

Selection Tests for Trainee Conference Interpreters

David Gerver, Patricia E. Longley, John Long et Sylvie Lambert

Volume 34, numéro 4, décembre 1989

URI : <https://id.erudit.org/iderudit/002884ar>
DOI : <https://doi.org/10.7202/002884ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (imprimé)
1492-1421 (numérique)

[Découvrir la revue](#)

Citer cet article

Gerver, D., Longley, P. E., Long, J. & Lambert, S. (1989). Selection Tests for Trainee Conference Interpreters. *Meta*, 34(4), 724–735.
<https://doi.org/10.7202/002884ar>

Résumé de l'article

This preliminary study devised and tested a series of psychometric tests to facilitate selection of simultaneous and consecutive interpreter-trainee candidates. Twelve tests, based either on text materials, linguistic subskills or speed-stress, were correlated with judges' ratings of the final interpretation examination. Students who passed the exam had higher mean scores on all tests than those who failed. Test scores were positively inter-correlated between completion-detection tests and simultaneous interpretation ratings and between recall tests and consecutive examination ratings. Text-based tests were more predictive than subskills or speed-stress tests. The relationship between test type and interpreting is discussed together with possible reasons underlying differential performance on tests of the same kind.

Problems in the recruitment and training of conference interpreters have been the frequent subject of debate by members of the profession (Keiser 1978; Longley 1978; Namy 1978). Although there have been some publications on psychological and other theoretical aspects of the interpreter's skills (Barik 1973; Gerver 1976; Seleskovitch 1976; Karmiloff-Smith 1978; Moser 1978; Chernov 1979; Lambert 1983), there has been no systematic research on student selection and training in this field.

What qualities and skills are required for success as an interpreter or as a trainee in the field? Although no previous empirical research has been carried out on this subject, there does appear to be some consensus among interpreters and teachers of interpreting as to the skills and qualities sought in new members of the profession. A survey of articles written by members of the profession, as well as transcripts of interviews with working interpreters, suggested some agreement of the following as being essential for success as a trainee or in the profession.

1. Profound knowledge of active and passive languages and cultures.
2. Ability to grasp rapidly and convey the essential meaning of what is being said.
3. Ability to project information with confidence, coupled with a good voice.
4. Wide general knowledge and interests, and a willingness to acquire new information.
5. Ability to work as a member of a team.

The specific aim of the present study was to develop and evaluate objective tests to assess the second item mentioned above, that is the candidate's ability to grasp rapidly and to convey the meaning of spoken discourse. It was expected that the tests would contribute to improving selection procedures, hence the number of students passing the final examination, as well as the course itself.

The present study arose from the decision of Patricia Longley, one of the co-authors and former director of the Interpretation Programme at the Polytechnic of Central London, to initiate the establishment of objective criteria for the selection of students for the postgraduate course in conference interpreting techniques at the School of Languages in London. The course, founded by Patricia Longley in 1963, and now headed by Jennifer MacKintosh, is an intensive six-month course in simultaneous and consecutive interpretation. Over 200 applications are received each year from candidates throughout the world, about half of whom are invited to London for interviews and informal tests, lasting a day. Only 25 to 30 candidates are selected for the course.

The informal tests consist of written translation tests in addition to tests of both aural and oral language skills in the candidates' other languages. During the tests, candidates are first asked to repeat texts in the source language (shadowing) before progressing to interpretation of texts from their passive languages into their active language. These texts progress from simple sentences to more complex, albeit non-technical, longer passages. Unusual and unexpected phrases are inserted into the more difficult passages for interpretation in order to assess candidates' skills in coping with the unexpected and with passages which cannot be translated word-for-word. The interviews are designed to assess candidates' general background knowledge and interests, motivation, presentation of self, and general suitability for the profession. Whether or not there is a current market for the candidate's particular language combination is also taken into account at the time of selection.

The intensive, six-month course involves training in both simultaneous and consecutive interpretation, and in the note-taking skills essential to the latter form of interpreting, in conference practices, and in a variety of subjects likely to be encountered in conference settings.

The final examinations are oral in both forms of interpretation in the candidate's principal working languages, and are conducted by a board of external examiners who are themselves senior practising interpreters in major international organizations such as the U.N., Council of Europe, I.L.O., and the W.H.O.

A secondary aim of the study was to explore alternative types of tests. Since the task of the interpreter is to transform spoken textual information from one language into another, either immediately (for simultaneous interpretation) or after some delay (for consecutive interpretation), the first type of test - termed text-based - assumed the processing of connected discourse to be a crucial feature of the interpreter's task. The approach was based on recent work in the area of text processing (Kintsch 1974; MacKintosh 1985). Text-based tests required either the recall of the information presented or the completion of individual target words in the text. The second type of test - termed subskill-based - assumed language subskills such as synonym generation, sentence re-expression and vocabulary selection to be an adequate reflection of the interpreter's task. The approach was based on the development of a set of factor referenced cognitive tests of verbal ability (Eckstrom et al. 1976). In addition, both simultaneous and consecutive interpretation are performed under conditions of speed-stress, since the interpreter is paced by the speaker over whom he has little or no control. Hence, the third type of test - termed stress-based-assumed speed stress to be a general performance factor not particular to, but exemplified by, interpreting. The approach was based on the development of a speed test requiring the solution of problems involving letter series performed under a time constraint (Furneaux 1956). It was expected that the present research would throw some light on whether text-based, subskill-based or speed stress-based tests reflected the crucial features of the interpreter's task for the purpose of testing trainee candidates' abilities.

SELECTION TESTS FOR TRAINEE CONFERENCE INTERPRETERS

† DAVID GERVER*

University of Stirling, Stirling, Scotland

PATRICIA E. LONGLEY

PCL's School of Languages

JOHN LONG

MRC Applied Psychology Unit, Cambridge

SYLVIE LAMBERT

University of Ottawa, Ottawa, Canada

ABSTRACT

This preliminary study devised and tested a series of psychometric tests to facilitate selection of simultaneous and consecutive interpreter-trainee candidates. Twelve tests, based either on text materials, linguistic subskills or speed-stress, were correlated with judges' ratings of the final interpretation examination. Students who passed the exam had higher mean scores on all tests than those who failed. Test scores were positively inter-correlated between completion-detection tests and simultaneous interpretation ratings and between recall tests and consecutive examination ratings. Text-based tests were more predictive than subskills or speed-stress tests. The relationship between test type and interpreting is discussed together with possible reasons underlying differential performance on tests of the same kind.

Problems in the recruitment and training of conference interpreters have been the frequent subject of debate by members of the profession (Keiser 1978; Longley 1978; Namy 1978). Although there have been some publications on psychological and other theoretical aspects of the interpreter's skills (Barik 1973; Gerver 1976; Seleskovitch 1976; Karmiloff-Smith 1978; Moser 1978; Chernov 1979; Lambert 1983), there has been no systematic research on student selection and training in this field.

What qualities and skills are required for success as an interpreter or as a trainee in the field? Although no previous empirical research has been carried out on this subject, there does appear to be some consensus among interpreters and teachers of interpreting as to the skills and qualities sought in new members of the profession. A survey of articles written by members of the profession, as well as transcripts of interviews with working interpreters, suggested some agreement of the following as being essential for success as a trainee or in the profession.

1. Profound knowledge of active and passive languages and cultures.
2. Ability to grasp rapidly and convey the essential meaning of what is being said.
3. Ability to project information with confidence, coupled with a good voice.
4. Wide general knowledge and interests, and a willingness to acquire new information.
5. Ability to work as a member of a team.

The specific aim of the present study was to develop and evaluate objective tests to assess the second item mentioned above, that is the candidate's ability to grasp rapidly

and to convey the meaning of spoken discourse. It was expected that the tests would contribute to improving selection procedures, hence the number of students passing the final examination, as well as the course itself.

The present study arose from the decision of Patricia Longley, one of the co-authors and former director of the Interpretation Programme at the Polytechnic of Central London, to initiate the establishment of objective criteria for the selection of students for the postgraduate course in conference interpreting techniques at the School of Languages in London. The course, founded by Patricia Longley in 1963, and now headed by Jennifer MacKintosh, is an intensive six-month course in simultaneous and consecutive interpretation. Over 200 applications are received each year from candidates throughout the world, about half of whom are invited to London for interviews and informal tests, lasting a day. Only 25 to 30 candidates are selected for the course.

The informal tests consist of written translation tests in addition to tests of both aural and oral language skills in the candidates' other languages. During the tests, candidates are first asked to repeat texts in the source language (shadowing) before progressing to interpretation of texts from their passive languages into their active language. These texts progress from simple sentences to more complex, albeit non-technical, longer passages. Unusual and unexpected phrases are inserted into the more difficult passages for interpretation in order to assess candidates' skills in coping with the unexpected and with passages which cannot be translated word-for-word. The interviews are designed to assess candidates' general background knowledge and interests, motivation, presentation of self, and general suitability for the profession. Whether or not there is a current market for the candidate's particular language combination is also taken into account at the time of selection.

The intensive, six-month course involves training in both simultaneous and consecutive interpretation, and in the note-taking skills essential to the latter form of interpreting, in conference practices, and in a variety of subjects likely to be encountered in conference settings.

The final examinations are oral in both forms of interpretation in the candidate's principal working languages, and are conducted by a board of external examiners who are themselves senior practising interpreters in major international organizations such as the U.N., Council of Europe, I.L.O., and the W.H.O.

A secondary aim of the study was to explore alternative types of tests. Since the task of the interpreter is to transform spoken textual information from one language into another, either immediately (for simultaneous interpretation) or after some delay (for consecutive interpretation), the first type of test — termed text-based — assumed the processing of connected discourse to be a crucial feature of the interpreter's task. The approach was based on recent work in the area of text processing (Kintsch 1974; MacKintosh 1985). Text-based tests required either the recall of the information presented or the completion of individual target words in the text. The second type of test — termed subskill-based — assumed language subskills such as synonym generation, sentence re-expression and vocabulary selection to be an adequate reflection of the interpreter's task. The approach was based on the development of a set of factor referenced cognitive tests of verbal ability (Eckstrom *et al.* 1976). In addition, both simultaneous and consecutive interpretation are performed under conditions of speed-stress, since the interpreter is paced by the speaker over whom he has little or no control. Hence, the third type of test — termed stress-based — assumed speed stress to be a general performance factor not particular to, but exemplified by, interpreting. The approach was based on the development of a speed test requiring the solution of problems involving letter series performed under a time constraint (Furneaux 1956). It was expected that the present research

would throw some light on whether text-based, subskill-based or speed stress-based tests reflected the crucial features of the interpreter's task for the purpose of testing trainee candidates' abilities.

METHOD

SUBJECTS

Thirty students participated in the Polytechnic of Central London's graduate course in conference interpretation in the academic term from September 1977 to April 1978; of these, 21 students were female and 8, male, between the ages of 21 and 35 and all possessing university degrees or equivalent qualifications.

The majority of tests were taken by the students either as part of the selection procedure during April and May 1977 or during the first month of the course. In some cases, the testing was interrupted when the student was called for an interview. Therefore, not all tests were taken by all of the students. The results of the tests are therefore always reported with the number of candidates taking the test.

The passive languages (or languages one interprets from) of the candidates were Dutch, English, French, German, Russian, and Spanish, and the active languages (or languages interpreted into) included Dutch ($n = 3$), English ($n = 6$), French ($n = 7$), and Spanish ($n = 5$). Eight students interpreted bilingually from French into English and from English into French. Whatever students' active or passive language combination, all shared English and/or French as common active or passive languages. Therefore, for all students, English or French were considered to be appropriate languages in which to conduct the tests.

TESTS

Twelve tests of three different types were administered to the candidates. They were based either on textual materials (eight tests), linguistic subskills (three tests) or speed stress (one test). Since some of these tests may be of interest to interpreter-trainers, some detail will be provided here.

A) Text-based tests

There were two types of tests based on textual materials. The first involved the recall of auditorily presented information. The second involved the completion/detection of individual target words in auditorily presented text. There were two different tests of each type. The four are described separately below.

1) Recall tests-text memory

The text memory test was based on the assumption that interpreters need to understand and to remember the information conveyed by the discourse, that is to respond to the meaning of the speaker. A grasp of the informational structure of discourse is assumed to facilitate comprehension in the case of simultaneous interpretation and, in addition, memory in the case of consecutive interpretation. Research has shown that memory for text as demonstrated by recall protocols and summaries tends to reflect the gist of the text rather than the details (Van Dijk 1980). Further, the correlation for individual subjects between gist and detail or subsidiary information is both positive and significant (Long and Harding-Esch 1978).

Following a modified approach to that advocated by Kintsch (1974), two texts of 1000 words each were composed, consisting of eight sections in the manner of rhetorical discourse: 1) setting; 2) main issue; 3) arguments for; 4) secondary issue; 5) arguments against; 6) refutation; 7) irrelevant issue; and 8) conclusion. Both texts were in the form of speeches delivered at an international meeting, one on "fishing limits" (Text Memory 1) and the other on "industrial pollution" (Text Memory 2).

The texts were translated into the mother-tongue of the students on the course and recorded by native speakers. The texts were presented auditorily by means of a tape-recorder. The candidates were instructed to write an information summary and given five minutes to do so, following presentation of the text. The summaries were in English. The candidates were not allowed to take notes during the presentation. The protocols were scored on the basis of propositional points recalled from the text.

2) Recall tests—"logical" memory

Two short text memory tests were chosen from the Wechsler Memory Scale — termed "logical" memory (Wechsler 1945). Logical Memory 1 consisted of 65 words divided into 24 "memory" units. Logical Memory 2 consisted of 52 words divided into 23 "memory" units. Candidates' protocols were scored in terms of the number of "memory" units correctly recalled. "Memory" units were similar to the notion of propositional points used to score the longer texts.

3) Completion/detection tests-Cloze

"Cloze" tests require subjects to restore words deleted from text. They have been used increasingly as tests of second language proficiency, and are known to be a good index of general language proficiency (Stubbs and Tucker 1974). They are assumed to evaluate lexical, syntactic and semantic aspects of language processing. They are also assumed to measure "integrative" skills, that is the anticipatory processes involved when native speakers of a language engage in normal conversation both in listening and speaking (Oller and Conrad 1971). Thus the cloze test was chosen since it reflected text processing. Most research using the cloze technique has involved written texts presented visually. An auditory form was devised here because interpreters normally work from spoken discourse and because auditory cloze effectively paces the hearer in the manner of a conference speaker.

Three short cloze passages were devised. The first was in English (Cloze 1) and consisted of a passage of approximately 500 words. Every tenth word was deleted from the passage. The passage was recorded in English by a native speaker. The candidates were instructed to listen to the texts and to write down the missing words signalled by a tone. The test was scored by counting the total number of exact responses. Two 500-word passages were written, one in English (Cloze 2) and its translation in French (Cloze 3).

4) Completion/detection tests-error detection

As an alternative approach to assessing candidates' speed and accuracy of response to discourse a test was devised which required the candidates to detect errors in auditorily presented text. In contrast to the cloze test, this test explicitly evaluated responses to lexical, syntactic and semantic aspects of the text by including targets of each type. For example, a lexical error was defined as a non-word which might result from a slip in mispronunciation, such as "foat" for "coat," or "tember" for "member." A syntactic error was defined, in a limited sense, as one involving the alteration of the tense of a verb, as in "had" for "have," or "goes" for "went." Semantic errors involved the insertion of words different in meaning from those in the text, as in "Simmonds who lived in that 'plant' for ('street')." There were 16 errors of each type. It was assumed that candidates' sensitivity to the linguistic components of discourse and their ability to keep up with the presentation would be reflected by their ability to monitor the material and to detect the errors.

The passages were recorded in English by a native speaker. Candidates were instructed to identify the errors and write them down. They wrote only the correct version and were told to spell incorrectly pronounced words as they sounded. It was emphasized that it was not a spelling test. Candidates' protocols were scored in terms of the number of errors correctly recorded.

B) Subskill-based tests

Three tests were chosen and adapted from the group of "factor referenced" tests developed by the Educational Testing Service (Princeton) for research purposes (Eckstrom *et al.* 1976). Such tests had high face validity with respect to interpreting in as much as the cognitive and linguistic subskills which they embodied were assumed to be required by the interpreter's task. Like the completion/detection tests, the subskills tests were easy to administer and to mark. All the tests were in English. They are described separately below.

1) Associational fluency (Synonyms test)

This factor is assumed to involve the ability to produce words which share a given area of meaning or some other common semantic property. According to Carroll (1976), associational fluency entails a search of a major portion of long-term memory with special attention to its semantic and associational aspects. The test manual indicates that subjects score higher on this type of test if they have more associations tied to a word and more flexibility in interpreting similarity. The fluency entailed by this subskill would seem an important requirement for interpreting. The particular test chosen was the "Controlled Associations" or "Synonyms" test (Fa-1). The test was in two parts and candidates had six minutes for each part in which they were asked to write as many synonyms as possible (up to 12) to each of four stimulus words. The test was scored on the basis of the total number of "acceptable" responses.

2) Expressional fluency (Rewriting test)

According to the test manual, this factor is assumed to involve 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. The tests do not require the production of new ideas, but rather sentences that fit restrictions imposed in terms of letters, words or ideas. Like associational fluency, expressional fluency also includes a search of lexico-semantic memory but, as pointed out by Carroll (1976), with special attention to the grammatical features of lexical items and different syntactic patterns of phrases and sentences. Re-expression is basic to the interpreter's task. The particular test chosen was the Rewriting test (Fe-3) which was in two parts, each lasting five minutes, in which candidates were asked to rewrite each of three sentences in two different ways. The test was scored in terms of the percentage of the original sentences conveyed by the candidates' version. (In practice, the scoring of this test was found to be subjective and difficult to accomplish, since it involved equating of meaning at the sentence level.)

3) Verbal comprehension (Extended Range Vocabulary Test)

This factor involves the ability to understand the English language. According to Carroll (1976), verbal comprehension is almost exclusively dependent on the contents of the lexico-semantic long-term memory store. Comprehension is an essential feature of the interpreter's task. The particular test chosen was the Extended Range Vocabulary Test (V-3). This is a five-choice synonym test with items ranging from very easy to very difficult. The candidates had six minutes for each of the two parts of the test. Each part contained 24 items. The test was scored on the basis of the total number of correct choices.

C) Speed stress-based test (Nufferno test)

This measures the effect of speed stress on a cognitive but non-linguistic task. The Nufferno test (Furneaux 1956) was chosen as it claims to be a relatively pure measure of speed stress. Speed stress in general was assumed to be a major determinant of interpreter performance.

The test involves the solution of three sets of letter series problems (*e.g.* Z,X,V,T,R,___) and is based on the assumption that scores on tests of intellectual ability are determined by three relatively independent factors: speed, accuracy and continuance (persistence). A number of scoring procedures exist, but the score used here was the "stress-gain" score which was derived by subtracting candidates' score for an untimed test from the second score of a timed test, the first timed test considered as practice.

D) Examination ratings

Final examinations were oral and involved both simultaneous and consecutive interpretation. Students were rated out of 10 by each examiner independently (usually by three examiners) for each principal working language combination and for each form of interpretation. In cases of disagreement, the examiners discussed the ratings until agreement was reached. No assessment other than by examiners' ratings was made. A pass mark for the diploma was a rating of 5 out of 10. The agreed ratings for students who passed and failed are reported in the results section.

RESULTS

Means and standard deviations of scores on all the tests, together with means and standard deviations of judges' ratings in the final examinations in simultaneous and consecutive interpretations are shown in Table 1. Tests were carried out on the differences between mean scores for students who passed and students who failed the course.

Table 1.
Means and standard deviations for all test results and for judges' rating in the final examinations in simultaneous and consecutive interpretation for candidates who passed and those who failed the examinations
(n= number of candidates contributing to each result).

Test type	Name of test	X	Pass	n	X	Fail	n	t	p
			S.D.			S.D.			
Text	Text Memory 1	8.33	2.06	15	7.73	2.15	11	72	n.s.
	Text Memory 2	9.13	3.55	16	6.36	1.91	11	2.73	0.01
	Logical Memory 1	14.76	4.41	17	10.25	4.49	12	2.68	0.01
	Logical Memory 2	14.47	4.66	17	6.83	3.46	12	5.06	0.001
	Cloze 1	18.19	8.49	16	12.82	4.85	11	2.08	0.05
	Cloze 2	18.13	7.84	16	12.71	3.99	7	2.15	0.04
	Cloze 3	26.66	7.71	16	21.30	8.91	8	1.38	n.s.
	Error Detection	20.63	13.61	16	10.00	8.39	11	2.51	0.02
Subskills	Synonyms	17.54	5.91	13	10.83	4.12	6	2.86	0.01
	Rewriting	73.62	7.88	13	64.75	11.35	8	1.94	n.s.
	Extended Range								
	Vocabulary	68.35	21.31	17	59.10	15.81	10	1.29	n.s.
Speed	Nufferno	25.07	21.43	16	16.70	17.78	10	1.06	n.s.
Stress									
Examinations	Simultaneous	5.83	1.02	17	4.56	0.87	12	3.90	0.001
	Consecutive	6.03	0.97	17	4.17	0.54	12	6.65	0.001

The values of *t* and associated *P* values also appear in Table 1. As would be expected, judges' ratings of students passing the examinations, both simultaneous and consecutive, were significantly higher than those who failed (*P* 0.001). Only one of the tests attained this level of significance (Logical Memory 2). However, those who passed the course had higher scores on all 12 tests than those who failed. These differences were significant in the case of 7 tests: Text Memory 2; Logical Memory 1 and 2; Cloze 1 and 2; Error Detection; and Synonyms. Of the eight text-based tests, six were significant — three recall tests and three completion/detection tests. Only one of the subskill-based tests was significant.

The remaining non-significant tests were: Text Memory 1; Cloze 3; Rewriting and the Extended Range Vocabulary test. Dissimilar performance, then, sometimes occurred for tests of the same kind: for example, Text Memory 1 and 2 and subskill-based tests. The point is taken up below in the discussion section.

Table 2.
Correlation coefficients between the seven tests from Table 1
which discriminated significantly between students who passed and
those who failed the interpreter examinations.

Test	TM2	LM1	LM2	ED	Cl.1	Cl.2	Syn	Sim
Text Memory 2								
Logical Memory 1	0.42 *							
Logical Memory 2	0.31	0.78 ***						
Error Detection	0.16	0.23	0.30					
Cloze 1	0.05	0.26	0.34	0.83 ***				
Cloze 2	0.31	0.44 *	0.50 **	0.69 ***	0.69 ***			
Synonyms	0.07	0.65 **	0.61 **	0.65 **	0.69 **	0.82 ***		
Simultaneous	0.21	0.25	0.45	0.40 *	0.56 **	0.44 **	0.43	
Consecutive	0.35	0.48 *	0.63 ***	0.43 *	0.37	0.46 *	0.50 *	0.5
** <i>P</i>	0.01; *** <i>P</i> 0.001							

Correlation coefficients were calculated between scores on the seven tests which discriminated significantly between the groups, together with the final examination ratings, and these are shown in Table 2. Of the 72 possible pairwise correlations between different scores, only one comparison was found to be negative, and that was marginally so (Text Memory 2 and Synonyms: $r = 0.07$). Text Memory 2 was least associated with other scores (significant only in the case of Logical Memory 1 and was the only test not significantly related either to the simultaneous or consecutive interpreting examination ratings ($r = 0.21$ and $r = 0.35$, respectively). All the other tests were significantly associated with at least one the ratings, and no test was significantly associated with less than four other scores. Cloze 1 was significantly correlated with only the simultaneous ratings; Logical Memory 1 and Synonyms, only with the consecutive ratings.

Logical Memory 2, Cloze 2 and Error Detection were significantly associated with both sets of examination ratings. All the tests were positively intercorrelated and seven of the 15 possible pairwise correlations were significant. Of the 12 possible pairwise

correlations between text-based tests and judges, ratings of the simultaneous and consecutive examinations, all were positively correlated and eight were significant.

The pattern of correlation coefficients suggested a systematic relationship between subsets of the tests and two forms of interpreting. Although the smallness of the subject population was acknowledged, a principal components analysis was carried out on test scores and examination ratings to explore any possible relationship. After varimax rotation, two factors with eigen values greater than one were extracted, and accounted respectively for 78.8 and 21.2 per cent of their variance. These are shown in Table 3. The simultaneous ratings loaded more highly on factor 1, along with Error Detection, Cloze 1 and 2 and Synonyms. The Consecutive ratings loaded more highly on factor 2, along with Text Memory 2, Logical Memory 1 and 2. The test loading highest on factor 1 was Cloze 1 and on factor 2 was Memory 2. The test loading least on factor 1 was Text Memory 2 and on factor 2 was Cloze 1. The highest h^2 values were associated with Logical Memory 2 and Cloze 1. Of the text tests, all the recall tests loaded more highly on factor 2 than factor 1, and all the completion/detection tests loaded more highly on factor 1 than factor 2. These results can be summarized as follows:

- (i) Candidates passing the interpreter examinations scored higher on all the tests than those who failed. The differences were in many cases significant.
- (ii) Test scores were positively intercorrelated and positively associated with the judges' examination ratings. The correlations were in many cases significant.
- (iii) Of the two factors characterizing the pattern of correlations, factor 1 loaded highly on completion/detection tests and ratings of the simultaneous interpretation examination, while factor 2 loaded highly on recall tests and ratings of consecutive interpretation examinations.
- (iv) Of the subskills and speed stress tests, only Synonyms significantly reflected differences between candidates who passed and those who failed the examinations.

Table 3.
Rotated factor matrix derived from the seven tests from Table 1
which discriminated significantly between students who passed and
those who failed the interpreter examinations.

Test typ	Test	Factor 1	Factor 2	h^2
Text	Text Memory 2	0.04	0.43	0.19
	Logical Memory 1	0.23	0.75	0.62
	Logical Memory 2	0.26	0.94	0.95
	Error Detection	0.85	0.10	0.73
	Cloze 1	0.93	0.08	0.87
	Cloze 2	0.75	0.38	0.71
	Cloze 3	0.75	0.38	0.71
Subskills	Synonyms	0.74	0.43	0.73
Examinations	Simultaneous	0.48	0.33	0.33
	Consecutive	0.38	0.58	0.48
	Percentage variance	78.8	21.2	

DISCUSSION

Let us first consider the relation between the tests and the examination ratings in the context of the interpreter's task. Scores for the pass groups of students were higher than for the fail group on all 12 tests, and the differences were statistically significant for seven of the tests. In a general way, therefore, good performance on the tests was associated with a good showing in the interpretation 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. Further, many of the tests which were significant were also significantly related among themselves. This suggests that some abilities required by interpreters are reflected by more than one test. However, not all tests were equally significantly related to both simultaneous and consecutive examination ratings, and the two most important factors loaded differentially on the two sets of ratings.

Although the subject population was small for a factor analysis, it is, nevertheless, tempting to interpret the pattern of factor loadings on factor 1 as indicating a general verbal or linguistic factor reflecting the lexical, syntactic and semantic components of the Cloze, Error Detection and Synonyms tests. All three tests require speed and accuracy of linguistic response. Error Detection reflects primarily the perception of linguistic details of discourse; Synonyms reflects primarily a productive facility; and Cloze reflects general linguistic performance. It is consistent that simultaneous interpretation, which depends as much on the speed and accuracy of both perceptual and productive linguistic processes, should have loaded more highly on this factor than consecutive interpretation. In contrast, consecutive interpretation ratings loaded more highly on factor 2 and particularly concerning memory for text, both in terms of memory for the longer text (Text Memory 2) and for the shorter texts of Logical Memory 1 and 2. This may be assumed to reflect the more important function of reconstructive memory processes for consecutive rather than simultaneous interpretation. In summary, the tests appear to have been successful in reflecting generally the abilities required for interpreting. There is also a suggestion, which needs confirming by future research, that they also reflect in particular the subsets of abilities specific to the two modes of interpretation.

The specific aim of the study was to evaluate objective tests which were intended to assess candidates' ability to grasp rapidly and convey the meaning of spoken discourse. On the basis of the data reported here, the tests would generally seem appropriate as aids for candidate selection and some of the tests appear distinctly promising. However, in the case of the students tested, would the tests have improved the prediction power of the existing selection procedures? Assuming that selection for the course was based on the prediction that a student had a more than even chance of passing, the successful prediction rate for those students taking the test battery was 59 per cent; that is 17 out of 29 students passed the final examinations. In order to assess whether use of the test could have improved this rate of correct predictions, distribution of scores on four of the tests were analyzed: Text Memory 2, Logical Memory 2, Cloze 1 and Error Detection. Text Memory 2 and Logical Memory 2 were selected because both discriminated significantly between pass and fail groups, both had loadings on the second factor alone, and there was a higher level of significance between scores for passes and fails on Logical Memory 2 than on Logical Memory 1. Cloze 1 and Error Detection tests were chosen for this analysis on the basis of significant differences between scores for passes and fails, and because of their unique and high loadings on factor 1. The distribution of scores for pass and fail groups were inspected and appropriate cut-off scores were selected. The cut-off scores were chosen so as to optimize the number of passes and minimize the number of fails below the cut-off. Appropriate decision matrices were then constructed. They are given in Table 4.

Table 4.
 Matrices of decision-making accuracy for Test Memory 2, Logical Memory 2,
 Cloze 1 and Error Detection showing percentage total predictions correct
 (P c Tot) and the percentage pass predictions correct (P c Pos)

Predicted performance	Actual performance		
	Pass	Fail	Subjects
Text Memory 2, for a cut-off score of 7:			
Pass	12	4	16
Fail	4	7	11
Subjects	16	11	27
P c Tot = 70%	P c Pos = 75%		
Logical Memory 2, for a cut-off score of 10:			
Pass	16	2	18
Fail	1	10	11
Subjects	17	12	29
P c Tot = 90%	P c Pos = 94%		
Cloze 1, for a cut-off score of 12:			
Pass	12	7	19
Fail	4	4	8
Subjects	16	11	27
P c Tot = 59%	P c Pos = 75%		
Error Detection, for a cut-off score of 9:			
Pass	12	4	16
Fail	4	7	11
Subjects	16	11	27
P c Tot = 70%	P c Pos = 75%		

Though there are two measures of the accuracy of decisions made on the basis of test scores — percentage of total correct decisions, that is correct identification of both passes and fails on the basis of scores falling above or below the predetermined cut-off score (P c Tot), and correct identification of passes alone (P c Pos), the appropriate comparison in the present context would appear to be P c Pos rates, with the actual successful prediction rate of 59 per cent (17 out of 29 candidates). Inspection of table 4 shows that all of the selected tests improved on the actual successful prediction rate: Cloze 1, Text Memory 2 and Error Detection by 16 per cent, and Logical Memory 1 by 35 per cent. Given approximately 200 applicants per year of whom about 100 are interviewed, and of these 30 accepted for training, and a failure rate of about one-third, extrapolation from the present data suggests the tests might be a useful additional procedure for the selection of candidates. On the basis of the present data, the tests appearing in Table 4 would seem to be the most promising for predictive purposes.

Concerning the appropriateness of the different types of test, as opposed to the tests themselves, the present results suggest that the processing of connected discourse constitutes a crucial feature of the interpreter's task which needs to be embodied in selection tests. Only two of the eight text-based tests failed to discriminate significantly between pass and fail candidates. Only one of the subskill tests significantly discriminated between them. The importance of connected discourse is further emphasized by the poor showing of the speed stress-based test. It was presumably insensitive to the language processing aspects of interpreting. Interestingly, both the recall tests and the completion/detection tests successfully discriminated between pass and fail candidates but loaded differentially on the two factors. This suggests that completion/detection tests may be more appropriate for consecutive interpretation. However, since most candidates are trained in both types of interpretation, whether or not they subsequently practise both, the two kinds of tests would seem to be necessary rather than constitute alternatives.

Concerning the recall tests of Text Memory 1 and 2, consistency was not helped by the relatively poor performance by both pass and fail groups. No group attained a mean score of 10 (i.e., 25 per cent) of the 40 possible points made in the speeches. The tests were too difficult to do well and therefore insufficiently sensitive. Of the two texts, however, the one on "pollution" was the more sensitive, producing a larger difference score between pass and fail groups. Examination of the "pollution" text suggested it might have given more opportunity than "fishing" for the active selection among points as well as the making of connections between points. A greater selection requirement would render the texts more similar to those used by Van Dijk (1980) and Long and Harding-Esch (1978) and might be expected to emphasize the ability to extract crucial information from a text rather than the ability to remember a large number of more or less equally important points constituting practically the whole body of a text. Emphasizing the gist content would render the texts less similar to the Wechsler logical memory tests, which by dint of their shortness, are almost verbatim recall tests. The effect might be to reduce the association between the two types of test although not necessarily to modify their common contribution to factor 2 (Table 3), particularly if this factor is more closely associated with consecutive than with simultaneous interpretation, as hypothesized earlier. The possible reduction in association between the Text Memory and Logical Memory tests is consistent with the lesser association between the "pollution" text and the apparently more sensitive second Logical Memory test.

Concerning speed-stress tests, the Nufferno stress-gain score failed to discriminate pass and fail groups. However, there can be no doubt that the speaker-paced nature of simultaneous interpretation involves speed stress of the classic kind (Conrad 1954). Since the auditorily presented Cloze tests and Error Detection test significantly discriminated pass and fail students, and since these tests involved external pacing and hence speed stress, it suggests that the particular stress experienced by simultaneous interpreters needs to be evaluated with linguistic materials rather than general purpose materials. This suggestion is consistent with the test results. It is also consistent with the view that the allocation of cognitive resources involves both general purpose and special purpose mechanisms (LaBerge 1981). Stress may interact with linguistic proficiency in a way which nullifies any direct correlation. It would be of some considerable interest to replicate the positive and negative stress effects reported here.

ACKNOWLEDGEMENTS

This project was funded by the Polytechnic of Central London.

Note

- * It is with great regret that we record here the sudden death of David Gerver. This work was initiated by him and carried out while he was at the Department of Psychology, University of Stirling, Stirling, Scotland. In completing this paper, the co-authors would like it to stand as a tribute both to him and his work.

REFERENCES

- BARIK, H.C. (1973): "Simultaneous Interpretation: Temporal and Quantitative Data," *Language and Speech*, 16, pp. 237-270.
- BROWN, J. (1976): *Recall and Recognition*. London, Wiley.
- CARROLL, J.B. (1976): "Psychometric Tests as Cognitive Tasks: A New 'Structure of Intellect'," in L.B. Resnick, *The Nature of Intelligence*, New York, Erlbaum.
- CARROLL, J.B. (1978): "Linguistic Abilities Translators and Interpreters," in D. Gerver and H.W. Sinaiko (eds.), *Language, Interpretation and Communication*, New York, Plenum.
- CHERNOV, G.V. (1979): "Semantic Aspects of Psycholinguistic Research in Simultaneous Interpretation," *Language and Speech*, 22, pp. 227-296.
- CONRAD, R. (1954): "Speed Stress," in W.F. Floyd and A.P. Welford (eds.), *Human Factors in Equipment Design*, London, H.K. Lewis.
- ECKSTROM, R.B., J.W. FRENCH, H.H. HARMAN and D. DERMEN (1976): *Manual for Kit of Factor-Referenced Cognitive Tests*, Princeton, N.J., Educational Testing Service.
- FURNEAUX, W.D. (1956): *Manual of Nufferno Speed Tests and Manual of Nufferno Level Tests*, London, Institute of Psychiatry.
- GERVER, D. (1976): "Empirical Studies of Simultaneous Interpretation: A Review and a Model," in R. Brislin (ed.), *Translation: Applications and Research*, New York, Gardner.
- KARMILOFF-SMITH, A. (1978): "Adult Simultaneous Interpretation: A Functional Analysis of Linguistic Categories and a Comparison with Child Development," in D. Gerver and H.W. Sinaiko (eds.), *Language, Interpretation and Communication*, New York, Plenum.
- KEISER, W. (1978): "Selection and Training of Conference Interpreters," in D. Gerver and H.W. Sinaiko (eds.), *Language, Interpretation and Communication*, New York, Plenum.
- KINTSCH, W. (1974): *Representation of Meaning in Memory*, Hillsdale, N.J., Erlbaum.
- LABERGE, D. (1981): "Automatic Information Processing: A Review," in J. Long and A.D. Baddeley (eds.), *Attention and Performance*, IX, Hillsdale, N.J., Erlbaum.
- LAMBERT, S.M. (1983): "Recall and Recognition among Conference Interpretation," Unpubl. doctoral dissertation, University of Stirling, Scotland.
- LONG, J. and E. HARDING-ESCH (1978): "Summary and Recall of Texts in First and Second Languages: Some Factors Contributing to Performance Differences," in D. Gerver and H.W. Sinaiko (eds.), *Language, Interpretation and Communication*, New York, Plenum.
- LONG, J. and E. HARDING-ESCH (in preparation): *Levels of Second Language Deficit*.
- LONGLEY, P.E. (1978): "An Integrated Programme for Training Interpreters," in D. Gerver and H.W. Sinaiko (eds.), *Language, Interpretation and Communication*, New York, Plenum.
- LYON, P. (1966): *French Short Stories*, vol. 1, London, Penguin.
- MOSER, B. (1978): "Simultaneous Interpretation: A Hypothetical Model and its Practical Application," in D. Gerver and H.W. Sinaiko (eds.), *Language, Interpretation and Communication*, New York, Plenum.
- NAMY, C. (1978): "Reflections on the Training of Simultaneous Interpreters. A Metalinguistic Approach," in D. Gerver and H.W. Sinaiko (eds.), *Language, Interpretation and Communication*, New York, Plenum.
- OLLER, J.W. Jr. and C. CONRAD (1971): "The Cloze Technique and ESL Proficiency," *Language Learning*, 21, pp. 183-196.
- PANSINI, F. (1979): *Interpretation in a Multilingual Parliament*. Report issued by the European Parliament.
- SELESKOVITCH, D. (1976): "Traduire: de l'expérience aux concepts," *Études de linguistique appliquée*, 24, Paris, Didier.
- STUBBS, J.B. and G.R. TUCKER (1974): "The Cloze Test as a Measure of English Proficiency," *Modern Language Journal*, 58, pp. 239-241.
- VAN DIJK, T. (1980): *Macrostructures*, Hillsdale, N.J., Erlbaum.
- WECHSLER, D. (1954): "A Standardized Memory Scale for Clinical Use," *Journal of Psychology*, 19, pp. 87-95.