Widdowson's model of communicative competence and the testing of reading: an exploratory study

G. Fulcher *

English Language Institute, University of Surrey, Guildford, Surrey GU2 5XH, UK

Abstract

This paper considers the validity of Widdowson's discourse model of communicative competence and performance (Widdowson, H., 1983. Learning Purpose and Language Use. Oxford University Press, Oxford) as the basis for developing tests of reading. The article describes the basic structure of the model, and attempts to add to the original description in such a way that it is possible to design an operational test of the model components. The purpose of the operational test is only to investigate the adequacy of the model as a theoretical basis for test design, but successful model validation is potentially very important for the interpretation of test scores. The results suggest that there is some empirical support for the position that one level of the model is valid for the stated purpose. © 1998 Elsevier Science Ltd. All rights reserved.

1. Introduction

The role of models of language, competence, and performance, in applied linguistic research has been well documented in recent years (Skehan, 1995; McNamara, 1995, 1996; Chalhoub-Deville, 1997), and yet there remains a paucity of research into the validity of these models in relation to their stated purpose. Writing of the Bachman (1990) model, Alderson (1997, p. 5) says:

The advantage of having an explicit model like the Bachman model is that it can be studied, criticised, its implications understood, its assumptions questioned, and it can then be improved. However, a model needs to be empirically tested: not just criticised speculatively. Bachman's model has been much referred to, but little operationalised, to date.

*Tel.: +44-1483-259910; fax: +44-1483-259507; e-mail: eli@surrey.ac.uk
This lack of operationalisation of models in general may be one reason why all models are necessarily transitory and incomplete, as observed by Alderson (1997). Theorists continue to develop new models without initiating empirical studies to investigate them, or their predecessors. It is for this reason that Chalhoub-Deville (1997, pp. 6-7) can legitimately claim that most models are descriptive; the task facing the model developer has been one of definition rather than empirical research and model validation.

This article argues that it is important to conduct research into the validity of models through the operationalisation of theory in language tests. It then reports on a study designed to operationalise an interpretation of the Widdowson model expounded most fully in his 1983 work, Learning Purpose and Language Use. The purpose of the study described in the paper is to empirically explore the usefulness of the model in the testing of reading. The specific research question addressed is whether the interpretation of the model presented in this paper is a valid basis for the design of reading tests.

2. The need for model validation studies

Models are important for the validation of language tests, in at least two ways. Firstly, a model may provide a framework for writing the test specifications that form the blueprint for a test (Alderson et al., 1995, p. 14); secondly, a model may offer test researchers a way of interpreting test scores in terms of competence, ability to perform or performance. That is, the model should provide the mechanism by which the inferences made from test scores can be traced back to the constructs the test was designed to measure (Messick, 1989). Yet, empirical studies of models are not frequently carried out. Indeed, models that have not been studied empirically are often used to conduct other studies (Bachman et al., 1995; Kunnan, 1995). It would not be unfair to claim that since the work of Oller (1979; see also Oller and Perkins, 1980), the only certainty to arise from empirical studies is that language competence is componential (Volmer, 1983).

It is therefore argued that a key part of the notion of construct validity, defined as "an integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores" (Messick, 1989, p. 13), is to establish the extent to which a test measures a component or trait as defined in a construct in the theoretical model. If this premise is accepted, it is essential to investigate the validity of models used in test design and construction.

There are a further two important reasons for an early start to the study of model validity. Anastasi (1988, p. 153) points out that "construct validation requires the gradual accumulation of information from a variety of sources." Information must be collected and compiled over a period of time, whilst the need for evidence of construct validity in many testing situations is urgent. Secondly, until models are empirically investigated, there is a danger that tests designed on the basis of under-defined models may lead to construct under-representation in the test, and this represents a major threat to test validity.
Just as Bachman (1988, p. 163) warned that having had years of experience working with a scale "in no way constitutes evidence for validity", it is appropriate to remind ourselves that years of familiarity with a model or family of models does not mean that they have any empirical foundations or psychological reality.

3. Widdowson's discourse model

3.1. A three tier model

Widdowson's (1983) model is primarily concerned with the theoretical background to teaching English for Specific Purposes, but it is a discourse model that is firmly based in schema theory. Widdowson (1983, pp. 34–35) describes schemata as:

... cognitive constructs which allow for the organisation of information in long term memory and which provide a basis for prediction. They are kinds of stereotypic images which we map onto actuality in order to make sense of it, and to provide it with a coherent pattern.

Schemata are part of communicative competence, or Hymes' "ability for use", which Widdowson distinguishes sharply from linguistic competence. The latter, the systemic level of the model, is considered to be a "second order abstraction", which fulfils a supportive role in actual language use.

To account for actual language use (performance), a third tier of the model is introduced. Procedural capacity is a model of the processes involved when a reader or speaker must negotiate or co-construct meaning because the schematic world of the text or the interlocutor is not the same as their own. The greater the difference between worlds, the more procedural work will be required to reach understanding.

The model therefore contains three levels:

- systemic level (linguistic competence);
- schematic level (ability for use, or communicative capacity); and
- procedural level (actual performance).

Before progressing to a more delicate definition of the model, and the interpretation of it provided in this paper, it is informative to look briefly at the relationship of this model to that of Hymes. Hymes (1972) is universally credited with expanding the notion of competence to include the ability for use—underlying traits that enable communication—in his definition of communicative competence. These abilities or traits, whilst not described in detail, clearly included both cognitive and metacognitive variables. Both aspects of communicative competence (knowledge and ability) were distinguished from performance, or actual use of language in communication. This tripartite division is reflected in the Widdowson model, although there is no place in Widdowson's structure for metacognitive variables. Widdowson reserves the term "communicative competence" for linguistic knowledge, and introduces the term "communicative capacity" or "procedural knowledge" for Hymes' ability for
use. It is this capacity that allows the speaker to actually communicate. In this respect, the Widdowson model is perhaps one of the closest in structure to Hymes' original insights than some other post-Hymes variants. This brief detour shows that the Widdowson model, although not frequently discussed in the language testing literature, is mainstream in applied linguistics, and deserves further attention by language testers.

In what follows, we are primarily concerned with operationalising the schematic and procedural levels for the purposes of testing reading. In order to do this, it is necessary to adequately define these levels.

3.2. The schematic level: ability for use

Widdowson distinguishes between two types of schemata, the "ideational", which are related to conceptual organisation, and "interpersonal", which are related to "patterns of participations in social life" (Widdowson, 1983, pp. 55–56). This division is familiar from the work of Halliday (1979). Widdowson associates ideational schemata with "frames", and interpersonal schemata with "plans" or "scripts". The former relates to conventional knowledge (a "restaurant frame" is the knowledge that it is a building for eating, that there are waiters and waitresses, that you pay for the food), whilst the latter are idealised, predictable routines in speech acts (asking for the menu, ordering, asking for the bill). Widdowson therefore refers to these as schematic routines.

As a theoretician, he did not develop the categories further, failing to provide anything that would allow language testers to use the model operationally, beyond the suggestion that the work of Eugene Winter and Michael Hoey may provide a description of "a basic rhetorical routine" (situation–problem–solution–evaluation) that could be used to describe interpersonal schemata (Widdowson, 1983, pp. 57–58).

It can be seen that textual organisation and the reader's ability to understand textual organisation is related by Widdowson not to ideational schemata but to routine schemata. That is, textual organisation and its recognition belongs to the interpersonal level. Initially, this may strike us as somewhat unorthodox, but the claim appears to be that it is textual organisation that sets up the relationship between text and reader. Hoey (1983a, b) refers to culturally popular patterns of organisation, such as problem–solution patterns, situation–evaluation patterns, and hypothetical–real patterns (Winter, 1978, 1979, 1986), as unmarked forms of discourse organisation. It is in these unmarked forms that we find the basic relationships that are set up between reader and text.

Finally, Widdowson used the term "scenario" to represent a combination of the two types of schematic knowledge. A scenario is defined as "a network of conceptual associations tied in with a co-occurring sequence of acts constituting a conventional episode or event" (Widdowson, 1983, p. 57). In this definition, we see the formulation of a unified concept of schemata incorporating frames and routines which, together, constitute the notion of communicative capacity, or the "ability for use".

Having outlined the notion of communicative capacity, and placed it within a model, it is nevertheless still necessary to consider the significance of the concept in
relation to our understanding of the reading process, and how it could inform the
design of a test of reading or the interpretation of scores from a test of reading.
Firstly, it is claimed that a reader establishes a model of the text that is being read by
reference to existing text scenarios. Secondly, but closely related to the first point,
the closer the model of the text to an existing text scenario, the better able the reader
is to anticipate or predict what is to come next in the text. The more aspects of the
scenario that are recoverable, the easier the text will become for a reader to com-
prehend (Fulcher, 1997). However, it is at the procedural level that the model deals
specifically with performance issues, or how the reader establishes a mental repre-
sentation of the text and is able to predict what is coming next in the text.

3.3. The procedural level

The procedural level is significantly different from the schematic level, in that it
concerns the ways in which a reader goes about interpreting the schemata of the
text. It is concerned with the on-line performance of competence. Widdowson dis-
tinguishes between “frame” and “routine” procedures, in the same way that he
distinguishes between “frame” and “routine” schemata. The procedural level repre-
sents an attempt to account for performance as well as the ability to perform. Frame
procedures are said to explain the ability to utilise frame schemata, and rou-
tine procedures are said to explain the ability to utilise routine schemata.

Frame procedures are defined as those that establish and maintain reference,
especially with regard to cohesion, and working out “given” from the “new” infor-
mation (Widdowson, 1983, pp. 41–42, 67). Through these procedures the reader fits
new information into a frame. The procedures involve tracing cohesive links and
understanding the relationship of information as belonging to a specific frame,
which is the engagement of schematic knowledge in processing text meaning.

Cohesion is the central concern of maintaining frame of reference, involving
reference, substitution, ellipsis, conjunction and lexical cohesion, as well as exo-
It may appear that these traditional categories of cohesion have little to do with the
process of comprehension in reading, but as Hasan (1984) has cogently argued, there
is a strong connection between cohesion and making sense of coherent text.

Routine procedures are concerned primarily with illocutionary intent, as described
by philosophers of language such as Austin and Searle, and the linguists Grice and
Labov. Associated with understanding interpersonal schemata, routine procedures
are also concerned with how readers make predictions and inferences. Schemata are
about product, whilst procedures, particularly routine procedures, are about pro-
cess. “Comprehension” in reading is an understanding of the illocutionary force of
the text, grasped through a dialogue with the text. Uncovering the reader’s dialogue
with the text is the researcher’s attempt to break into the on-line processes.
Widdowson (1979, p. 60) provides a rare illustration of what he means by this:

The Greek revolution deserves its fame. It is unique in the annals of mankind.
What makes it unique is precisely the directed efforts, the continued and
systematic modifications of the schematic of conceptual art, till making was replaced by the matching of reality through the new skill of mimesis.

This can be derived from the following discourse:

The Greek revolution deserves its fame.
Why?
It is unique among the annals of mankind.
In what way unique?
What makes it unique is precisely the directed efforts, the continued modifications of the schematic of conceptual art, till making was replaced by the matching of reality through the new skill of mimesis.

Widdowson believes this example to reveal the discourse that underlies the text in terms of the illocutionary intention of the writer, text being a product of discourse. What is being demonstrated in this rather simplistic example is the way in which reader and text interact in the reading process, as described in more detail by Hoey (1983a). Questions can be used to clarify clause relations or some aspect of the schematic patterning, but Hoey (1983a, p. 170) also argues that “they are intended to reflect a crucial aspect of the interpretative process.”

Crucially for the testing of reading, Hoey continues:

... it is suggested that on the basis of (1) cultural and linguistic expectations about the type of discourse encountered and (2) what the writer/speaker has already said (including the title and place of occurrence), a reader/listener hazards guesses as to the content to come and its relationship to what has preceded. In so far as they guess correctly, they have a smooth ride; in so far as they guess wrongly, their comprehension is slowed down to some extent. If they consistently guess wrongly, it can be doubted whether they properly comprehend at all, though the fault may lie with either encoder or decoder.

The work of Hoey may be compared directly with that of Rumelhart, especially in the technique of revealing text line by line, with the intention of tapping on-line real-time comprehension (Rumelhart, 1984, pp. 6–13). In his work, Rumelhart argues that bottom-up and top-down processing interact in a complex mechanism of triggering and anticipation. However, Rumelhart deliberately removes signals from the texts, using full lexical items in his investigation of the role of knowledge of the world in prediction, whilst Hoey looks at the impact on prediction of precise linguistic phenomena. The two aspects of the investigation are entirely complementary. Rumelhart is investigating frame procedures through lexical and exophoric cohesion, whilst Hoey is looking at routine procedures through prediction triggered by clause relational signals.

In Hoey’s work, these precise linguistic phenomena relate to the interpretation of text organisation at the level of interpersonal schemata. The ability to process matching and logical sequence relationships, for example, shows us that routine
procedures are activated by linguistic input in the text, and how the reader creates a more accurate model of the text through the process.

In order to understand procedural routines, we will very briefly outline the concepts of matching and logical sequence relationships.

These relationships are said to be categories of clause relations, the latter being described as "the cognitive process whereby we interpret the meaning of a sentence or group of sentences in the light of its adjoining sentences or group of sentences" (Hoey, 1983a, p. 18). In matching relations, information is "matched" against other information to indicate that it is similar, or that it is contrastive. Hoey (1983a, p. 20) provides the following examples of matching compatibility and matching contrast, respectively:

Upstairs Fred thumped and bumped and tossed and turned. And downstairs Ted moaned and groaned and crashed and thrashed all over the bed.

The skirmish was not taken seriously at Vienne. What was taken seriously, however, was the fact that Constantius had named me his fellow consul for the New Year.

In the first sentence Fred and Ted are "matched" for their similar behaviour, whilst in the second example there is a contrast between the seriousness of the two events.

Logical sequence relationships include indicators of time (... next, they investigated the blue door), condition–consequence (If the computers are left on overnight, the students will have access first thing in the morning), instrument–achievement (She passed the module by submitting an assignment she bought from an Internet store), and cause–consequence relations (They were ill because they had fallen into the icy water). Logical sequence relations can be indicated through the use of subordinators, conjuncts, lexis, repetition, or paraphrase. These matching and logical sequence relations are the precise phenomena that the reader works on in order to comprehend text.

In Section 4, we turn our attention to the design of a reading test that attempts to reflect this interpretation of the Widdowson model in its structure. That is, there has been an overt attempt to operationalise the model for the purpose of exploratory investigation.

4. Operationalising the model

Once the elements of a model have been described, it is necessary to operationalise them in test tasks. This is one of the most difficult processes that is undertaken in this type of research, because the link between the theoretical description of the model and the test task relies, to a greater or lesser degree, on the interpretation of the task designer. In Table 1, we present the format of the test that was designed for this particular study. It is made up of a number of sub-tests, each of which is designed to measure a particular component of the model. The component is listed
### Table 1

<table>
<thead>
<tr>
<th>Level</th>
<th>Item types</th>
<th>Number of items</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Part 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systemic 1</td>
<td>Matching sentence halves</td>
<td>10</td>
</tr>
<tr>
<td>Schematic 1(A)</td>
<td>Matching titles to stories</td>
<td>10</td>
</tr>
<tr>
<td>Schematic 1(B)</td>
<td>Open-ended questions</td>
<td>5</td>
</tr>
<tr>
<td>Procedural 1(A)</td>
<td>Grid completion</td>
<td>10</td>
</tr>
<tr>
<td>Procedural 1(B)</td>
<td>Within text M/C</td>
<td>10</td>
</tr>
<tr>
<td><strong>Part 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systemic 2</td>
<td>Jumbled sentences</td>
<td>10</td>
</tr>
<tr>
<td>Schematic 2(A)</td>
<td>Intruding sentences</td>
<td>10</td>
</tr>
<tr>
<td>Schematic 2(B)</td>
<td>Grid completion</td>
<td>5</td>
</tr>
<tr>
<td>Procedural 2(A)</td>
<td>Cloze</td>
<td>10</td>
</tr>
<tr>
<td>Procedural 2(B)</td>
<td>Jumbled text</td>
<td>10</td>
</tr>
</tbody>
</table>

In the left-hand column, the item types selected are listed in the middle column, and the number of items in that sub-test is given in the right-hand column.

The left-hand column of Table 1 requires further explanation. The sub-tests are separated into two halves, labelled “Part 1” and “Part 2”. This is because the test contains two sub-tests of each component (with the exception of the systemic level). Thus, Schematic 1A and 2A were designed to measure frame schemata, and Schematic 1B and 2B were designed to measure routine schemata. Similarly, Procedural 1A and 2A were designed to measure frame procedures, and Procedural 2A and 2B were designed to measure routine procedures. Finally, there are two systemic sub-tests which were designed to tap a single aspect of linguistic knowledge: word order within sentences.

The research described here depends to a large degree on the extent to which the operational sub-tests can be claimed to reflect the theoretical components of the model. Appendix A contains the sub-test Schematic 2A, which was designed to measure frame schemata. The item type is termed “intruding sentences”. A text on the topic of pollution in Rome is presented with each sentence beginning on a new line. However, a number of sentences from other texts are placed in between the sentences of the complete text. The task is to identify the intruding sentences. Sentence (a) introduces the topic of pollution on the streets of Rome with an example of pollution, and in Sentence (b) “air pollution” is given as the topic of the rest of the passage. Sentence (c) is cohesive because of the repetition of the word “pollution” from the end of Sentence (b), and there is lexical cohesion between Rome–Europe–France–Germany–Netherlands. Sentence (c) is not only cohesive, it also contains an internal cause–consequence relationship that appears to fit easily into the developing description of the seriousness of pollution. However, the test taker must recognise that “waterway” does not fit the frame that has been established in the first two sentences; the second clue lies in the phrase “as well as the Netherlands”, which
should activate geographic schemata. Each of the intruding sentences in this test is cohesive with the sentences before and after it to some degree, but does not comply with the expected frame of reference. These are examples of sentences that are generally cohesive but not coherent.

Appendix B contains sub-test Schematic 1B, which was designed to measure interpersonal (routine) schemata. The item type is the short open-ended response, in which the test taker is required to make a list of 6 problems, followed by the solutions offered to these problems by four organisations. This is based upon a text which has a clear iterative problem–solution–evaluation structure. The test takers must identify the problems and solutions signalled in the text. Take for example the problem which could be listed as "increase in chances of death from cirrhosis", which is explicitly signalled as a "health problem", and also by the second part of the previous sentence: "...but it's been very bad for us." I would contend that this item type, used with a text that has a culturally popular pattern of organisation, does adequately represent an operationalisation of the model concept, as currently formulated.

An example of a sub-test designed to tap frame procedures is provided in Appendix C. Frame procedures are concerned, as we have seen, with the maintenance of frames of reference through cohesion. The text selected for the sub-test describes a process, and topic is suitably general. One weakness of the text is that although it is ideal for testing pronominal reference, it is difficult to use it to test for other types of reference. In the first sentence, it is necessary for the test taker to identify "it" respectively with: the ready-filleted fish (anaphoric), washing in cold water (cataphoric), and the ready-filleted fish (anaphoric). A failure to be able to understand this string of references would cause a reader considerable difficulty in maintaining the frame of reference, and comprehension would suffer considerably.

Routine procedures are hypothesised to be the highest level of processing, consisting of prediction, understanding illocutionary intent, and clause relational meaning. Cooper (1984, pp. 131–132) conducted a study in which prediction, as signalled by matching, contrast and logical sequence relationships, was found to be the most accurate discriminator between practised and unpractised readers. It was for this reason that Cooper considered this to be one of the highest levels of discourse processing, and this is supported theoretically in the operationalisation of the Widdowson model as described here.

An example of a sub-test designed to measure routine procedures is provided in Appendix D. The sub-test is a sentence level cloze, where the items consist of the original sentence as the key and a number of alternatives as distracters. Once again, we will consider an example. The correct answer to the first item in the text entitled "Philosophy" is 1b. How would a test taker come to the conclusion that this answer is correct? Sentence 1b stands in a cause–consequence relation to the first sentence of the text. The text asserts that humans are curious, and therefore humans ask questions, wonder and speculate. What do we ask questions, wonder or speculate about? The answer to this question comes in the third sentence. The conjunction therefore is the key to selecting the correct answer, as it helps us to make sense of the relationship between the first and third sentences.
Each of the items in this sub-test measures the degree to which the test taker is capable of decoding the features of discourse that enable us to understand clause relations, and hence comprehend the meaning of sentences and discourse.

We have now considered, in some detail, the rationale for the set of sub-tests, and the item types. It has been argued that these sub-tests are effective operationalisations of the model. What remains to be done is to empirically investigate this operationalisation in order to comment on the validity of the model on which it was constructed.

5. **Method**

In this research a test was constructed with three levels: systemic, schematic and procedural. The content of each part of a test is determined by the three tier model of Widdowson, as interpreted in this paper. The item types and content of the sub-tests were carefully considered, and the rationale for item types explained earlier.

The test was given to a group of 121 students drawn from the upper form of six secondary schools in Cyprus. The mean age of the students was 16.6 years. Test taking was completed in 4 days. The sample is presented in Table 2. It is divided (here as in Cooper, 1984) into two groups. Practised readers are those who have received their secondary education in English from the age of 11 years or earlier, and unpractised readers are those who have undergone primary language education since the age of 11 years, but have at least 8 hours of English tuition per week. The students in the sample spoke either Greek or Armenian as their primary language, and this information is also provided in Table 2.

In order to answer the main research question regarding the validity of the model, it is necessary to collect a range of evidence.

The first piece of evidence relates to the reliability of the test and the sub-tests. This comes in two parts. We wish to know if the test as a whole is reliable and able to discriminate between experienced and inexperienced readers, as defined already. If the test as a whole is reliable and capable of discriminating between students who have been classified as experienced and inexperienced readers on independent grounds, it can be argued that we may rely on the scores as measures of something, and that these scores are sensitive to reading ability. The reliability of the sub-tests is important because it is the relationships between sub-tests that enable us to decide whether, and to what extent, the sub-tests are measuring what they claim to measure. If a sub-test has a low reliability, any conclusions based on correlations between that sub-test and other sub-tests must be extremely tentative.

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample characteristics</td>
</tr>
<tr>
<td>Practised readers</td>
</tr>
<tr>
<td>74 (61.16%)</td>
</tr>
<tr>
<td>Greek Cypriot</td>
</tr>
<tr>
<td>106 (87.6%)</td>
</tr>
</tbody>
</table>
The second piece of evidence that is offered in this study is the correlation matrix of the sub-tests. The principle of convergent validity demands that there is a high and significant correlation between sub-tests that are designed to measure the same construct (Schematics 1 and 2, for example). The principle of divergent validity demands that the correlation between sub-tests that are designed to measure different constructs (Schematic 1 and Procedural 1, for example) will be lower. In a multitrait-multimethod study, it would also be possible to look at the impact of test method on the correlation matrix, but in this study there is no systematic use of the same method across traits. Although this is a disadvantage, it is not a fatal flaw in a methodology designed to be exploratory.

6. Results and discussion

6.1. Reliability

The mean score on the test total was 52.98 (± 2.42), with a standard deviation of 13.61. Reliability was estimated for the whole test and each sub-test, using Cronbach’s alpha. The results are provided in Table 3.

With the exception of one of the systemic sub-tests, the sub-tests in this battery appear to be adequate for their purpose, although some sub-tests would need to be improved or replaced, should it be used for diagnostic profiling. With hindsight, the low reliabilities for the systemic sub-test is not surprising. The focus of this research was on the schematic and procedural levels of the model, and it was initially assumed that a word order sub-test would adequately tap “linguistic knowledge”. However, this is clearly not the case. The concept of linguistic knowledge is much more complex than this, and with hindsight it appears more obvious that defining linguistic knowledge operationally is a difficult task. The assumption that we do know what linguistic knowledge means is taken from Widdowson. For him the real question is the higher levels of the model rather than the “second order abstraction”. This would appear to be an assumption that we should not make.

Test total score is adequately reliable for the purposes of this research.

| Table 3 |
| Reliability estimates |
| Component | Reliability |
| Systemic 1 | 0.74 |
| Schematic 1 | 0.71 |
| Procedural 1 | 0.82 |
| Systemic 2 | 0.34 |
| Schematic 2 | 0.88 |
| Procedural 2 | 0.89 |
| Test total | 0.89 |
6.2. Discrimination

Using the total test score, it was possible to accurately discriminate between practised and unpractised readers, as defined in Section 5. A z-test was used to compare the means of the two groups, shown in Table 4, which proved significant at \( p < 0.001 \). This provides evidence that the test as a whole is sensitive to the difference between practised and unpractised readers.

6.3. Correlational analysis

The main method of investigation selected at this exploratory level is correlation. The correlation matrix can reveal indications of the structure of the test, providing some evidence for the convergence of sub-tests of the same construct, and divergence between sub-tests of different constructs. It has already been noted that this study could not account for test method, because of the wide range of item types used in the test. However, suggestive results would nevertheless indicate whether further investigation of the model would be worthwhile. The correlation matrix is presented in Table 5.

The diagonal in Table 5 contains the reliability coefficients for each of the sub-tests, and thus is a copy of the information provided in Table 3. The upper right triangle contains the un-corrected correlation matrix, and the three shaded boxes represent the correlations between different sub-tests that were designed to measure similar constructs. These correlation coefficients can be treated, tentatively, as validity coefficients for the constructs. Each of the coefficients is significant, and

<table>
<thead>
<tr>
<th>Table 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Table 4</strong></td>
</tr>
<tr>
<td>Descriptive statistics for practised and unpractised readers</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Standard deviation</td>
</tr>
<tr>
<td>( Z (x_1-x_2) = 4.6758 (\pm .9410) )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Table 5</strong></td>
</tr>
<tr>
<td>Correlation matrix</td>
</tr>
<tr>
<td>Sub-test</td>
</tr>
<tr>
<td>Systemic 1</td>
</tr>
<tr>
<td>Schematic 1</td>
</tr>
<tr>
<td>Procedural 1</td>
</tr>
<tr>
<td>Systemic 2</td>
</tr>
<tr>
<td>Schematic 2</td>
</tr>
<tr>
<td>Procedural 2</td>
</tr>
</tbody>
</table>
higher than coefficients between sub-tests designed to measure different constructs. The coefficient of 0.49 between the two systemic sub-tests is unsurprisingly low, because of the unreliability associated with the Systemic 2 sub-test. Despite this encouraging finding, it should be stressed that 0.42 and 0.55 represent only 18% and 30% shared variance, respectively, for the schematic and procedural level sub-tests. Although each sub-test has more in common with its partner than other sub-tests, the shared variance is not impressively large.

The bottom left triangle contains the disattenuated correlation matrix. That is, the correlation coefficients have been corrected in each case to take the varying reliability of the sub-tests into account. It has the effect of holding reliability steady, and asking what correlation coefficient could have been expected if each of the sub-tests were perfectly reliable.

The first factor to notice is the extremely high correlation between the two systemic sub-tests. A correlation of 0.98 is intuitively unsatisfying, and appears to be an artefact of raising the reliability coefficient from 0.32 to 1.0 on the Systemic 2 sub-test and its impact on the initial significant correlation of 0.42. Correction for attenuation should not be used when sub-tests have extremely low reliability coefficients, or when the reliability coefficients between two sub-tests are not reasonably close together. We must therefore discount from this discussion all corrected correlation coefficients between the Systemic 2 sub-test and other sub-tests. These are marked with the dagger symbol (†) in Table 5.

It is therefore unsafe to conclude anything about the coherence of the construct “linguistic knowledge” on the basis of these results, or these sub-tests. We have already argued that one of the flaws in this research, and in applied linguistic theory, is the assumption that we know what “linguistic knowledge” is, and that it does not need careful definition.

With regard to the schematic level, or communicative capacity, there is mixed evidence. We have noticed that the correlation matrix indicates that the sub-tests share approximately 17% of variance, and this increases somewhat to 0.53 or 28% once the correlation is corrected for attenuation. Nevertheless, this is still low, and is very close to the correlation between Schematic 2 and Procedural 1, at 0.51. The distinctiveness of this level cannot therefore be convincingly argued from these results.

Evidence for the existence and independence of the procedural level, on the other hand, is much clearer. The corrected correlation is 0.75, or 56% of shared variance, and this figure is much higher than the coefficients between either of the procedural sub-tests and all other sub-tests. As a similar pattern was observed in the uncorrected correlation matrix, we can be more confident that this finding is stable and interpretable. We noted in Section 4 that Cooper (1984) had previously designed a test using a similar construct, and he had concluded that the procedural level was the most powerful in discriminating between students of different reading abilities. The findings of this research may indicate that Cooper’s findings are linked to the fact that this level of the model is more coherently defined, more easily operationalised, and so can be used more effectively in the design of reading tests.
7. Conclusion

This study has taken a discourse model of communicative competence that was set out over a decade ago, and attempted to test that model empirically. This is a practice which is not often followed, and for this reason many of the models in current use remain underdefined and non-operational. With hindsight, the study has suffered from not being able to account for test method, but it has been argued that this is not a fatal error in an exploratory study of this nature. As an early attempt to investigate a mainstream model, the design is adequate in the process of generating more advanced research designs.

Although exploratory in nature, the study has produced empirical support for the likelihood of the existence of the procedural level at work in comprehending texts. That it could not offer empirical support in favour of the existence and coherence of the schematic level is worrying, in that the level of capacity or ability for performance, requires further definition. That is, a significant part of the theoretical basis of the model has been called into question.

This article set out to investigate the validity of the model as a potential basis for the design and development of reading tests. It has discovered empirical evidence for the potential existence of the highest level of processing, indicated that the level of communicative capacity may be underdefined, and argued that we cannot take for granted that we know what linguistic knowledge is. Further research is required. This may involve improving the definition of communicative competence at the level of knowledge and ability for use, in a way that makes operationalisation of those concepts easier. This process will encourage more empirical research into models, rather than continued theorising and defining without investigation. Secondly, empirical studies must take account of variables such as test method, that were not included in this study. Such studies would help to reduce the noise that is probably present in the correlation matrix of Table 5. Finally, it is important to design and pilot reading tests that are based on what we do think may be empirically established constructs; all too often the nature of constructs and their relationship to other constructs cannot be established, because there is no agreed stable point at which to begin. Establishing one or more bases on rock rather than shifting sands is a prerequisite to making progress in model evaluation.

References


Hoey, M., 1983b. Three metaphors for examining the semantic organisation of monologue. English Language Research, Birmingham.


Appendix A: Schematic 2(A)

Read the following newspaper story about the problems of pollution in Rome. It contains 26 sentences, each marked with a letter (a–z). 10 of the 26 sentences do not belong to this story, and have been printed here by mistake. When you read a sentence which does not belong in the story, write the letter at the beginning of the sentence on your answer paper.

Here is an example:

(a) Whether on the ground or in the air, the high-speed collision of aircraft is every pilot’s worst fear.
(b) The world’s leaders had to meet in conference.
(c) Yet, each day in the U.S. fears grow. “Near midairs,” is the safety expert’s term for when two planes come dangerously close to each other in the air.

Answer: b.

1. When in Rome—don’t breathe

(a) Traffic police donned surgeon’s masks this month to protest about sickening exhaust fumes.
(b) And irate magistrates threatened to close Rome’s historic centre to traffic if the city did not adopt a plan to fight air pollution.
(c) So serious has the pollution become in Europe’s largest and busiest waterway, that it is affecting France and West Germany as well as the Netherlands.
(d) Baroque church facades, medieval houses and ancient ruins are stained blacker and blacker each year with the soot and exhaust from the city’s infamous traffic.
(e) The city state has one of the newest airports, super-highways and numerous air conditioned shopping centres.
(f) But, the stronger calls each year for the city to close the historic centre to traffic still have not been taken up by the city authorities.
(g) City officials say that it would only worsen the problem in the “Periferia”, the vast outskirts of Rome where getting around is at least as difficult as it is downtown.
(h) This is the conclusion of a Commonwealth report by a group of experts on technological change in society.
(i) The white-gloved traffic police staged their protest last week to draw attention to the health hazard from the exhaust spewing from the endless stream of cars.
(j) “We traffic police are perhaps more exposed, but everybody who works in the centre runs the same risk to their health,” said Marco Ciaffi, 30, who directs traffic at the busy Piazza Venezia.
(k) Others condemn their prosperous land as a cultural desert.
(l) The Dutch Minister of Transport said this week that he would be seeking for compensation.
(m) Judge Gianfranco Amendola threatened to call out the "Carabinieri" to block traffic in the city centre if the city did not adopt a new plan by tomorrow to reduce pollution.

(n) "Things definitely look better now than ten days ago when there were fears of a drought," he said.

(o) He later withdrew the threat.

(p) Several unions have been promoting "traffic-free day" on November 28th.

(q) They are asking everyone to leave their cars at home to see how well the city can move with public transportation that isn’t hindered by private car traffic.

(r) The mayor has said that experts are working on the problem, focusing on the creating of more pedestrian islands, pay parking lots and one-way main thoroughfares.

(s) However, no date has been set for the finalisation of such a plan.

(t) One problem exists: more than 76% of the population is Chinese.

(u) So, the forecast is that it will be rather cloudy with isolated showers around the hills.

(v) The central Piazza Navona has remained closed off to traffic since 1968 and the Spanish Steps area since 1983.

(w) Historical experience clearly suggests that heavier reliance on high technology encourages growth.

(x) But other "emergency" measures have had little success.

(y) So far, eighty two cases have proved fatal.

(z) New measures must soon be introduced if historical Rome is to be preserved for the future.

Appendix B: Schematic 1(B)

Read the following questions, and then read the passage quickly. Write your answers to the questions on the answer paper.

In the second paragraph of the passage we are told that "Alcohol intake, with its attendant social and health problems, has gone up and down according to cost and availability more than anything else."

1. Make a list of 6 social or health problems mentioned in the passage that are associated with drinking too much alcohol.

2. Four organisations have tried to offer solutions to these alcohol associated problems. What is the solution suggested by each of these four organisations:

   (a) The conference of Medical Royal Colleges;
   (b) National Drinkwatchers’ Network;
   (c) Alcoholics Anonymous (AA);
   (d) The Health Education Council.
What's your poison?

The everyday story of drinking in America is a familiar one. More than one in ten Americans now drinks more than one bottle of alcohol a day. Alcohol consumption has doubled in the last 20 years. This growth is primarily due to the increased availability and consumption of alcohol, which has been climbing steadily. The number of bars, pubs, and other places where alcohol is served has increased dramatically. The problem is that drinking has increased in the last few decades, and the public is not aware of the dangers of alcohol consumption.

According to the national figures, more than 50% of the population drinks more than one bottle of alcohol a day. Alcohol consumption has doubled in the last 20 years. This growth is primarily due to the increased availability and consumption of alcohol, which has been climbing steadily. The number of bars, pubs, and other places where alcohol is served has increased dramatically. The problem is that drinking has increased in the last few decades, and the public is not aware of the dangers of alcohol consumption.

Appendix C: Procedural IA

Look quickly over this passage from a cookbook. Pay attention to words that have been underlined. Fill in the table on your answer paper. The first one has been done for you as an example.

1. How to clean fish

If you have bought the fish ready-filleted, etc., from the fishmonger, he will, of course, have washed it there. Even so, it is better to wash it in plenty of cold water yourself—pat dry on kitchen paper.

With whole fish you intend to bake, you may find there are loose scales on the skin. Scrape these off gently with a knife. You will find they come away more easily if you scrape towards the tail.
Most fishmongers will fillet and skin on request, but you may sometimes find you have to do it yourself. These step-by-step pictures show plaice being filleted and skinned; the same rules apply to any white fish of similar shape.

_Note._ If you dip the edge of your knife in a little salt, you will find you have a better cutting edge.

Picture 4 shows the correct way to skin fish, but if you are in a great hurry and you put the fillets under a hot grill for just a moment, the skin comes away with no difficulty.

1. Cut off head with a sharp knife. (_It_ provides a great deal of flavour for fish stock.)

2. Make a deep incision down the back, i.e. on the side with the dark skin. Make a smaller cut round the edge of the flesh, avoiding the bones and fins.
3. Insert the tip of the knife and gradually loosen flesh away from the backbone, folding it over gently with your fingers. Do not hurry or pull, or you may waste some by leaving it on the bone. Turn the fish the other way and repeat the process, working from the tail rather than the head. Then, turn the fish over and do the same on the white side, giving four fillets.

4. To skin, take hold of the fillet in your left hand, holding it by the tip. Make a very firm cut with a sharp knife just at the tip, then lift the fish with your knife away from the skin; continue like this until all the flesh is free.

Appendix D: Procedural 1(B)

Read the following passage. Within the passage you will find a number of alternative sentences. Choose the sentence which best joins what went before in the passage with what comes afterwards. Record your answer (a, b or c) on your answer paper.

For example:

"We may say that there are various reasons for this problem."

a. One of these is simply the fact of high unemployment.  b. This is not a great problem in itself.  c. Several other reasons could no doubt be found.

"It is not necessary to discuss other reasons here."

The answer to this question is "a".

1. Philosophy

Most human beings are curious. Not, I mean, in the sense that they are odd, but in the sense that they want to find out about the world around them, and about their own part in this world.
1a. But they cannot
do this easily.  
1b. They therefore ask 
questions, they wonder,  
they speculate.  
1c. Or, on the other hand, 
they may wish to ask many 
questions.

What they want to find out may be quite simple things: What lies beyond that range of hills? Or they may be rather more complicated inquiries: How does grass grow? Or they may be more puzzling inquiries still: What is the purpose of life? What is the ultimate nature of truth?

To the first question the answer may be obtained by going and seeing. The answer to the next question will not be so easy to find, but the method will be essentially the same.

2a. So, he is forced to 
observe life as he sees it.  
2b. Although, often it may 
not be the same.  
2c. It is the method of 
the scientist.

A method that may reasonably be summed up by the phrase: "Going and Seeing". The last set of questions would normally be thought of as philosophical, and it would not be easy to find answers to them that would command general agreement. Some people would say that they are unanswerable.

3a. But those who have tried 
to answer them in the past 
have used the method of 
speculation rather than of 
investigation, of sitting and 
thinking rather than going 
and seeing.  
3b. Therefore, they have 
sat at home and not been 
too concerned about such 
questions.  
3c. The reason for this is 
clear to see: they are 
indeed, even in principle, 
unanswerable.

"Leisure," as Thomas Hobbes remarked, "is the mother of Philosophy"; the same relationship, it will be noted, as that which proverbially exists between necessity and invention. This should not be taken to imply that philosophers are not busy people, but their activity is likely to be mental rather than physical.

4a. This is not to say that all 
of the scientists do nothing 
but think about problems of 
science and physics.  
4b. Therefore, the philosopher 
must be free from all more 
practical duties and problems 
of everyday life.  
4c. It would be a misleading 
over-simplification, however, 
to identify science with the 
investigation and philosophy 
with thinking.

The scientist who is investigating the world around him will certainly do some sitting and thinking about the results of his inquiries; and mathematicians, often as a result of pursuing lines of thought which seemed at first to have no practical applications, have made enormous contributions to modern discoveries in physics. The philosopher who is speculating about the nature of truth, though he may not do much going, is likely to do a certain amount of seeing. He must have some data for his reflections, even if it is only that which is provided by the fact that he is
reflecting. And modern philosophers especially often undertake detailed investigations into the ways in which language is used.

5a. Nevertheless, it is on the whole true that for science the emphasis has been investigation, and for philosophy on speculation.  
5b. Reflecting is thinking, while it is obvious that investigating involves doing things.  
5c. Thus, science is a matter of pure investigation, and philosophy is merely thinking, for which it is to be criticised.

It will be useful now for us to examine more closely what the word “philosophy” has been and is used to describe.