

Loss of coherence was scored by assigning one point for each occurrence in the candidate's linguistic output. As explained above, this parameter is reverse-scored by comparison with synonyms: a high score on loss of coherence quantifies a *negative* feature of the paraphrasing test performance. The expectation is therefore that it will correlate negatively with good performance on the outcome measures: in other words, it is a *low* score on loss of coherence which should be predictive of success on the course.

2.3 Outcome measures

As stated in the introduction, the outcome measures were the average mark in first and second year interpreting exams and the number of sessions taken by each student to pass his/her first and second year interpreting exams.

The students enrolled in this study were required to pass four interpreting exams for each year, and for each of the two 'B' languages: (1) simultaneous

interpretation (SI) into Italian; (2) SI from Italian; (3) consecutive interpretation (CI) into Italian; (4) CI from Italian. Marks were expressed according to the Italian system: out of 30, with a pass mark of 18/30. In the present study "average mark in first and second year interpreting exams" (referred to from now on as "examination mark") is the arithmetic mean of the 16 marks of the interpreting exams in the two languages, and in the two years envisaged for the M.A. course.

Since a student who fails an exam can continue to follow courses and then resit, the number of exam sessions needed to complete the course can vary considerably. There are three exam sessions per academic year (i.e. 6 regular exam sessions to complete the M.A. course), the first of them just after the end of the year's courses. Thus, for the academic year 2004–2005, the first session ("session 1") was in June 2005, the second ("session 2") in September 2005, and the third ("session 3") in January 2006. In the present study the number of sessions (referred to from now on as "number of exam sessions") taken by each student to pass interpreting exams was calculated by adding the number of the session at which the student completed the first year exams to the number of the session at which s/he completed the second year exams (e.g., 2 + 1 = 3, indicating that the student completed the exams in sessions 2 and 1 of the first and second years respectively — September 2005 for the first year exams, and June 2006 for the second year exams).

2.4 Statistical analysis

The two outcome measures "examination mark" and "number of exam sessions" were correlated with the scores given, in the independent evaluation of the candidates' recorded paraphrase test, for the two assessment parameters "synonyms" and "loss of logico-semantic coherence". The correlation of the outcome measures with the actual admission test score for paraphrase and recall tests (referred to from now on as "admission test score") was also examined.

Correlations were first assessed by Kendall's rank correlation test (Kendall 1938), a distribution-free method (Armitage & Berry 1987) to study the strength of association between each of the two outcome measures (the response variables) and each of the three test scores (the explanatory variables). This involved calculating the Kendall's correlation coefficient τ along with the significance level, i.e. the *p*-value (probability value), obtained by testing the hypothesis that each of the two response variables (i.e. outcome measures: average exam mark; number of exam sessions) was independent of the three explanatory variables (i.e. synonyms score; loss of coherence score; admission test score). Kendall's rank correlation coefficient (τ) measures the extent to which, as one variable increases, the other variable tends to increase or decrease, without requiring that increase or decrease to be represented by a linear relationship. As a distribution-free method, it does

not rely on the assumption that data are drawn from a given probability distribution, while the linear correlation method (see next paragraph) assumes that both variables are normally distributed, i.e. they have a Gaussian distribution.

As to the second part of the statistical analysis, the linear correlation method (Edwards 1984) was used to study the strength of the linear association between each of the response variables and each of the explanatory variables. The product-moment correlation coefficient r was used to measure the linear association between the two variables concerned, the p-value again being obtained by testing the hypothesis that they were independent.

The third step was to investigate the linear trend in the association of the response variables with the explanatory variables, by the robust linear regression analysis (Rousseeuw & Leroy 1987): the robust approach was chosen in order to reduce the effect of occasional outlying observations. The prediction for each response variable, based on its linear regression against each explanatory variable, was calculated and the resulting lines were plotted. By identifying the regression coefficient b (the slope of the regression line) and its 95% confidence interval (CI), it was possible to obtain the average increase (or decrease) in the response variable corresponding to a unit increase (or decrease) in the explanatory variable. The p-value was obtained by testing the hypothesis that $b = 0$, i.e. that the average value of the response variable did not change with the explanatory variable.

To complete the statistical analysis, students were classified as "high" (≥ 24/30) or "low" mark (< 24/30) according to their average interpreting exam mark, 24/30 being the median value for the group of 64 interpreting students. Similarly, students were classified as "fast" (≤ 6 exam sessions to pass all interpreting exams) or "slow" (> 6 sessions).

For each of the explanatory variables, the cut-off values for high-mark vs. low-mark students, and fast vs. slow students, were then calculated. This was done by identifying the mean scores of each test for high-mark vs. low-mark students, and for fast vs. slow students, after which the following formula was applied:

$$\text{Cut-off value}_{\text{Examination mark}} = [(\text{Mean score}_{\text{high mark students}}) + (\text{Mean score}_{\text{low mark students}})] : 2$$

and

$$\text{Cut-off value}_{\text{Number of exam sessions}} = [(\text{Mean score}_{\text{fast students}}) + (\text{Mean score}_{\text{slow students}})] : 2$$

The two cut-off values of each test score were used to determine the number of students correctly allocated to the groups of high-mark and low-mark students and the groups of fast and slow students.

The 95% CIs of the proportions of the correctly allocated low-mark or slow students (sensitivity of the test) and of the correctly allocated high-mark or fast students (specificity of the test) were computed according to the score method, incorporating continuity correction as described by Newcombe (1998).

Statistical analysis was performed with the Stata/SE 8.0 (StataCorp 2003) package.

3. Results

3.1 Scores on all variables

The mean scores obtained for the 64 enrolled students in synonyms, loss of coherence and the admission test are shown in Table 3.

Table 3. Mean values with standard deviation (*SD*), and minimum and maximum values for test scores, exam mark and number of exam sessions needed to pass all exams

Variable	Mean (*SD*)	Min. – Max.
Scores for:		
Synonyms/equivalent expression	30 (11)	9–56
Loss of coherence	2 (2)	0–8
Admission test (paraphrase/recall)	15 (2)	8–19
Exam mark	24 (2)	20–28
Number of exam sessions	6 (2)	2–12

3.2 Correlation and regression analysis

3.2.1 *Outcome measure: Examination mark*

As shown in Table 4, Kendall's rank correlation test indicated that the average exam mark correlated significantly with use of synonyms and admission test scores ($p = 0.006$). The linear correlation analysis and robust linear regression analysis were consistent with this finding (linear correlation, $p = 0.04$ for synonyms, 0.02 for admission test score; robust linear regression, $p = 0.004$ for admission test score, $p = 0.02$ for use of synonyms). On the other hand, there was no statistically significant correlation for exam mark vs. loss of coherence.

3.2.2 *Outcome measure: Number of exam sessions*

For this outcome measure, Kendall's rank correlation test (Table 4) showed the strongest correlation with use of synonyms ($p < 0.0001$) and the admission test

Table 4. Correlation between outcome measures and explanatory variables, measured by Kendall's correlation coefficient (τ), product-moment correlation coefficient (r) and regression coefficient (b)

Explan. variable	Outcome measure													
	Examination mark							Number of exam sessions						
	τ	p	r	p	b	95% CI	p	τ	p	r	p	b	95% CI	p
Synonyms/ equivalents	0.2	0.006	0.3	0.04	0.05	0.01; 0.09	0.02	−0.3	<0.0001	−0.4	0.0005	−0.08	−0.1; −0.04	<0.0001
Loss of coherence	−0.06	0.5	−0.03	0.8	−0.05	−0.4; 0.3	0.7	0.1	0.2	0.06	0.6	0.2	−0.1; 0.5	0.1
Admission test	0.2	0.006	0.3	0.02	0.3	0.1; 0.5	0.004	−0.2	0.02	−0.2	0.06	−0.3	−0.5; −0.1	0.02

score ($p = 0.02$). In the linear correlation test, the strongest correlation was again with use of synonyms ($p = 0.0005$). The robust linear regression analysis indicated significant correlations with use of synonyms ($p < 0.0001$) and admission test score ($p = 0.02$). As for the previous outcome measure, there was no statistically significant correlation with loss of coherence.

3.3 Sensitivity and specificity of tests

3.3.1 *Low-mark and high-mark students*

Tables 5 and 6 show that assessment of synonyms and equivalents correctly allocated 66% (95% CI = 54–77%) of all students to the low-mark and high-mark groups. Sensitivity for this assessment was 73% (95% CI = 52–88%), this being the percentage of low-mark students correctly identified as a result; specificity was 61% (95% CI = 43–76%), this being the percentage of high-mark students correctly identified. The 'loss of coherence' parameter correctly allocated 56% (95% CI = 44–68%) of

Table 5. Allocation of students to low- and high-mark groups, by cut-off values

Actual student group	Group to which students are allocated, by cut-off value for each explanatory variable					
	Synonyms/equivalents (cut-off value = 29.8)		Loss of coherence (cut-off value = 2.1)		Admission exam score (cut-off value = 14.5)	
	Low-mark	High-mark	Low-mark	High-mark	Low-mark	High-mark
Low-mark ($n = 26$)	19	7	10	16	17	9
High-mark ($n = 38$)	15	23	12	26	14	24

Table 6. Number and proportion of low- and high-mark students correctly allocated by cut-off value for each explanatory variable

Explanatory variable	Correctly allocated students			
		Low-mark ($n = 26$)	High-mark ($n = 38$)	All ($n = 64$)
Synonyms/equivalents	Number	19	23	42
	%	*73*	*61*	*66*
Loss of coherence	Number	10	26	36
	%	*38*	*68*	*56*
Admission	Number	17	24	41
	%	*65*	*63*	*64*

all students to the low- and high-mark groups, with 38% sensitivity (95% CI = 21–59%) and 68% specificity (95% CI = 51–82%). Finally, the admission test score for paraphrase/recall correctly allocated 64% (95% CI = 52–76%) of all students, with 65% sensitivity (95% CI = 44–82%) and 63% specificity (95% CI = 46–78%).

3.3.2 *Slow and fast students*

Assessment of synonyms and equivalents correctly allocated 75% (95% CI = 64–86%) of all students to the groups of slow and fast students (Tables 7 and 8). Sensitivity was 80% (95% CI = 61–92%), this being the percentage of slow students correctly identified; specificity was 71% (95% CI = 52–84%), i.e. the percentage of fast students correctly allocated. For loss of coherence, the percentage of correctly allocated students was 59% (95% CI = 47–71%), with a sensitivity of 43% (95% CI = 26–62%) and a specificity of 74% (95% CI = 55–86%). Finally, the admission test score correctly allocated 55% (95% CI = 42–67%) of all students: sensitivity was 53% (95% CI = 35–71%), while specificity was 56% (95% CI = 38–72%).

Table 7. Allocation of slow and fast students, by cut-off values

Actual student group	Group to which students are allocated, by cut-off value for each explanatory variable					
	Synonyms/equivalents (cut-off value = 29.9)		Loss of coherence (cut-off value = 2.1)		Admission exam score (cut-off value = 14.6)	
	Slow	Fast	Slow	Fast	Slow	Fast
Slow ($n = 30$)	24	6	13	17	16	14
Fast ($n = 34$)	10	24	9	25	15	19

Table 8. Number and proportion of slow and fast students correctly allocated by cut-off value for each explanatory variable

Explanatory variable	Correctly allocated students			
		Slow (*n* = 30)	**Fast** (*n* = 34)	**All** (*n* = 64)
Synonyms/equivalents	Number	24	24	48
	%	*80*	*71*	*75*
Loss of coherence	Number	13	25	38
	%	*43*	*74*	*59*
Admission	Number	16	19	35
	%	*53*	*56*	*55*

4. Discussion and conclusions

The results for loss of coherence did not show the same trend as in the earlier sample (Russo & Pippa 2004): in the present study, there was no statistically significant correlation between the outcome measures and the score for this parameter.

By contrast, results for assessment of synonyms/equivalents in the present study were consistent with those reported by Russo and Pippa (2004): a significant correlation was identified with both outcome measures. Of the three explanatory variables considered (synonyms/equivalents, loss of coherence, admission test score), use of synonyms and equivalents was the one which most clearly distinguished high- vs. low-mark students and, even more so, fast vs. slow students.

This result is of great interest, for several reasons. First, it is consistent with Carroll's definition of aptitude (1961, quoted in Bowen & Bowen 1989: 110): "The learner's aptitude, defined as an inverse function of the amount of time which, other things being optimal, will be required for him [the learner] to attain a criterion mastery in the task to be learned." Second, calculating the amount of time needed to complete exams is a more objective measure than marks at interpreting exams, given the lack of a standardized marking procedure across different languages and evaluators (see also Donovan 2003). Third, taking six sessions as the cut-off value between fast and slow students reflects the actual length of the two-year M.A. in Italian universities and is crucial for both the student and the institution: for the student, because taking longer to graduate means s/he will have to pay tuition fees for more than the statutory two years; and for the institution, because students' failure to graduate within the normal time affects the Department's academic rating.

Our study gives further support to the finding, reported in previous studies (Gerver et al. 1984, 1989; Pöchhacker, in this volume; Russo & Pippa 2004), that

use of synonyms and equivalent expressions indicates high verbal fluency and is a strong predictor of success in completing interpreter training. Gerver et al. (1984, 1989) found that, of the subskills and speed stress tests in their battery, only the synonyms test significantly reflected differences between candidates who passed and those who failed their examinations.

Recently, Pöchhacker (in this volume) used his SynCloze assessment (a combination of cloze testing and use of synonyms) on 116 students attending an introductory course in interpreting. His aim was to investigate interpreting aptitude by assessing production of synonyms in end-of-sentence position. For institutional reasons, he was not able to correlate test scores with marks for consecutive or simultaneous interpreting exams proper, but he did assess correlation with performance on the introductory course exam, a short consecutive from German into German. The correlation was strong in 24 students, and moderate in 56.

Synonym production generally reflects different cognitive and linguistic skills at individual level, related to mental flexibility and expressive ability. The interpreting students who passed their exams quickly in the present study had initially proved better than the slower students at producing synonyms and equivalents under time pressure, which simultaneously testifies to their ability to rapidly grasp the meaning, to their associational fluency (which entails a search of a major proportion of long-term memory, according to Carroll 1976 quoted in Gerver 1989: 728) and to their expressional fluency.

As to expressive ability, this is undeniably considered a fundamental interpreting skill and professional interpreters usually have excellent command of active vocabulary. Thus, Straniero Sergio (2007) documents constant enhancements and multiple solutions in the interpreter's target language expression, while Russo et al. (2006) observe greater lexical variety in the production of Italian simultaneous interpreters at the European Parliament than in the speeches of many Italian delegates addressing the Parliament. In a study by Setton and Motta (2007), in which users very familiar with conference interpreting were asked to score experienced and novice interpreters' performances, lexical quality was identified as a quality feature of experienced practitioners. In particular, sophisticated and contextualised lexical choice was the factor that most positively correlated with users' overall quality ratings (ibid: 217).

However, it is still debatable which are the most suitable procedures to elicit this type of linguistic output from candidates for admission to interpreting courses. There is an ongoing debate about whether a simultaneous intralinguistic interpretation (paraphrasing) can justifiably be compared to a simultaneous interlinguistic interpretation (SI proper). Those who think so consider paraphrasing a natural process (Anderson 1994), which may even occur in tired conference interpreters who start interpreting into the *source* language (Christoffels & de Groot

2004; Moser 1997). In addition, the same cognitive skills are involved (Danks et al. 1997). Those who disagree maintain that paraphrasing is more difficult than SI, because the interpreter is prevented from using the same words to convey a message that is already perfectly expressed in the source text, and therefore actively inhibits his/her linguistic production (Christoffels & de Groot 2004; Malakoff & Hakuta 1991). Whether paraphrasing is really more difficult than SI may depend on the source text, as stated by Christoffels and de Groot (2004: 236, note 5):

> Note that with certain manipulations of stimulus material, it may be possible to reverse the relative performance in the interpreting and paraphrasing tasks. Some types of material may be easier to interpret whereas others may be easier to paraphrase. For example, sentences with low frequency words that have high frequency synonyms may be relatively easy to paraphrase. The same sentences may be relatively difficult to translate because low frequency target words must be found.

In the present study, the text type and the impromptu mode of delivery in the source texts for the paraphrasing test arguably limited the difficulty of paraphrasing. There are two reasons for not considering the task particularly difficult in this context: (i) the source texts' semantic redundancy; (ii) the original speakers' tendency to use lexical items and expressions for which high-frequency equivalents could generally be found.

There is a further point to highlight in the debate on paraphrasing. Though it is not considered to be a component of simultaneous interpreting (Pöchhacker, in this volume), Straniero Sergio's (2007) unique *Corit* corpus of media interpreting contains many instances of interpreters continuing to reformulate the same concept: they produce what the author calls 'coppie sinonimiche' (synonymic pairs), or even full alternative sentences, which do not fall under the categories of self-repairs (Petite 2003, 2005) or false starts (Bendazzoli et al. 2011). This prompts the following remark from Straniero Sergio (2007: 473): "Tramite il ricorso a sinonimi, antonimi, iponimi e iperonimi, l'interprete rielabora l'elemento lessicale antecedente, formando vere e proprie catene parafrastiche" ("By using synonyms, antonyms, hyponyms and hyperonyms, the interpreter re-elaborates the previous lexical item, thus forming authentic paraphrastic chains." — my translation).

In conclusion, developing aptitude tests for interpreting is an important ethical concern and an economic necessity for training institutions. Unfortunately, various factors hamper test validation procedures in our field. First, legislative, institutional, and organisational differences among interpreter training programmes make it hard to compare and transfer successful aptitude testing procedures. Second, there are difficulties in developing objective evaluations to make methodologies replicable and prevent subjective bias. Third, it is more difficult than in other disciplines relying on tests, such as psychology or sociology, to enrol large

and homogeneous samples of candidates so as to have a sounder basis for practical assessment of aptitude parameters. Last but not least, interpreter training institutions may encounter difficulties in keeping track of candidates' performances and in correlating results. This may be for a variety of reasons: changes in syllabus, turnover in teaching staff, and lack of methodological continuity among educator-researchers.

Despite possible limitations such as sample size, the present study sheds light on the value of two interpreting-related skills, assessed under conditions as similar as possible to SI, as selection criteria for potential trainee interpreters: mental flexibility and expressive ability.

References

AIIC (2010a). http://www.aiic.net/ViewPage.cfm/page60.htm (accessed 13 April 2013).

AIIC (2010b). Advice to students wishing to be conference interpreters. http://aiic.net/page/56 (accessed 13 April 2013).

Anderson, L. (1994). Simultaneous interpretation: Contextual and translation aspects. In S. Lambert & B. Moser (Eds.), *Bridging the gap: Empirical research in simultaneous interpretation*. Amsterdam: John Benjamins, 101–120.

Angelelli, C. V. (2007). Assessing medical interpreters: The language and interpreting testing project. *The Translator* 13 (1), 63–82.

Armitage, P. & Berry, G. (1987). *Statistical methods in medical research*. 2nd ed. Oxford: Blackwell Scientific Publications.

Bendazzoli, C., Sandrelli, A. & Russo, M. (2011). Disfluencies in simultaneous interpreting: A corpus-based analysis. In A. Kruger, K. Wallmach & J. Munday (Eds.), *Corpus-based translation studies: Research and applications*. London/New York: Continuum, 282–306.

Bowen, D. & Bowen, M. (1989). Aptitude for interpreting. In L. Gran & J. Dodds (Eds.), *The theoretical and practical aspects of teaching conference interpretation*. Udine: Campanotto Editore, 109–125.

Carroll, J. B. (1976). Psychometric tests as cognitive tasks: A new 'structure of intellect'. In L. B. Resnick (Ed.), *The nature of intelligence*. New York: Erlbaum, 27–56. DOI: 10.1007/978-1-4615-9077-4_12

Carroll, J. B. (1978). Linguistic abilities in translators and interpreters. In D. Gerver & H. W. Sinaiko (Eds.), *Language interpretation and communication*. New York: Plenum Press, 119–130.

Christoffels, I. K & de Groot, A. M. B. (2004). Components of simultaneous interpreting: Comparing interpreting with shadowing and paraphrasing. *Bilingualism: Language and Cognition* 7 (3), 227–240. DOI: 10.1017/S1366728904001609

Danks, J., Shreve, G. M., Fountain, S. B. & McBeath, M. K. (Eds.) (1997). *Cognitive processes in translation and interpreting*. Thousand Oaks, CA: Sage Publications.

Donovan, C. (2003). Entrance exam testing for conference interpretation courses: How important is it? *Forum* 1 (2), 17–45.

Edwards, A. L. (1984). *An introduction to linear regression and correlation*. 2nd ed. New York: Freeman.

Gerver, D., Longley, P., Long, J. & Lambert, S. (1984). Selecting trainee conference interpreters. *Journal of Occupational Psychology* 57, 17–31. DOI: 10.1111/j.2044-8325.1984.tb00144.x

Gerver, D., Longley, P., Long, J. & Lambert, S. (1989). Selection tests for trainee conference interpreters. *Meta* 34 (4), 724–735. DOI: 10.7202/002884ar

Henderson, J. (1987). *Personality and the linguist*. Bradford, UK: Bradford University Press.

Herbert, J. (1952) *Manuel de l'interprète. Comment on devient interprète de conférences*. Genève: Georg.

Keiser, W. (1965a). Admission dans les Ecoles d'interprétation. In AIIC, *Colloque sur l'enseignement de l'interprétation*, Paris 18–19 décembre 1965, 3–8.

Keiser, W. (1965b). Définitions d'aptitude. In AIIC, *Colloque sur l'enseignement de l'interprétation*, Paris 18–19 décembre 1965, 28–37.

Keiser, W. (1978). Selection and training of conference interpreters. In D. Gerver & H. W. Sinaiko (Eds.), *Language interpretation and communication*. New York: Plenum Press, 11–24. DOI: 10.1007/978-1-4615-9077-4_3

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika* 30, 81–93.

Lambert, S. (1992). Aptitude testing for simultaneous interpretation at the University of Ottawa. *The Interpreters' Newsletter* 4, 25–32.

Lederer, M. (1975). Le colloque sur les examens d'entrée aux Ecoles d'Interprètes. *AIIC Bulletin* 3 (2), 25–28.

Longley, P. (1989). The use of aptitude testing in the selection of students for conference interpretation training. In L. Gran & J. Dodds (Eds.), *The theoretical and practical aspects of teaching conference interpretation*. Udine: Campanotto Editore, 105–108.

Malakoff, M. & Hakuta, K. (1991). Translation skills and metalinguistic awareness in bilinguals. In E. Bialystok (Ed.), *Language processing in bilingual children*. New York: Oxford University Press, 141–166. DOI: 10.1017/CBO9780511620652.009

Moser-Mercer, B. (1985). Screening potential interpreters. *Meta* 30 (1), 97–100. DOI: 10.7202/003631ar

Moser-Mercer, B. (1994). Aptitude testing for conference interpreting: Why, when and how. In S. Lambert & B. Moser-Mercer (Eds.), *Bridging the gap: Empirical research in simultaneous interpretation*. Amsterdam: John Benjamins, 57–68.

Moser-Mercer, B. (1997). Methodological issues in interpreting research: An introduction to the Ascona workshops. *Interpreting* 2 (1/2), 1–11. DOI: 10.1075/intp.2.1-2.01mos

Newcombe, R. G. (1998).Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Statistics in Medicine* 17 (8), 857–872. DOI: 10.1002/(SICI)1097-0258(19980430)17:8%3c857::AID-SIM777%3e3.0.CO;2-E

Paneth, E. (1962). The interpreter's task and training. *The Incorporated Linguist* 1 (4), 102–109.

Petite, C. (2003). Repairs in simultaneous interpreting: Quality improvement or simple error correction. In A. Collados Aís, M. M. Fernández Sánchez & D. Gile (Eds.), *La evaluación de la calidad en interpretación: Investigación*. Granada: Comares, 61–71.

Petite, C. (2005). Evidence of repair mechanisms in simultaneous interpreting: A corpus-based analysis. *Interpreting* 7 (1), 27–49. DOI: 10.1075/intp.7.1.03pet

Pippa, S. & Russo, M. (2002). Aptitude for conference interpreting: A proposal for a testing methodology based on paraphrase. In G. Garzone & M. Viezzi (Eds.), *Interpreting in the 21st century: Challenges and opportunities*. Amsterdam: John Benjamins, 247–258.

Rousseeuw, P. J. & Leroy, A. M. (1987). *Robust regression and outlier detection*. New York: John Wiley and Sons. DOI: 10.1002/0471725382

Russo, M. (1989). Text processing strategies: A hypothesis to assess students' aptitude for simultaneous interpreting. *The Interpreters' Newsletter* 2, 57–64.

Russo, M. (1991). Ipotesi valutativa dell'attitudine all'interpretazione simultanea: risultati preliminari. *Zeta* 12 (14/15/16), 388–399.

Russo, M. (1993). Testing aptitude for simultaneous interpretation: Evaluation of the first trial and preliminary results. *The Interpreters' Newsletter* 5, 68–71.

Russo, M. (1995). L'interpretazione simultanea ed i test attitudinali. *Il Traduttore Nuovo* 44 (1), 37–46.

Russo, M., Bendazzoli, C. & Sandrelli, A. (2006). Looking for lexical patterns in a trilingual corpus of source and interpreted speeches: Extended analysis of EPIC (European Parliament Interpreting Corpus), *Forum* 4 (1), 221–254.

Russo, M. & Pippa, S. (2004). Aptitude to interpreting: Preliminary results of a testing methodology based on paraphrase. *Meta* 49 (2), 409–432. DOI: 10.7202/009367ar

Seleskovitch, D. (1978) *Interpreting for international conferences*. Washington, DC: Pen and Booth.

Setton, R. & Motta, M. (2007). Syntacrobatics: Quality and reformulation in simultaneous-with-text. *Interpreting* 9 (2), 199–230. DOI: 10.1075/intp.9.2.04set

Sofr, O. (1976). Test of aptitude for interpreting. *L'interprète* 31 (3/4), 3–8.

StataCorp. (2003). *Stata statistical software: Release 8.0*. College Station, TX: Stata Corporation.

Straniero Sergio, F. (2007). *Talkshow interpreting. La mediazione linguistica nella conversazione spettacolo*. Trieste: EUT.

Sunnari, M. (2002). Aptitude tests and selection criteria for interpreting students. In *Teaching simultaneous interpretation into a "B" language: Proceedings of EMCI Workshop*, Paris, 23–26. Also available at: http://www.emcinterpreting.org/repository/pdf/EMCI-TeachingSimultaneousIntoB-vol1.pdf (accessed 26 April 2009).

Timarová, Š. & Ungoed-Thomas, H. (2008). Admission testing for interpreting courses. *The Interpreter and Translator Trainer* 2 (1), 29–46.

Tryuk, M. (2002). Les tests d'aptitude et leur rôle dans la formation des interprètes de conférence. In B. Lewandowska-Tomaszczyk & M. Thelen (Eds.), *Translation and meaning*, Part 6. Maastricht: UPM, 421–434.

Zannirato, A. (2013). Multidisciplinary reflections on aptitude and affect in interpreter selection and training. *The Interpreter and Translator Trainer* 7 (1), 107–127.

Assessing aptitude for interpreting

The SynCloze test*

Franz Pöchhacker
University of Vienna

Based on a review of some of the most promising approaches to aptitude testing in the literature this paper proposes a novel task piloted at the Center for Translation Studies of the University of Vienna. The SynCloze test combines an auditory cloze exercise with a task requiring high expressional fluency, that is, rapidly finding contextually appropriate synonymic sentence completions. The rationale and design of the SynCloze test as well as the scoring method, which takes into account both the degree of accuracy and the speed of response, are described. The results of four rounds of testing involving some 120 students in the final stage of their undergraduate studies show that the test effectively discriminates between undergraduate novices and a control group of interpreting students, and students for whom the test language (German) is the A vs. the B language. Most significantly, the test scores correlate, albeit moderately, with students' performance on an intralingual consecutive interpreting exam at the end of the course.

1. Introduction

The issue of admission testing in interpreter education was raised already in the preface to one of the most widely read early publications on conference interpreting — the 1968 book by Danica Seleskovitch:

> It is probably not necessary to train large numbers of interpreters. However, we should seek out those young people who are most likely to succeed at this very difficult, worthy and socially important profession. (Gravier 1968/1978: vii)

* I am grateful to two anonymous reviewers for valuable comments on an earlier draft of this paper. Most importantly, I wish to acknowledge my debt of gratitude to Dominika Tondera for her diligence and enthusiasm in assisting me with this project.

While the demand for interpreters may of course vary from one market to another, the basic objective remains the same — that is, seeking out those most likely to succeed.

Determining a person's aptitude for interpreting means confronting two major challenges: the complex nature of aptitude, and the difficulties of implementing a test under the legal and organizational constraints of a given educational context.

1.1 Aptitude

As is clear from even a standard dictionary definition of aptitude — as "any constellation of measurable characteristics known to predispose to the learning of certain skills" (Webster's 1986: 108) — assessing aptitude for interpreting requires an understanding of the traits and abilities involved and of their relative weight in shaping a trainee's probability of success.

The awareness that interpreting requires a complex set of abilities and skills can be traced back to the very first scientific study among professional interpreters. Based on interviews with some twenty conference interpreters in Geneva, Spanish educator Jesús Sanz (1930) listed a dozen requirements, including cognitive abilities, such as intelligence, intuition and memory, and moral and affective qualities, such as tact, discretion and poise.

In subsequent decades, as conference interpreters came to do most of their work in the booth, the balance between cognitive and affective characteristics apparently shifted towards the former. Keiser, for instance, emphasizes knowledge (of languages and of the world) and personal qualities, including analytical, memory and public speaking skills as well as concentration and good "nerves" (1978: 17). A focus on cognitive skills is also evident from the influential study by Gerver et al. (1989) among conference interpreting trainees in London. While these authors were clearly aware of the multifactorial nature of aptitude for interpreting, only one out of the five success factors in their list — teamwork ability — relates to the interpersonal dimension. The others — knowledge of languages and cultures; ability to quickly grasp and convey meaning; speaking skills; and wide general knowledge and intellectual curiosity — are essentially cognitive and are evidently geared towards (simultaneous) conference interpreting as the skill to be acquired.

As we have come to broaden our notion of interpreting — and our training goals — the role of affective characteristics has acquired new relevance, and relatively little is known as yet about the (inter)personal qualities needed to cope with the conflictual or otherwise emotionally taxing situations that are likely to arise in face-to-face dialogue in institutional settings.

The extent to which different occupational profiles need to be taken into account in assessing aptitude for interpreting will of course depend on the nature of

the program, which in turn is shaped by the overall educational and professional context. In the case of the MA in Interpreting at the University of Vienna, for instance, students can choose between two different specializations, conference and dialogue interpreting, and therefore would need to be screened for aptitude in either.

1.2 Context

As stated clearly by Bowen & Bowen (1989: 109), aptitude testing "must be seen as part of a whole educational system." A multitude of factors come into play, from secondary-school-leaving exams to (changing) legal provisions for higher education. Rather than attempt an analytical review, I would like to give a brief sketch of the particular educational and institutional context in which my aptitude testing efforts are set — both to give an illustration of possible contextual constraints and to explain the options available in our particular curriculum.

On the overarching European level, the major constraint is of course the Bologna Process, with its three-tiered structure. Our previous curriculum for interpreter education, which lasted as long as 30 years, was a single-tier MA model, with a first four-semester cycle (focused on language enhancement, LSP and translation), after which students could opt freely to specialize in either translation or interpreting and earn an MA degree after a total of four years of study. Drop-out rates over the entire four-year course were as high as 80 to 90 percent. In other words, selection took place during the degree course itself.

One of the key principles in Austria's essentially public higher education system is free access for anyone with a secondary school certificate. With few exceptions (e.g. for medical school), there is no selective admission. However, given the obvious problems with this liberal approach, the trend is now for screening during the first phase of the BA curriculum, so that failure to pass a defined set of core courses in the first one or two semesters will bar students from going further. This probation principle is also applied in our current 3 + 2 BA/MA curriculum, which includes an eliminatory language test as part of an introductory course for all first-term BA students.

In the new policy context, drop-outs are no longer considered a hallmark of high standards but a sign of institutional failure, both in our three-year BA degree and in the two-year MA programs. The latter are open to anyone with a BA degree, which puts a great strain on teaching resources. Our only advantage is that most students enter the MA level after completing our own BA program, the content of which — language, culture, communication — is tailored to the needs of future translators and interpreters without providing professional translating and interpreting skills as such.

It is in this context that I have attempted my assessment of students' aptitude for interpreting, though my efforts are a matter of individual interest and initiative rather than one of institutional policy.

The BA curriculum includes a one-semester course that introduces students to the basics of interpreting. As the instructor for that course until recently, I used it for my experiment in aptitude assessment. The course is generic (non-language-pair-specific) and based on monolingual practical exercises. The intended learning outcome is consecutive interpreting (German–German) of a two-minute speech, with note-taking as needed. The course, which meets once a week for 90 minutes throughout the 15-week term, introduces students to such components of the interpreting process as active listening, message analysis, memory and note-taking. In the third week of the course, after an explanation of and exercises in concentration and active listening, students are asked to take the test described in this paper, after the following review of existing approaches and options.

2. Approaches to aptitude assessment

Two broad options can be distinguished: (1) a test, or set of tests, administered at a given point in time, and (2) an extended curricular component, or course, in which students' performance will be taken to reflect their aptitude.

2.1 Course-based assessment

The latter, course-based option was pioneered in the late 1970s by Barbara Moser (1978). Drawing on her model of simultaneous interpreting, she proposed exercises relating to various processing steps, such as abstraction of ideas (keyword), paraphrasing, prediction, dual-tasking and shadowing. In a correlational study based on data for 114 students over a four-year period (1978–1981), Moser-Mercer (1985) used four types of exercises within a ten-week monolingual course designed to introduce students to simultaneous interpreting: shadowing, dual-tasking (listening while counting, followed by recall of main ideas), paraphrasing, and number processing (shadowing text and noting down numbers). Assessment for the course involved such additional criteria as English language proficiency, assertiveness and resilience, and consisted in either a recommendation, a qualified recommendation, or no recommendation for training in SI. There was a high correlation between the type of recommendation issued at the end of the introductory course and students' results on the first-year (qualifying) and final (professional) exams.

Unfortunately, there are no further descriptions (e.g. of the scoring method for the paraphrase task), nor an explanation for dropping the prediction test used in the first, test-based study.

At any rate, course-based assessment as proposed by Moser-Mercer clearly has advantages over one-off tests. However, it is also costly in terms of time and teaching resources, especially in one-year MA programs and when applicant numbers are high.

2.2 Tests

Selective pre-admission examinations have been described by various authors, and a distinction can be made between psychometric tests (as carried out by Gerver et al. 1989 or, more recently, López Gómez et al. 2007) and those that seek to approximate the interpreting task as such and thus offer more validity, at face value, in measuring what the test is supposed to measure. With ample reference to their experience at Georgetown University, for instance, Bowen and Bowen (1989) discussed translational tasks, such as written translation, sight translation and interlingual summary, as well as intralingual recall tasks. The latter task, in particular, which, in terms of meaning processing, could be construed as monolingual consecutive, was said to reflect general knowledge and memory skills as well as linguistic proficiency.

There seems to be agreement that oral tasks — or, for the sake of the signed-language interpreting community, non-written tasks — offer higher face validity in assessing aptitude for interpreting, which is after all a real-time speech comprehension and production activity. Yet written tasks are usually easier to administer and score, and are therefore used especially for eliminatory screening prior to the testing of interpreting-specific skills.

Donovan (2003) reviewed exam procedures at twelve European institutions providing post-graduate interpreting courses. Aside from oral interviews with the applicants, the only test used by all institutions was interlingual oral reformulation, described as rendering a presentation of two or more minutes' duration in another language (among the applicant's active languages). Though also referred to as a "recall test", this task may be viewed as an approximation of consecutive interpreting and as such tests the type of skill that is to be acquired. The same applies to sight translation, of course, an intermodal, interlingual simultaneous processing task that cannot easily be described as a natural bilingual skill.

This holds true also for the intralingual oral simultaneous paraphrase task developed and tested by Russo and Pippa (2004). This approach avoids the interlingual dimension but hinges on simultaneous listening and speaking as in simultaneous interpreting. Aside from tapping into linguistic (lexical and syntactic)

flexibility, this task is essentially meaning-based, and test takers' output is scored for the propositional as well as pragmatic appropriateness with regard to the original speech.

As demonstrated in Russo & Pippa (2004), scoring the paraphrase protocols is a major issue, and the authors have moved from an initial attempt to assign a single numerical score (Pippa & Russo 2002) to a three-tier model in which every reformulation chunk can be classified at three levels of analysis, i.e. syntax and lexicon, semantics, and pragmatics. The level of syntax and lexicon has seven categories (including reduction, expansion and permutation), and there are four categories each under the rubrics of semantics and pragmatics.

The scoring approach described by Russo and Pippa (2004) is clearly at the more labor-intensive end of the spectrum, at least when compared to that of Bernstein and Barbier (2000/01), who propose a set of tests suitable for automatic scoring. Not surprisingly, most of the test items are of a linguistic nature (shadowing, word translation) so that the output can be handled by the speech recognition and matching system. The authors also mention a sentence translation and summarizing task (passages of 15 to 60 words) but are not very explicit on the scoring approach for these tests.

Scoring is clearly a major issue in designing an aptitude test. Another is validity with regard to the aptitude or skill component to be measured. On both counts, the paraphrase test proposed by Russo and Pippa (2004), though relevant and inspiring, seems less than ideal for asssessing a candidate's basic aptitude for the core skill of interpreting, that is, message comprehension followed by re-expression. Attention management for simultaneous listening and speaking in a meaning-based comprehension and reformulation task seems geared towards simultaneous interpreting, of which simultaneous paraphrasing is arguably the closest approximation.

For an MA program like the one at the University of Vienna, which offers a specialization in dialogue interpreting as an alternative to (simultaneous) conference interpreting, it seemed desirable to focus on a component skill that is relevant to either processing mode without mirroring one or the other. Hence the idea for the SynCloze test.

3. SynCloze test

3.1 Sources

The SynCloze test is chiefly inspired by Moser (1978), Gerver et al. (1989) and Russo & Pippa (2004), with further encouragement from the work of Bernstein & Barbier (2000/01) and López Gómez et al. (2007). (Other authors who proposed

and described the use of cloze exercises include Anaya & Lopez 1990, Lambert 1992 and Viaggio 1992.)

Given the focus on the simultaneous mode in Barbara Moser's (1978) seminal work, not all her tasks seem appropriate for testing generic core skills of interpreting. The most promising candidates seem to be keyword analysis, paraphrasing and prediction. The latter was implemented by Moser-Mercer (1985) as oral message completion based on knowledge from prior reading of a relevant text. Such gap filling was also an important component in the test battery administered by Gerver et al. (1989). Three out of the eight text-based tasks used by Gerver et al. (1989) were cloze exercises, one of which was presented as a recording, with every tenth word deleted and signaled by a tone. Upon hearing the tone, subjects were to write down the missing word. It was this semi-oral version of the cloze test (aside from two text-based memory tasks and an error detection task) that proved particularly predictive of successful completion of the six-month training course (as determined by consecutive and simultaneous interpeting exams).

While Moser (1978) and Gerver et al. (1989) do not offer a detailed description of their scoring method for these completion exercises, it is conceivable that more than one response could be correct. Gerver et al. (1989) speak of "counting the total number of exact responses", which would favor straight lexical matching over meaning-based appropriateness.

The same applies to their scoring of the "synonyms test", which consisted in writing down as many synonyms as possible to four stimulus words. Scoring based on "acceptable" responses was apparently unproblematic in the case of synonyms, though not so in their test of expressional fluency (rewriting test), for which the authors admit that "the scoring of this test was found to be subjective and difficult to accomplish, since it involved equating of meaning at the sentence level" (1989: 728).

3.2 Rationale

As highlighted by the paraphrase task, suggested by Moser (1978) and elaborated by Russo and Pippa (2004), interpreting, like translating, means expressing meaning "in other words". Expressional fluency, especially under time stress, would therefore be a crucial skill in interpreting. Before meaning can be re-expressed, it must of course be understood. Interpreters must therefore derive as complete an online mental representation as possible based on the input speech as well as on prior linguistic, contextual and encyclopedic knowledge. This online meaning processing requirement is basic to the consecutive as well as the simultaneous mode of interpreting, concurrently with storage operations and target-language production, respectively.

In an effort to tap into both the online comprehension and expressional skills, the test presented here combines a semantic cloze with a lexical or phrase-level paraphrase task. Due to its purely auditory/oral administration, it is crucially dependent on memory, both working memory — for processing and storage — and long-term memory, for lexical and content-related knowledge.

3.3 Design

A German text of approximately 660 words (prior to deletions) on the topic of "Mobility and Health" was recorded at a moderate speed (100 wpm), with a total of 24 sentence-final gaps of 5 to 7 seconds duration signaled by a tone (beep). The gaps are distributed evenly throughout the text, with at least one full sentence between incomplete ones.

Most gaps are based on content rather than on fixed linguistic collocations, and items with the verb-final structure that German is famous for are complemented by items with noun-based completions. Here is one example, stretching English syntax in the translation for the sake of illustration:

e.g.: *Der Anteil übergewichtiger Kinder und Jugendlicher in Österreich hat sich in den letzten Jahrzehnten* * [verdoppelt, stark erhöht, rasch vermehrt, nach oben entwickelt…]
[The percentage of children with obesity in Austria has in recent decades * [doubled, risen sharply, increased significantly…]

The completion hinges not on a fixed collocation but on prior context: The previous sentences stated that more and more people did not get enough exercise and that lack of exercise was a reason for being overweight.

Instructions are provided in a 200-word passage (2 min.) preceding the actual text (9 min. 40 sec.). This serves as a warm-up for the listening condition (students take the test in a booth wearing microphone-headsets) and offers two sample gaps relating to the explanatory introduction. In presenting test takers with the combined semantic cloze and lexical/phrase-level paraphrase task, they are instructed to complete the gaps at the end of sentences "as quickly as possible and with as many contextually appropriate variants as possible".

Performance was therefore to be measured by the number of synonymic completions as well as by the speed of reaction.

3.4 Administration

As explained in the Introduction (Section 1), the test was administered in the third week of my generic BA-level introductory course in interpreting. The first

test run took place in October 2005 with 35 students (Group V). Shortly thereafter, the test was administered to a control group of 11 MA students with one or two semesters of training in consecutive and simultaneous interpreting (Group I). In March 2006, the test was taken by another 24 students enrolled in the introductory interpreting course (Group W), and two more groups of close to 30 students were added in October 2006 (Group X) and spring 2007 (Group Y). The data for the latter groups were analyzed mainly for correlations (see Section 4.5 below). The protocols of these 127 participants were scored as described below.

3.5 Scoring

Students' performances in the booth were digitally recorded. Using a transcription sheet, individual completions were transcribed from the recording, using an audio player with oscillographic visualization to facilitate identification of the cloze items.

Contrary to somewhat naïve initial expectations, the varied nature of the responses made it evident that scoring would need to account not only for the quantity of completions but also for their quality. The following scoring system was therefore devised:

Two points were given for each completion, with one-point deductions for (1) grammar mistakes; (2) lack of coherence (contradictory semantics); (3) faulty collocation (e.g. "a risk is … strong"); and (4) amendments, i.e. completions that merely modify or extend rather than fully paraphrase a prior completion (e.g. "clearly visible", "distinctly visible").

All scoring was done by a graduate student using part of the material for her MA thesis (Tondera 2007). A small sample of doubtful cases was submitted to me for independent scoring, and my scores for such doubtful cases matched those proposed by the main scorer.

In addition to the quantitative and qualitative completion score, following a recommendation by Barbara Moser-Mercer (personal communication, 11 Sept. 2005), response times were measured using standard audio software (Audacity). Response-time values were subsequently added to the plain scores as a speed bonus as follows: For response times of up to 600 msec, a bonus of 50% was applied to the score for the first completion; for response times from 601 to 1600 msec (= max.), a graded bonus was calculated using the formula f(RT) = (Max − RT)/2. For example, a response time of 800 msec yields a bonus of 40%: f(800) = (1600 − 800)/2 = 40%. A response time of 1300 msec yields a bonus of 15%: f(1300) = (1600 − 1300)/2 = 15%.

4. Findings

4.1 Beginners vs. interpreting students

The mean scores and *SD* for the three groups are shown in Table 1.

Table 1. Mean scores for groups V and W vs. the control group (I)

	Group V (*n* = 35)	**Group W** (*n* = 24)	**Group I** (*n* = 11)
Mean	67.7	69.4	92.5
SD	25.9	27.1	23.8
Range	106.0	119.0	89.0

The scores for the two groups of BA students without exposure to training in interpreting are very similar, and markedly lower than the score for the interpreting students, who also have the smallest standard deviation and range. A *t*-test confirms that there is no significant difference between the mean scores of the two groups of beginner students (V, W). When comparing the means for the beginners ($n = 59$) and the interpreting students ($n = 11$), on the other hand, a one-tailed *t*-test is significant at the 1% level: $t(68) = -2.845$, $p = 0.006$.

4.2 Results by language profile

In the group of BA students, roughly one third had an A language other than German (V: 12/35; W: 8/24); Group I had only 2 non-German A students (out of 11). Students' language combination had been documented on the assumption that the SynCloze scores for students with German A would be higher than those for students with German as their B language. This hypothesis was confirmed (Table 2); a one-tailed *t*-test proved significant at the 1% level: $t(55) = 4.194$, $p = 0.000$.

Table 2. Mean scores by students' language profile

	Group V+W German = A (*n* = 38)	**Group V+W German = B** (*n* = 20)
Mean	76.6	48.8
SD	24.3	19.4
Range	109.0	87.0

Table 3. Response times

	mean RT Group total	**mean RT German A**	**mean RT German B**
Group V	1.87 sec	1.70 sec	2.30 sec
Group W	1.85 sec	1.69 sec	2.23 sec
Group I	1.46 sec	1.24 sec	2.48 sec

4.3 Response times

Response time measurements for the three groups show a significant difference analogous to the difference in mean scores (Table 3).

The more advanced interpreting students clearly had the shortest response times whereas the two groups of beginner students were nearly equal in this respect. This trend is also seen for the comparison by language status, with German-A students giving their initial cloze response significantly faster than those with German as a B language.

When the plain scores are recalculated by applying the speed bonus described above, the difference in mean scores by group becomes even more pronounced, as might be expected from the superior results on both mean scores and mean response times in Group I.

4.4 Results by cloze item

Though a qualitative analysis of the cloze items with the highest and lowest scores is beyond the scope of this presentation, it is interesting to note that several items (nos. 10, 13, 21) proved most productive throughout the three groups. One example is item 13, which was cited by way of illustration in the description of the test design. Others are items no. 10 and 21, which describe an opposite development ("less driving, more …") and a parallel increase, respectively. Cloze item no. 9, on the other hand, which relied on a logical inference from the immediate sentential context, yielded particularly low scores in nearly all groups.

4.5 Correlation

Whereas the SynCloze test has been shown to discriminate as expected between native-German and non-German students, and between BA-level beginners and MA-level interpreting students, who can be assumed to have developed their concentration, comprehension, anticipation and production skills in the course of one or two semesters of training in consecutive and simultaneous interpreting, the

ultimate test for the SynCloze approach is of course its predictive power for a given student's aptitude for interpreting.

According to Carroll (1961, cited in Bowen & Bowen 1989: 110), a learner's aptitude is defined as "an inverse function of the amount of time, other things being optimal, which will be required for the learner to attain a criterion mastery in the task to be learned." The assumption of "other things being optimal" was unfortunately not applicable to our curricular circumstances at the time, given the complex challenges of implementing the new curriculum, harmonizing teaching approaches and establishing final examination procedures.

Rather than a longitudinal study correlating SynCloze test results with program completion time or final exam grades, what may be attempted is to show a correlation between SynCloze test results and students' performance on the end-of-term exam in the introductory course. In keeping with the learning outcome defined for that course, students had to give a consecutive rendition in German of a German speech on a general topic (V: globalization; W: the life of Columbus). The speech was divided into two takes of less than two minutes. As with the SynCloze test, the final exam was taken in the booth, with students listening to the source speech and recording their output via a microphone headset. As the course also included an introduction to note-taking, students were allowed to take notes, but admonished to trust their memory skills as practiced for several weeks in the course. Exams were graded by the course leader on a seven-point scale (7 = max.).

Based on students' performance and informal feedback, it was clear that the authentic speech excerpts used in the final exam for Group V were extremely difficult. However riveting the anti-globalization tirade by Jean Ziegler, his eloquence, speed and Swiss accent proved overwhelming for all but the top students in the class. Indeed, the relationship between SynCloze test and final exam results for Group V is entirely skewed, with no German-B student obtaining more than 5 points on the exam while German-A students with the highest SynCloze scores account for all the top-graded exams (6 or 7 points).

In contrast, Group W took the final exam with a less challenging but nevertheless authentic speech on the life of Christopher Columbus, again in two takes of less than two minutes. Exam results showed a significant correlation with SynCloze test scores (Spearman's rank coefficient $r_s = .754$; $n = 18$, $p < 0.01$, one-tailed).

For the subsequent test groups, the correlation was only moderate — Group X: $r_s = .545$, $n = 29$, $p < 0.01$; Group Y: $r_s = .591$, $n = 27$, $p < 0.01$. It seems that a number of confounding variables may be at play, including the extent of students' independent practice of monolingual consecutive during the course and their use, or overuse, of note-taking during the exam, but it is not clear why their impact would vary in the different groups. Moreover, as pointed out by Gile (2005), intrinsic variability is always a factor to be reckoned with.

5. Conclusions

Based on the findings for the first five groups of test takers, including a control group of MA students with one or two semesters' experience in interpreting, one can conclude with some confidence that the SynCloze test measures interpreting-related subskills, such as online comprehension, oral expressional fluency constrained by contextual appropriateness, and fast reaction times. The latter are presumably associated with a chance for higher completion scores, given the risk of memory decay in slow responders. Indeed, the SynCloze test has proved quite challenging even for advanced and experienced students. This is probably due to the high cognitive load resulting from the need to search for and retrieve synonyms from long-term memory while at the same time matching alternative completions against the truncated input sentence that must be held in working memory. Failure to do the latter may explain why many students, particularly with German as a B language, manage no more than one completion for many of the test items.

From a task-oriented perspective, this difficulty might be explained by the closure effect resulting from producing a successful completion. Interpreters do not normally offer synonyms for their output (though they do repair it more often than one might expect, as shown by Petite 2005). In other words, having said something well might discourage a search for (possibly inferior) alternatives. This, however, could also be held against the simultaneous paraphrase task, which forces test takers to move away from a formulation that may be highly appropriate.

As regards the test's predictive power, it appears to apply at least to a moderate degree to a monolingual consecutive interpreting task despite the rather different nature of the test and the end-of-term exam, which involves less speed stress and no online response. One might therefore hypothesize that the SynCloze test could predict aptitude also for a simultaneous interpreting task. Unfortunately, our present curriculum does not include a course of that nature in which the test could be embedded.

The findings from this test series also suggest that the embedded test approach may be preferable to an attempt at correlating aptitude with successful program completion, at least in times of profound curricular change, staff turnover, and uneven teaching standards in the twelve languages in which interpreter training is offered at the University of Vienna.

Even under comparable optimal conditions across language sections, aptitude test results may be hard to correlate with program completion times or final grades, given such factors as study-abroad periods, ambitious MA thesis projects, part-time work, parenthood or, quite generally, individual learning styles and motivation, as investigated in several contributions in this volume. Which brings us

back to the fact that aptitude for interpreting is a multidimensional construct, only a part of which, however crucial, may be addressed by the SynCloze test.

References

Anaya Dávila Garibi, G. & Lopez Islas, J. (1990). Oral cloze: A backup exercise for interpreting. *Meta* 35 (3), 647–651.

Bernstein, J. & Barbier, I. (2000/01). Design and development parameters for a rapid automatic screening test for prospective simultaneous interpreters. *Interpreting* 5 (2), 221–238.

Bowen, D. & Bowen, M. (1989). Aptitude for interpreting. In L. Gran & J. Dodds (Eds.), *The theoretical and practical aspects of teaching conference interpretation*. Udine: Campanotto, 109–125.

Donovan, C. (2003). Entrance exam testing for conference interpretation courses: How important is it? *Forum* 1 (2), 17–45.

Gerver, D., Longley, P. E., Long, J. & Lambert, S. (1989). Selection tests for trainee conference interpreters. *Meta* 34 (4), 724–735.

Gile, D. (2005). Empirical research into the role of knowledge in interpreting: Methodological aspects. In H. Dam, J. Engberg & H. Gerzymisch-Arbogast (Eds.), *Knowledge systems and translation*. Berlin/New York: Mouton de Gruyter, 149–171.

Keiser, W. (1978). Selection and training of conference interpreters. In D. Gerver & H.W. Sinaiko (Eds.), *Language interpretation and communication*. New York: Plenum Press, 11–24.

Lambert, S. (1992). The cloze technique as a pedagogical tool for the training of translators and interpreters. *Target* 4 (2), 223–236.

López Gómez, M. J., Bajo Molina, T., Padilla Benítez, P. & Santiago de Torres, J. (2007). Predicting proficiency in signed language interpreting: A preliminary study. *Interpreting* 9 (1), 71–93.

Moser, B. (1978). Simultaneous interpretation: A hypothetical model and its practical application. In D. Gerver & H.W. Sinaiko (Eds.), *Language interpretation and communication*. New York: Plenum Press, 353–368.

Moser-Mercer, B. (1985). Screening potential interpreters. *Meta* 30 (1), 97–100.

Moser-Mercer, B. (1994). In S. Lambert & B. Moser-Mercer (Eds.), *Bridging the gap: Empirical research in simultaneous interpretation*. Amsterdam/Philadelphia: John Benjamins, 57–68.

Petite, C. (2005). Evidence of repair mechanisms in simultaneous interpreting: A corpus-based analysis. *Interpreting* 7 (1), 27–49.

Pippa, S. & Russo, M. (2002). Aptitude for conference interpreting: A proposal for a testing methodology based on paraphrase. In G. Garzone & M. Viezzi (Eds.), *Interpreting in the 21st century: Challenges and opportunities*. Amsterdam/Philadelphia: John Benjamins, 245–256.

Russo, M. & Pippa, S. (2004). Aptitude to interpreting: Preliminary results of a testing methodology based on paraphrase. *Meta* 49 (2), 409–432.

Sanz, J. (1930). Le travail et les aptitudes des interprètes parlementaires. *Anals d'Orientació Professional* 4, 303–318.

Tondera, D. (2007). *Zur Eignung für das Konferenzdolmetschen: Der SynCloze-Test in Erprobung*. Diplomarbeit, Universität Wien.

Viaggio, S. (1992). Cognitive clozing to teach them to think. *The Interpreters' Newsletter* No. 4, 40–44.

Putting interpreting admissions exams to the test

The MA KD Germersheim Project

Catherine Chabasse and Stephanie Kader
Johannes Gutenberg University Mainz

With increasing numbers of students wishing to become conference interpreters, but limited capacities in most university degree programs, accurate admission testing is an important means of predicting an applicant's chances of completing the program successfully. This article focuses on three aptitude tests for simultaneous interpretation: Pöchhacker's SynCloze test; Chabasse's cognitive shadowing test; and Timarová's personalized cloze test. The test battery was administered at the start of the 2009/2010 academic year to students beginning the two-year Master's program in conference interpreting (MA KD) at Germersheim. Correlations between test performance and subsequent exam grades at the end of the second semester were examined for all three tests. Given the large number of applicants each year, practical feasibility of the tests was taken into consideration with a view to scheduling the format and content of the entrance exam for the 2012/2013 academic year. In this perspective, cognitive shadowing was identified as the most useful test under the existing time constraints.

1. Introduction

In an increasingly globalized world, the demand for professional interpreters is constantly rising. To meet this demand, universities around the world are offering conference interpreting qualifications in many different languages. At the same time, the need to test steadily increasing numbers of applicants faces these universities with a growing challenge.

In 2004 the School of Translation, Interpretation, Linguistics and Cultural Studies of the University of Mainz in Germersheim (FTSK) organized a pre-admission aptitude test for the first time. This occurred when the interpreting study program was restructured, as part of the Bologna Process: the traditional *Diplom*

(earned after eight semesters of post-secondary school study) was replaced by a Master's program in four semesters (MA KD), following a BA degree in any field. As students came from differing backgrounds, the FTSK had to assess whether they had the necessary aptitude to succeed in the interpreting program.

The first admission exam at the FTSK in 2004 was in two parts: written tests (multiple choice and essay in the B and C languages), which students had to pass before being admitted to the oral part of the examination. One year later, in 2005, the sequence was reversed. Neither of these arrangements was entirely satisfactory: the results were subjective, and the oral tests focused only on aptitude for consecutive interpreting (CI), not for simultaneous interpreting (SI).

To remedy this situation, a research study was undertaken (Chabasse 2009a). This study used various interpreting aptitude tests among students at three German universities: Germersheim, Heidelberg and Saarbrücken. The written and oral exercises were grouped into three categories: language skills (in the A language and B language/s), cognitive skills, and personality traits.

The paraphrasing exercises and cloze tests used in the study had also been part of previous admission exams for conference interpreting at the FTSK. However, there had not yet been uniform implementation standards for the different languages, and there was thus no quantifiable basis which would have allowed overall assessment of the tests. Hence the importance of developing a systematic and empirically validated approach. In this respect, the 2009 "Aptitude for Interpreting — Towards Reliable Admission Testing" conference in Antwerp had provided a timely statement of the need for greater research "before researchers can hope to supply educators with a full range of reliable and practical methodologies for testing the aptitude of applicants to interpreter training programs" (Shlesinger & Pöchhacker 2011: 4).

The research team working on this topic at Germersheim evaluated a series of aptitude tests to determine which would be best suited to the practical requirements of entrance examinations for conference interpreting courses. Conclusions were drawn by correlating test scores and exam grades after one year of study.

Against this background, the aim of the present study was to assess the predictive qualities of the tests used and examine the possible need for booth-dependent tests, previously not used in the context of the admission exams. A secondary objective was to find out how feasible booth-dependent tests are with large numbers of participants, in terms of both logistics and grading.

2. Interpreter aptitude testing

2.1 Aptitude and aptitude testing

In the context of this study, aptitude is understood to be the interaction of cognitive competences and capacities and non-cognitive personality features. These are detailed in Chabasse's (2009a: 122) aptitude model for SI, which is based on Thurstone's Multiple-Factor Theory (1938). The model is reproduced in Figure 1.

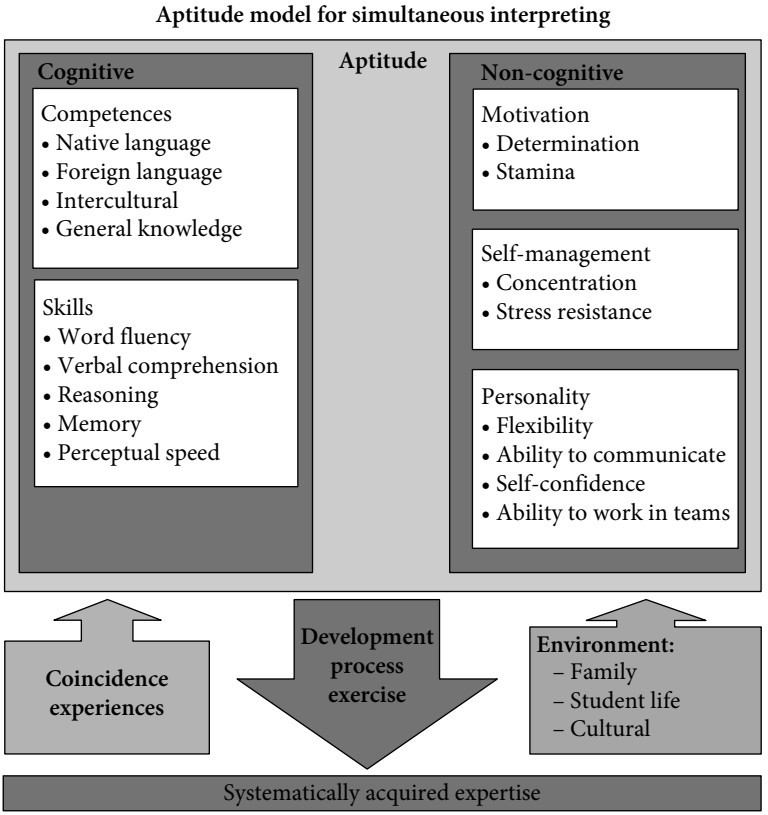

Figure 1. Aptitude model for simultaneous interpreting (Chabasse 2009a)

The competences shown on the left side of the table can be acquired (e.g. foreign languages, intercultural and general knowledge) and are knowledge-based, whereas the items listed under "skills" represent the student's ability to use his or her acquired competences. The cognitive competences shown are necessary prerequisites for successfully completing the degree program in the allotted time frame. Candidates' language proficiency with a view to the MA degree program should be at the C1/C2 level in the Common European Framework of Reference

for Languages, in order to prevent biased results and to ensure that the admission tests actually reflect aptitude for conference interpreting rather than language skills as such (see Chabasse 2009a: 183, 197). Language proficiency tests can be administered in written or oral form; however, since evaluating written tests would be impractical in light of the large number of applicants each year, assessment of language proficiency at the FTSK admission exams has always been based on oral presentations by the candidates.

The non-cognitive aptitudes on the right side of the table include motivation, personality and self-management. While the concentration and stress resistance identified under "self-management" are basic requirements for SI, concentration also forms the basis for a candidate's language transfer capabilities (cf. Chabasse 2009a: 108–113). The exercises used for the study were designed to test these aptitudes. Since the study was meant to identify suitable means of assessing overall aptitude for SI, the decision was made to simulate the SI environment by including an exercise in an interpreting booth.

Several factors have to be considered when compiling a test battery that is designed to determine candidates' aptitude and their likelihood to perform and learn well. A multitude of publications exists on the topic of interpreting aptitude, taking into account the different skills aspiring interpreters need. A detailed survey of these studies can be found in Russo's paper on "Aptitude testing over the years" (in this volume).

Timarová and Ungoed-Thomas (2008) conducted a survey among 18 schools offering interpreting degrees. They compiled a list of tests for the different skill categories tested by the schools, which is reproduced in Table 1.

The problem with these test batteries is that nearly all of them require the candidate to have interpreting skills already. They also seem to neglect SI completely.

"Teachability", a term used in the 1970s by Seleskovitch and Lederer, and readiness to perform are the cornerstone of the Germersheim tests. Students wishing to become conference interpreters need to enter the program already equipped with certain skills; these skills will form the basis of the abilities to be acquired subsequently, through the degree program — particularly its practical components. The students do not, however, need interpreting experience or prior knowledge in the field before beginning the program. That is why only tests that do not require explicit interpreting skills were selected for this study.

2.2 Current testing at Germersheim

Admission exams at Germersheim are usually taken over a period of two days, with the different tests organized according to the candidate's language combination. Since applicants come to Germersheim from all over the world, considerable

Table 1. Overview of major skill categories and most popular associated tests (Timarová & Ungoed-Thomas 2008: 39)

Skill category	Test	Schools using the test/method
Language	Short consecutive	14
	Short speech made by candidate	9
	Interview	7
	Summary	7
	Translation	7
Communication	Short speech made by candidate	11
	Short consecutive	10
	Summary	6
	Interview	5
Comprehension	Summary	8
	Translation	8
	Short consecutive	5
Analysis	Summary	11
	Short consecutive	6
	Translation	5
General Knowledge	Interview	8
	Written test	4
	Short speech made by candidate	3

effort is made to schedule the tests accordingly, thus facilitating visa and travel arrangements for international applicants. Test results are announced on the same day, or on the last day of testing at the latest, before the students depart from campus.

The current admission tests are based loosely on the AIIC criteria for aptitude in conference interpreting (AIIC 2006), and include the activities listed below.[1]

a. Presentation: In order to test cultural knowledge of the countries and regions in which the 'B' language is spoken, candidates are given 10 minutes to prepare a short speech in response to a question relating to cultural topics or current events in the part(s) of the world concerned. Candidates then have 10 minutes to present the speech in the 'B' language, allowing the assessors to evaluate 'soft' skills (presentation, stress management, communication,

1. The same tests are administered to all candidates, irrespective of language combination, educational or personal background. In order to qualify for the admission exams, candidates must have completed a BA, MA or equivalent degree from an accredited university or college. Exemption from testing is granted only to those candidates who can prove that they have previously passed interpreting tests, either at other universities or during their BA interpreting modules (where applicable). Any such exemptions are decided on a case-by-case basis.

improvisation, ability to react to questions and/or criticism), verbal fluency (vocabulary, grammar, coherence) and knowledge of the subject (facts, ability to reflect).

b. Listening comprehension: A three- to five-minute text in the 'B' language is read aloud, during which time the candidate is allowed to take notes. S/he is then asked to summarize the text in the 'A' language and be prepared to answer questions regarding the content of the text. This test affords insight into the candidate's listening comprehension skills (degree of completeness and precision), 'A' language production skills (coherence, cohesion, idiomatic correctness) and ability to process content while performing language transfer. While this exercise does not require any prior knowledge of CI, students from undergraduate programs that offer introductory courses in CI are generally at an advantage.
c. Cloze exercise: The candidates are given a text in their 'A' language, from which certain words and phrases have been deleted. Candidates are initially asked to read the text out loud and to fill the gaps with words or phrases in their 'A' language. Afterwards, they are expected to translate their suggested solutions into the 'B' language and to add synonyms, again in the 'B' language (5–10 minutes). Assessment criteria include the number of synonyms, collocations and speed.
d. Personal interview: During the interview, the aim is to gain an overview of the candidate's interpersonal (e.g. ability and willingness to work in a team) and intra-personal skills (e.g. self-confidence, motivation). This is done by asking about his/her personal background, experience away from home, and education. Assessors can gauge the candidate's motivation to enter the program and his/her degree of self-reflection, which gives an overall sense of whether or not s/he would benefit from the program.

In view of the growing number of applicants to the MA KD program and the logistic needs associated with booth-dependent tests, no such tests were included until 2011. While drop-out rates for the MA KD program have been negligible so far, failure rates continue to increase in alarming numbers, although most candidates pass their intermediate or final exams the second time around. This has prompted the question of whether a booth-dependent SI admission test could provide a more detailed perspective on the individual student's initial prospects of success in the program.

3. The new FTSK admission test

One of the main problems identified by the project team was that current testing lacked scientific parameters according to which test sections would be graded. This made the evaluation of a candidate's performance highly subjective and assessor-dependent. Another perceived shortcoming was the absence of tests geared to SI. In this respect, one of the main questions discussed at the Antwerp conference was whether different tests were necessary to evaluate a prospective student's aptitude for CI as opposed to SI. In the test regime which was later implemented, the project group focused first and foremost on SI: no specific tests were included to test at this stage for CI aptitude, which will be examined in a later study.

3.1 Piloting the new FTSK admission tests

In addition to German native speakers, the interdepartmental project team included a native speaker from each of the foreign languages offered within the MA KD program.

The study took place in November 2009. Given the small number of potential test participants in some departments, only English, Italian, and French were considered for the study. To ensure comparability of results and limit the number of potential confounding variables, only individuals with German as their 'A' language were included. The group comprised students who had just enrolled in the Germersheim MA KD program for the 2009/2010 academic year, having successfully passed the admission tests in both their foreign languages. The 'B' languages of the 24 participating students were: English (12 students), French (8), or Italian (4). The study cohort was fairly homogenous: three of the participants were male, and only three participants were over the age of 30 at the start of the study. Participation was voluntary.

The tests were conducted early in the first semester of the graduate program. Students participated in the tests during classes in the 'B' language; their instructors administered and graded the different tests. Participation was almost 100%, students having been told that their performance on the tests would not impact their course grades.

3.1.1 *Correlation parameters*

The grades that students achieved in the intermediate exams at the end of the second semester were chosen as correlation parameters. Taking into account that students inevitably acquire and consolidate skills at their own pace, a follow-up study will re-evaluate findings once all students have completed the MA

program. For this purpose, a third correlation parameter will be used: final exam grades.[2]

3.1.2 *Institutional considerations*

The individual tests included in the project were chosen with regard to their practical feasibility for large numbers of applicants and the extent of grading required. Uniform criteria were provided for the grading.

During the run-up to the tests, questions were raised as to how many tests are actually needed to determine a person's suitability for the program, and how many tests can feasibly be administered if the applicants are to be given the results on the same day. The final choice of tests is indicated in the next section. All are booth tests.

3.2 Tests chosen for our candidates

3.2.1 *Tests in the 'A' language (German)*

1. SynCloze: The study included the SynCloze test, presented at the Antwerp conference by Pöchhacker (in this volume), which is a booth test to assess powers of anticipation and vocabulary. This choice of test was based on the promising findings reported in Antwerp, an additional advantage being the possibility of comparing groups in Vienna and Germersheim.
   Test: Candidates heard the 660-word Pöchhacker recording of a text on mobility and health, and were asked to fill the gaps with words of their choice. The recording speed was 100 words per minute; the total length of the text was 9 minutes and 40 seconds. There were 24 five- to seven-second gaps in end-of-sentence position, which students were instructed to fill with as many solutions as possible. A 200-word (two-minute) introduction included two practice sentences, so that students could familiarize themselves with the exercise beforehand. All students were recorded, for subsequent evaluation.
   Scoring: Students received two points per completed sentence. The maximum score was thus 48 (24 × 2). One point was deducted for grammatical mistakes, lack of coherence, non-standard collocations, or elaborations. A difference from the original setup of the test in Vienna was that we did not take into account response times or lengthy décalage.

---

2. Student data may not all be available at the same time, since students who failed the intermediate exams need to retake them after one semester before being allowed to move on to the next class. Periods of study abroad can also mean that some students may take their final exams later than others.

2. Cognitive shadowing: This test, based on Ingrid Kurz' interpreter training exercise (1992), was adapted by Chabasse to evaluate language proficiency, fluency and ability to concentrate (Chabasse 2009a, 2009b). The test was chosen to broaden the range of skills included in the 2009 study, making it possible to assess ability to listen, speak and think at the same time without actually requiring any previous experience in SI.
   Test: There were six sections of 30 questions[3] each: 1a, 'yes' or 'no' questions, A-A; 1b, 'yes' or 'no' questions, B-B; 2a, 'why' questions, A-A; 2b, 'why' questions, B-B; 3a, 'why' questions, B-A; 3b, 'why' questions, A-B. All questions were recorded, with short introductions and instructions for each section. No interruptions were allowed: students were expected to answer the first question while already listening to the next, and so forth. All answers in the first two sections had to consist of complete sentences, either "*Yes* + repetition of original concept in affirmative form" or "*No* + repetition of original concept in negative form". In Sections 2a-3b, the original concept expressed in the question did not have to be explicitly repeated in the answer. Questions in Sections 1a and 2a were devised to test the candidates' concentration and retention skills, while Sections 1b and 2b focused on concentration and the quality of 'B' language oral production. Questions in Sections 3a and 3b included language transfer from the 'A' to the 'B' language and vice versa, thus testing concentration and language transfer capabilities.
   Scoring: One point was awarded per correct and complete answer. One point was deducted if the answer was wrong, incomplete, or included grammatical and/or collocational errors. Students were allowed to answer "I don't know if/why…" if they were unsure about the right answer: answers of this kind were awarded one point, as long as they were complete and grammatically sound. The maximum possible score was 180.

3.2.2 *Tests in the 'B' language (English, French, Italian)*

3. SynCloze: Based on the Pöchhacker text, this test involved the same criteria as for the 'A' language in terms of text length, rendition speed, numbers of gaps and scoring. Some adjustments were made to the Italian, French and English texts to optimize positioning of the gaps in the text. Again, the students' response time was not recorded.

---

3. The factual content of the question has to be inconsequential, meaning that the students should be able to answer the questions without too much cognitive effort being spent on grasping the content and finding a correct answer (see Kurz 1992).

4. Personalized cloze: The test was first presented by Timarová at the conference in Antwerp and was originally conducted with interpreting students at Charles University, Prague.
   Test: Listening to a short text describing the life and career of a fictional person, students were asked to shadow the text freely: they were not required to repeat it verbatim, but to substitute the information they heard with corresponding information about their own life and career.
   Scoring: Two points were awarded for each substituted item of information if the student said anything at all, while one point was deducted for collocational and/or grammatical errors. The maximum number of points depended on the number of substitutions to be made in the text.

Weighting

The texts used in the different language departments varied in length and the number of necessary substitutions. To allow a comparison, the points obtained for each test were calculated as a percentage of the maximum possible score. The percentages were then used to calculate the correlations with examination grades.

4. Results

Table 2 shows the correlations between the individual test sections and the grades the students achieved in each of the individual examinations (CI A-B, SI A-B, CI B-A, SI B-A), after one year in the program, at the end of the 2010 summer term (second semester). Out of scientific curiosity, correlations were also calculated with regard to CI, although the study's main focus was SI. In the German school and university system, the top grade is 1.0, while a grade of 4.0 is the minimum pass. Since the evaluation scale in the project tests worked the other way (i.e., the higher the score, the better), the ideal demonstration of their predictive value would be a negative correlation between success in the proposed tests (high score) and in the end-of-term exams (low score): the higher the negative correlation, the better the predictive value.[4] Correlations, based on Pearson's linear coefficient (one-tailed, $n = 24$), will be discussed here only at a level of $\geq -.35$.

4. The interpretation of the strength of correlation is based on Cohen (1992): 0.1–0.3 = weak; > 0.3–0.5 = moderate; > 0.5–1.0 = strong.

Table 2. Correlations between test section scores and grades

Test	CI A-B	SI A-B	CI B-A	SI B-A
1. Cognitive shadowing[5]				
1.1 Concentration	−.13	−.44*	−.26	−.44*
1.2 Concentration and language proficiency	−.56**	−.46*	−.16	−.26
1.3 Concentration and language transfer	−.57**	−.39*	.03	−.06
2. SynCloze				
2.1 SynCloze A (German)	−.28	−.39*	−.14	−.13
2.2 SynCloze B (English, Italian, French)	−.57	−.39*	−.17	−.27
3. Personalized cloze				
Personalized cloze B (English, Italian, French)	−.35*	−.52**	−.14	−.14

Note. *$p < .05$; **$p < .01$.

At first glance, all tests used in the study showed significant correlations against SI A-B, but not SI B-A. In the following section, the results for the individual tests will be discussed.

4.1 Cognitive shadowing

Part 1.1 is designed to test concentration in relation to items processed in the 'A' language ('yes'/'no' and 'why?' questions). The correlation of 0.44 with both SI A-B and SI B-A is consistent with the need for concentration for the comprehension and anticipation components of SI, irrespective of directionality.

Interestingly, Section 1.2 ('yes'/'no' and 'why?' questions, B-B) and Section 1.3 ('why?' questions, B-A and A-B) correlate significantly not only with SI A-B (0.46 and 0.39 respectively), but also with CI A-B (0.56 and 0.57 respectively). Although the cognitive shadowing test as such is an exercise geared to SI, CI too (albeit with different dynamics) involves retention and retrieval of information under pressure of time. However, the correlation identified here with students' subsequent CI grades could equally be a function of their language proficiency. Possibly the language transfer in Section 1.3 ('why?', in A-B and B-A) involves cognitive processes which are a distinctive feature of CI as opposed to SI — e.g., memorization, segmentation and subsequent production with décalage of more than a few words.

5. In the table used for data processing and depiction, the system is as follows: Sections 1a and 2a = 1.1 (concentration), Sections 1b and 2b = 1.2 (concentration and language proficiency), Sections 3a and 3b = 1.3 (concentration and language transfer).

4.2 SynCloze (A and B)

There is a -0.39 correlation between the SynCloze test, in both the 'A' and 'B' languages, and SI A-B. For SynCloze A, this correlation is consistent with the role of good comprehension (in this case, of 'A') as an essential basis of SI. Regarding SynCloze B, the correlation with SI A-B is readily explained in terms of the need for proficient expression in 'B'.

This exercise tests for two related but discrete aspects: anticipation and language proficiency. A combination of these two factors seems to be involved: SynCloze A first and foremost tests ability to anticipate in 'A', while SynCloze B focuses on 'B' language proficiency. Together, these results create a complete picture with regard to the correlation of SynCloze A and B with SI A-B. Essentially an anticipation exercise, SynCloze A shows only a negligible correlation with SI B-A, while the correlation between SynCloze B and SI B-A is somewhat higher. Students who perform well in SynCloze A demonstrate good anticipation and overall language skills in their 'A' language, whereas a high rating in SynCloze B reflects 'B' language proficiency. If students who show good anticipation skills in their 'A' language do not achieve good grades in the subsequent SI B-A exam, this is possibly the result of problems with 'B' language comprehension.

It follows that the SynCloze test is less suitable as an admission test with regard to SI B-A than SI A-B. Pöchhacker's suggestion (in this volume) that the test might be indicative of SI aptitude is thus supported only partially by the results of the Germersheim tests.

4.3 Personalized cloze

In scoring the test, the assessors disregarded the factual truth of the information given and focused on linguistic factors. Students were told at the outset that they could make up answers if they preferred, as long as linguistic standards were respected. The test's correlation with SI A-B is presumably based on linguistic factors: listening to a text in the 'B' language, the students were required to provide substitutions in the 'B' language under controlled time constraints.

Comprehension and analysis of the source text was of primary importance for successful completion of the test. The opportunity to rephrase the sentence and break up the syntactic structure where necessary mimics a SI situation, with the added challenge of having to insert new information into the text and deleting the original content — a process that is admittedly very different from real-life SI. The reason for the differing degrees of correlation with CI A-B and SI A-B (–.35 vs –.52) could be the exercise's essentially simultaneous nature. 'Real-time' synonym

production and processing of connected discourse are particularly important in SI (see also Russo & Pippa 2004).

The personalized cloze, though disliked by the students, gave correlation results that were among the most predictive in the study.

5. Discussion

The **cognitive shadowing** test is one of the most promising indicators of interpreting aptitude, but will have to undergo some changes in order to be useful and feasible as part of an admission exam.

During the admission tests in 2012, this test took up too much of the time allotted for the 'B' language section of the overall admission exam. Since it would be unwise to sacrifice any of the other tests, cognitive shadowing needs to be shortened to ten minutes. Given that the test can be graded while it is in progress, the results are available as soon as the candidate leaves the booth, which is undoubtedly an advantage.

The SynCloze A and B tests will have to be discarded, as a result of feasibility issues: testing would need to encompass both the 'A' and 'B' languages in order to garner conclusive results. Further difficulties with SynCloze A involved the syntactic differences between German, French, English and Italian. Whereas German allows for the gaps to be placed toward the end of the sentence, the same was not always possible for the other languages. This led to varying levels of difficulty with regard to the required anticipation skills. In addition, the test requires recording of candidates and the process of evaluating the recordings is too long for the practical needs of the admission exam.

While the **personalized cloze** test showed considerable predictive value, the main problem with administering this kind of test during the admission exams is that results are needed on the same day, making immediate scoring necessary. 'Real-time' assessment of candidates' performance is virtually impossible and subsequent evaluation of recordings would (as for SynCloze) take too long, in view of the large number of applicants.

Since the personalized cloze yielded significant results for SI A-B, an equivalent test in the 'A' language might also be of interest. However, the correlations evidenced by this test in its present form essentially provide the same information as the cognitive shadowing test — which is far easier to evaluate and therefore preferable for the purposes of an admission exam.

6. Conclusions

Among professional interpreters and interpreting instructors, ideas about the "perfect" interpreter have not changed much over the years, providing a fairly stable archetype as a parameter for aptitude-testing research. The tests conducted within the framework of this project were geared towards testing aptitude for SI, not CI. Although no specific CI tests were administered, the correlations which emerged with CI exam grades nevertheless showed that the tests for SI were to a certain extent predictive of CI aptitude.

The project has shown that significant correlations exist between certain tests and exam performance in the conference interpreting MA, leading the project team to believe that future admission testing should ideally include at least one booth-dependent segment. Such tests were not only promising in terms of predictive capabilities, but they also created a reasonably realistic interpreting environment in which to test the applicants. Booth-dependent testing can be especially informative for those applying for a non-consecutive MA degree, i.e. those students who have had no previous exposure to interpreting and have not graduated from a translation-related BA program. Considering the time and space constraints under which admission testing is conducted, the logistic demands of organizing booth tests for an ever-growing number of applicants to Germersheim's MA program can seem daunting. Such tests should nevertheless be incorporated, for the sake of effective, reliable aptitude testing.

To complement the test battery with regard to 'A' language proficiency, the 'simultaneous' paraphrase test described by Russo (in this volume) could be used to evaluate deep semantic processing and memory capacity. This approach might provide further insight into SI aptitude. Accordingly, a paraphrase test was administered at the beginning of the 2011/2012 academic year. Since the results of the test will again be correlated to the results achieved during the student's intermediate exams two semesters later, the final analysis is still awaited.

References

AIIC (2006). Advice to students wishing to become conference interpreters. *Communicate!* May/June 2006. http://aiic.net/page/56 (accessed 7 October 2010).

Chabasse, C. (2009a). *Gibt es eine Begabung für das Simultandolmetschen? Erstellung eines Dolmetscheignungstests*. Berlin: Saxa Verlag.

Chabasse, C. (2009b). Eignungstest für das Simultandolmetschen: Erstellung und Erprobung eines Dolmetscheignungstests mit Schwerpunkt Simultandolmetschen. In: M. Albl-Mikasa, S. Braun & S. Kalina (Eds.), *Dimensions of second language research*. Tübingen: Narr, 183–194.

Cohen, J. (1992). A power primer. *Psychological Bulletin* 112 (1), 155–159. DOI: 10.1037/0033-2909.112.1.155

Kurz, I. (1992). 'Shadowing' exercices in interpreter training. In C. Dollerup & A. Loddegaard (Eds.), *Teaching translation and interpreting: Training, talent and experience*. Amsterdam: John Benjamins, 245–250.

Russo, M. & Pippa, S. (2004). Aptitude to interpreting: Preliminary results of a testing methodology based on paraphrase. *Meta* 49 (2), 409–432. DOI: 10.7202/009367ar

Shlesinger, M. & Pöchhacker, F. (2011). Introduction: Aptitude for interpreting. *Interpreting* 13 (1), 1–4. DOI: 10.1075/intp.13.1.01int

Timarová, Š. & Ungoed-Thomas, H. (2008). Admission testing for interpreting courses. *The Interpreter and Translator Trainer* 2 (1), 29–46.

Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.

ABOUT THE AUTHORS

Daniel Bauwens graduated in 1981 from Erasmus University College in Brussels as a conference interpreter with Dutch, English and French as working languages. Since then he has been a freelance interpreter, and has added Danish and German to his repertoire. He is the head of the Danish section at the Applied Linguistics Department of the Free University of Brussels (VUB), where he teaches Danish and French and coordinates the master's programme in interpreting. He is also a guest lecturer for French at the postgraduate programme for conference interpreters at the Catholic University of Leuven (KU Leuven, campus Brussels). He is a member of AIIC and of the legal interpreters' screening commission for Danish at KU Leuven, campus Antwerp.
Address: Vrije Universiteit Brussel – TTKA • Pleinlaan 5 • 1050 Brussels • Belgium
E-mail: daniel.bauwens@vub.ac.be

Karen Bontempo has nearly 25 years experience as an Auslan (Australian Sign Language) interpreter and has worked as an interpreter educator for the past 18 years. Her academic qualifications span the disciplines of psychology, linguistics and education. Karen holds a PhD from Macquarie University, where she is an Honorary Associate of the Linguistics Department. In addition to her work as an interpreter and an interpreter educator, Karen works part time as a teacher of deaf students in a secondary school. She is the national chairperson of the Interpreter Trainers' Network in Australia; serves on the national interpreter examination panel, and the national interpreting accreditation authority's qualifications assessment committee. Karen's research interests include aptitude for interpreting, evaluation of interpreter performance, mentoring, and issues surrounding interpreting pedagogy.
Address: Department of Linguistics • Macquarie University • Sydney NSW 2109 • Australia
E-mail: karen.bontempo@mq.edu.au

Catherine Chabasse studied German and French Studies in Paris from 1972 to 1977 and worked as a teacher for 8 years. She then graduated from the School of Translation, Interpreting, Linguistics and Cultural Studies of the University of Mainz in Germersheim, Germany in 1989 and since then has been working as a freelance conference interpreter and teaching translation and interpreting in Germersheim. She received her PhD in 2009, with a dissertation on aptitude for conference interpreting.
Address: FTSK / FB 06, Department of Intercultural German Studies • Johannes Gutenberg University Mainz • An der Hochschule 2 • 76726 Germersheim • Germany
E-mail: chabasse@uni-mainz.de

Andrew R. A. Conway is a Senior Lecturer in the Department of Psychology at Princeton University. He is also the Principal Investigator of the Human Working Memory Lab at Princeton and is an affiliate faculty member of the Princeton Neuroscience Institute and the Princeton Statistics and Machine Learning program.
Address: Peretsman-Scully Hall • Princeton University • Princeton, NJ 08540 • USA
E-mail: aconway@Princeton.edu

June Eyckmans obtained her PhD at the Radboud University of Nijmegen (the Netherlands) in 2004 on the methodology of L2 vocabulary assessment. She is currently Lecturer of English at

Ghent University and her research interests include cognitive approaches to foreign language learning and the methodology of interpreting and translation assessment. She publishes in national and international journals on foreign language acquisition, translation assessment and interpreting studies. She is the supervisor of research assistant Alexandra Rosiers.
Address: Department of Translation, Interpreting and Communication • Groot-Brittanniëlaan 45 • B-9000 Ghent • Belgium
E-mail: june.eyckmans@ugent.be

Stephanie Kader graduated from the School of Translation, Interpreting, Linguistics and Cultural Studies of the University of Mainz in Germersheim, Germany in 2004 and works as a freelance interpreter. She began teaching in the English Department in Germersheim in 2006, and since 2010 she has also taught in the Dutch Department. Her focus is on teaching conference interpreting from English and Dutch into German.
Address: FTSK / FB 06, Department of Dutch Studies • Johannes Gutenberg University Mainz • An der Hochschule 2 • 76726 Germersheim • Germany
E-mail: kader@uni-mainz.de

Judy A. Kegl obtained her doctorate in linguistics in 1985 from the Massachusetts Institute of Technology. She is a professor of linguistics at the University of Southern Maine, where she also directs the Signed Language Research Laboratory and coordinates the ASL/English Interpreting Concentration of the linguistics major. She has been a certified, practicing ASL/English interpreter since 1978 and has been an interpreter educator for over 16 years. She serves as Commissioner and Secretary of the Collegiate Commission on Interpreter Education (CCIE), the accrediting body for ASL/English interpreting programs. Her areas of research include ASL linguistics, neurolinguistics, and language emergence. She is best known for her discovery of an emerging signed language in Nicaragua, which she has been researching for almost 30 years.
Address: Linguistics Department • University of Southern Maine • P.O. Box 300 • Portland, ME 04104 • USA
E-mail: kegl@maine.edu

Brooke N. Macnamara received her PhD in cognitive psychology from Princeton University in 2014. She is a certified ASL-English interpreter and is currently a professor in the Department of Psychological Sciences at Case Western Reserve University.
Address: 10900 Euclid Avenue • Cleveland, OH 44106–7123 • USA
E-mail: brooke.macnamara@case.edu

Adam B. Moore received his PhD in cognitive psychology from Princeton University in 2011 and is currently a professor of psychology at the University of Edinburgh.
Address: 7 George Square • Edinburgh, Midlothian EH8 9JZ • United Kingdom
E-mail: amoore23@staffmail.ed.ac.uk

Jemina Napier is Professor and Chair of Intercultural Communication at Heriot-Watt University, where she teaches spoken and signed language interpreting students. She is also an Adjunct Professor at Macquarie University, Australia and an Honorary Research Associate at the University of Free State in Bloemfontein, South Africa. She practices as a signed language interpreter in several languages, and has published research on various aspects of signed language interpreting, interpreting pedagogy and interpreting research methods. She was the inaugural editor of the *International Journal of Interpreter Education*.

Address: Department of Languages & Intercultural Studies • Heriot-Watt University • Edinburgh • Scotland EH14 4AS • UK
E-mail: J.Napier@hw.ac.uk

Franz Pöchhacker is Associate Professor of Interpreting Studies at the University of Vienna. He has worked as a conference and media interpreter and published research on various domains of interpreting as well as general issues of interpreting studies as a discipline. He is the author of *Introducing Interpreting Studies* (Routledge, 2004) and co-editor of *The Interpreting Studies Reader* (Routledge, 2002) and of *Interpreting: International Journal of Research and Practice in Interpreting*.
Address: Center for Translation Studies • University of Vienna • Gymnasiumstrasse 50 • 1190 Vienna • Austria
E-mail: franz.poechhacker@univie.ac.at

Alexandra Rosiers studied Applied Linguistics at Erasmus University College in Brussels, where she obtained her Master's degree in translation in 2006 and a Master's degree in interpreting in 2007. In 2008 she started teaching at Erasmus University College and joined the interpreting research unit, focusing on psycholinguistic analyses of interpreters' and translators' attitudinal profiles. In 2012 she started working on a doctoral dissertation on individual difference variables among translators, interpreters and multilingual communicators in the Department of Translation, Interpreting and Communication of Ghent University.
Address: Department of Translation, Interpreting and Communication • Ghent University • Groot-Brittanniëlaan 45 • B-9000 Ghent • Belgium
E-mail: alexandra.rosiers@ugent.be

Mariachiara Russo graduated in Conference Interpreting from the SSLMIT of the University of Trieste in 1987 and is a freelance conference interpreter. In 1993 she became an Associate Professor at the SSLMIT of Trieste, teaching simultaneous and consecutive interpreting from Spanish into Italian. In 2001 she moved to the SSLMIT (now Department of Interpreting and Translation) of the University of Bologna at Forlì, where she also teaches interpreting theory.
Address: Department of Interpretation and Translation • Università di Bologna a Forlì • Corso della Repubblica 136 • 47100 Forlì • Italy
E-mail: mariachiara.russo@unibo.it

Heidi Salaets is head of the Interpreting Studies Research group at KU Leuven, campus Antwerp (Belgium), where she teaches dialogue interpreting and Interpreting Studies. She is also a research associate at the University of the Free State, Bloemfontein, South Africa. She trains conference interpreters at KU Leuven, campus Antwerp and community interpreters at institutional level, where she also assesses interpreters during certification exams. Her research concerns aptitude for conference interpreters, assessment in interpreting, interpreting in mental health settings, legal interpreting and training for interpreter users (service providers, legal personnel, etc.).
Address: KU Leuven, Faculty of Arts, campus Antwerp • Sint-Andriesstraat 2 • B-2000 Antwerp • Belgium
E-mail: heidi.salaets@arts.kuleuven.be

Sherry Shaw, EdD, CSC, is a Professor at the University of North Florida, USA, and Program Director of the Bachelor's and Master's degree programs in American Sign Language/English Interpreting. She has been an interpreter educator for 25 years and is actively engaged in research

that explores similarities and differences between signed and spoken language interpreting students. Her research interest since 2002 has centered on cognitive and dispositional characteristics of interpreting students, and she has also published in the areas of student support systems, transitioning from language acquisition to interpreting, and quality of life for Deaf seniors. She is author of *Service Learning in Interpreter Education: Strategies for Extending Student Involvement in the Deaf Community* (Gallaudet University Press), co-editor of the *Journal of Interpretation*, and reviewer for the *Journal of Community Engagement and Scholarship*. Sherry is a Commissioner on the Commission on Collegiate Interpreter Education, the accrediting body of post-secondary signed language interpreting programs in the USA.
Address: University of North Florida • 1 UNF Drive, Building 57/Room 3159 • Jacksonville, FL 32224 • USA
E-mail: sherry.shaw@unf.edu

Šárka Timarová is research assistant at KU Leuven, campus Antwerp (Belgium). Her research interests include cognitive aspects of the interpreting process, process research methodology and (mainly cognitive) aptitude for interpreting, with specific application to aptitude testing and admission screening in interpreting schools. She trained, and is professionally active, as a conference interpreter.
Address: KU Leuven, campus Antwerp • Translation and Journalism Studies • Sint-Andriesstraat 2 • B-2000 Antwerp • Belgium
E-mail: sarka.timarova@arts.kuleuven.be

Index