Chapter 7
Prior Knowledge State Tests: Background and description

1 Introduction

It has been demonstrated in our ex post facto research report 2 (chapter 6), that there are no grounds for assuming that prior knowledge can be measured by means of indicators. In order to be able to study the precise mechanisms of the prior knowledge effects and in order to draw conclusions of relevance to students and educators, it seems worthwhile trying to gain an insight into the prior knowledge state through the development and use of PKS tests. Earlier we argued that other possible evaluation techniques such as schema construction (cognitive structure test), portfolio assessment and error analysis are more time consuming, more expensive and less promising (Wagemans and Dochy, 1989; Dochy and Bouwens, 1990e).

In this connection, attention must be paid to the generally recognized "requisite and... foundational role played by domain knowledge" (Alexander and Judy, 1989). Research in cognitive psychology during the past two decades has produced at least one undisputed finding about academic performance: those who know more about a particular domain generally understand and remember better than those with only limited background knowledge (e.g., Chi, 1985; Glaser, 1984).

This chapter deals with the question of assessing the PKS, more specifically it looks at knowledge state tests. First, we will give a general outline of our prior knowledge state tests, assigned as PKST or KST. In the second part of this chapter the background and construction of these tests will be described. For those tests which, as far as we know, have never been used before for this purpose, we will give some supplementary background and support for their construction and use (e.g. the optimal requisite KST and the domain-specific KST). Finally, comment is given on the psychometric qualities.

Prior knowledge state tests

Carrying out a study of the literature (see chapter 3), it was determined how prior knowledge is measured in research and experiments. Taken alongside our earlier observation that prior knowledge is rarely defined, it is striking that prior knowledge is often measured by means of tests which have not been specifically developed for this purpose (Dochy, 1988). In other cases, tests measure a very small part of the PKS, such as preconceptions about a specific subject. An example of the latter type of test is asking for a definition of ten words used in golf or basket-ball. Again this stresses the need for more ecologically valid research in this area (De Corte, 1990b). In addition, different sorts of prior knowledge have been measured. Some of the tests measure metacognitive knowledge, other tests focus on domain-specific knowledge or other parts of content knowledge. Finally, in most of the studies we screened there was no attention paid to the quality (in terms of validity and reliability) of the assessment instrument used to assess prior knowledge.

The research used different PKS tests, varying mainly in terms of content. The subsequent investigations reported in chapter 10 will also focus on the influence of educational, epistemological and psychometric dimensions. In order to define the PKST variants indicating different content components, content experts (i.e. economists) were asked to ascertain types of content knowledge that influence learning results. This was extended with the above stated review of the types of tests used in literature. Within the PKS tests identified, we can distinguish three groups: Subject-oriented Knowledge State Tests (SO KST) (i.e. covering a part of a subdomain), Cross-domain Knowledge State Tests (CD KST)(i.e. covering a part of all subdomains) and Domain-specific Knowledge State Tests (DS KST)(i.e. covering the whole domain).
In this investigation, a set of short tests (i.e. the SO KST and CD KST) were developed, based on the following considerations:

1. The need to detect the differential role of the components of prior knowledge.
2. The need to restrict the load placed upon the student by the test.
3. The need to confine the research procedure to a single day.
4. The need to assure that, within the limitations of consideration 3, the students had sufficient time to complete the learning task.

In developing our tests, we have tried to make a clear distinction between the different components of prior knowledge, and we have developed a test for each individual sort that, according to the experts was relevant in this case. The individual sorts of test were: Subject-oriented Knowledge State Tests (SO KST), an Optimal Requisite Knowledge State Test (OR KST) and a Mathematics Knowledge State Test (MA KST), which are Cross domain Tests, plus a Domain-specific Knowledge State Test (DS KST). Figure 1 gives an overview of these types of different tests in relation to the domain of economics. It will be clear that all PKST relate to the domain-specific prior knowledge, although some are directed towards one or several parts of the DS PKS and others towards the whole domain. The description of the tests as given below concerns the 1.0 version, as used in our next investigation. For later studies there were updated versions of the tests, which are described later.

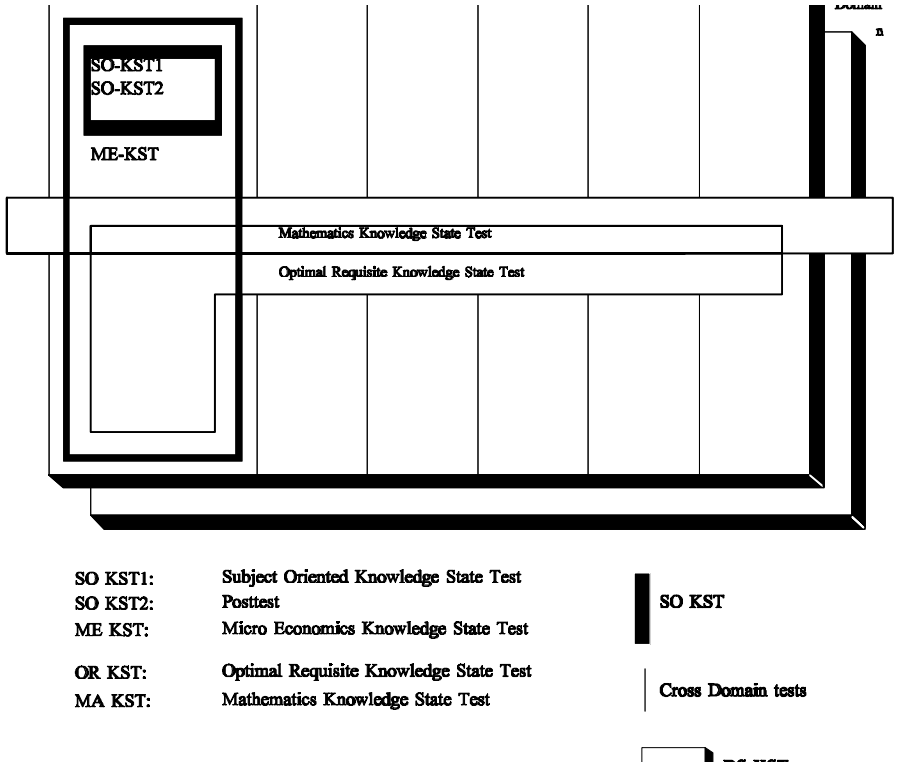

Figure 1: Different knowledge state tests in relation to the domain.

2 Background and description of the PKS tests

2.1 The Subject-oriented Knowledge State Test (SO KST)

The term 'subject-oriented prior knowledge state test' is used in this study to identify a test which is of direct relevance for the material to be studied (in this case the study modules or blocks to be studied for the Economics and Money course).

Four experts (economists) from the University of Limburg (UL) screened the items from the block tests of the UL to ascertain whether they were of direct relevance with regard to the aims of blocks 3 and 4 of the Economics and Money course. In addition, the relevance of the selected items was assessed once more by an expert from the OU. Finally, a number of these items were collected to form a representative and well-balanced test in agreement with the aims, as stated in the course. This study used subject-oriented prior knowledge state tests for the modules "The supply of goods: the costs" and "The supply of goods: producer behaviour and market forms". These tests (SO KST1 and SO KST2)(first version) each consisted of 12 items of the four-choice type. The choice was partially influenced by the possibility of converting the existing material (tests) to similar prior knowledge state tests.

Obviously the subject-oriented prior knowledge state test for economics examines what the student is already capable of and what he knows with regard to the course to be studied. The tests are goal-related standardized tests.
Further a Micro-economics Knowledge State Test was constructed. This ME KST is a subject-oriented test covering the subject-matter of micro-economics and consists of 11 items. This is a representative sample of the itembank of the Maastricht Economics Faculty. Items are of the true/false type.

2.2 The Cross-domain Knowledge State Test (CD KST)

Prior knowledge tests that do not representatively cover the whole domain but do exceed the content of a subdomain, are called cross-domain tests. They make a cross-section of several subdomains focusing on a specific aspect, for example the mathematical knowledge (MA KST) or the optimal requisite prior knowledge related to a certain subject-matter (OR KST).

The Mathematics Knowledge State Test (MA KST)

A prior knowledge state test for mathematics was constructed at pre-university (VWO) level. This test comprises 28 multiple-choice items cutting across the domain and is based on the mathematics test of the University of Antwerp (Dyck, 1976). The original test by Dyck is still used in the University of Antwerp as a instrument for the self-evaluation of starting level. It has been shown that the test has a high degree of validity and reliability (>.80). Moreover the test score had a strong correlation with study success in first year at university.
In order to decide if this test was appropriate for our purpose, it was set to four economics students (along with the other tests). From that try-out, it was clear that the original test containing over a hundred items could not be completed within several hours. We had to reduce the test to a maximum of 30 items. This enabled students to take the test in one hour and without a considerable loss of attention.

The Optimal Requisite Knowledge State Test (OR KST)

Prior knowledge, however, is much broader than knowledge of the content of a subject in the narrow sense. It also includes optimal requisite prior knowledge. This is the content related knowledge that a student must possess if he is to start his course of study in optimal circumstances. The OR KST was constructed by asking six of the general economists from the Economics Product Group of the University of Heerlen and ten economists from the Economics Faculty of the University of Limburg to describe the optimal prior knowledge needed to study the above mentioned modules. In addition, they were asked to present concrete themes on this knowledge and to name articles and books in which it is treated. On the basis of the answers, a multiple-choice test was constructed, consisting of 8 items, each representing a set of sub-items. Construction of this test was based on the opinion of economics experts who identified the optimal requisite knowledge required for the execution of the learning task.

Support for the Optimal Requisite KST

Apart from the assumption that prerequisite knowledge could be a good measure of prior knowledge and study success, there is very little information in the literature relevant to constructing an optimal requisite (OR) test. We did find some support for our assumption in earlier work of Hively et al. (1973), Boekaerts (1979), Gagné (1977), Weinert (1989) and in the elaboration theory (Reigeluth and Stein, 1983).

If a domain, subdomain or the content of a certain course is defined, it must be possible to collect the prerequisite knowledge, necessary for its study.
Gagné has drawn attention to the fact that learning a specific skill is impossible without specific prerequisite knowledge (Gagné, 1977). Moreover, Weinert and his colleagues showed that general aptitude cannot replace prerequisite knowledge in whole or in part (Weinert, 1989). Hively et al. (1973) stated that, even if prerequisite behaviour is not taught in a unit, it probably should be tested, since failure that might be attributed to the unit may be due to inadequate preparation, i.e. prior knowledge. This is in accordance with the elaboration theory of instruction (Reigeluth and Stein, 1983). This theory states that a learning prerequisite sequence is based on a learning structure, which gives the sequences in which certain knowledge and skills should be mastered. More support was found in the view of Boekaerts (1979) concerning the domain test. She defends the position that a teacher, even if he has formulated objectives for a lesson or a sequence of lessons, must still do some things before he can plan the teaching-learning process. Among these things, Boekaerts identifies the assessment of the domain-based knowledge which is a prerequisite for attaining the objectives (also called "required domain-based knowledge"), and the assessment of the students' knowledge of the domain in question.

2.3 The Domain-specific Knowledge State Test (DS KST)

Construction of the DS KST

There is evidence that learning is much more domain-specific that earlier learning theorists believed (Shuell, 1986) and in chapter 3 we argued our choice for directing mainly on domain-specific knowledge. It is obvious that domain-referenced testing provides a reasonable possibility of measuring an individual's knowledge status and of tracking his progress within a certain domain. Apart from that main reason, there is more evidence to account for increasing application of domain-specific PKS tests.

First, from the above-stated view of Boekaerts, among others, it can be assumed that the learning process is also influenced by prior knowledge that is broader than strictly subject-specific prior knowledge. For this reason, a domain-specific test was developed covering the whole domain (up to a certain degree of difficulty of specialization). In our case we are concerned with the domain of economics.

Second, differences between students concerning specific subjects are sometimes rather large. In this case, a test at beginners level would not be able to bring to light all of the differences between the students. The chances of doing this is greater

when using a test related to end terms, i.e. one whose level corresponds to the end of the second year university study.

Third, because some students have already gained a great deal of experience in their working environment or have already attained a relatively high educational level (some higher vocational education degree or a university degree), a test set at beginner's level (final VWO level) would not be capable of measuring part of the prior knowledge state.

Fourth, the recent trend known as 'open learning' which uses flexibility as a key concept, is strongly related to the use of DS tests. Open learning tries to take the students' prior knowledge into account and allows them to study at their own place and pace. Students have a large degree of freedom in choosing educational media and objectives. A sudden openness related to objectives and other choices needs an appropriate assessment instrument. Domain-specific tests seem to be appropriate. Other instruments rather neglect the primary conditions of open learning. DS tests not only reveal measurement results, but also information and guidelines to attack deficiencies. According to Glaser (1990) testing should serve the learning process.

A fifth argument is the psychometric quality of the item bank for DS tests. The amount of available items for a course is often restricted and does not allow the removal of items on the basis of insufficient psychometric quality (validity, reliability, p-value). This problem disappears when using the larger item banks for DS tests.

A sixth reason in favour of DS tests is the possibility of using them for different functions, i.e. assessment of entrance level, of progress and certification. Further, it allows to focus on different dimensions that exceed beyond the content level (see chapter 10).

Finally, there is the trend towards internationalization of higher education and co-operation between European universities. In this respect, DS tests enable comparisons of individual students and comparisons of institutions to be made.

For these reasons, a domain-specific test was developed. This test, which is aimed at the whole domain of economics (as shown in figure 1), is set at the level which should be attained by the end of the second year of university study. The heterogeneity of the test population (or student population) is so great that a test at beginner's level would not be able to bring to light all of the differences between the students. After all, it can be assumed that students with years' of working experience in, for instance, the financial sector, or students who have obtained other academic (WO) or higher vocational education (HBO) diplomas will have advanced further than the beginners level in certain areas, and may achieve a score approximating to the final 'economist' level. In other words, because some students have already gained a great deal of experience in a working environment or have already attained a relatively high educational level, a test set at beginner's level (final VWO level) would not be able to measure some of the prior knowledge state.

The University of Limburg possesses a wealth of experience in constructing tests, especially tests associated with end terms. Studies into the importance of these end terms tests or progress tests for the first study year at the University of Limburg have been carried out by Imbos (1982, 1989) and Wijnen (1984). Our DS Knowledge State Test was constructed as a representative random test of items

selected from the item bank of the Economics Faculty of the University of Limburg. The items in the data bank are classified in 9 subject areas, viz.: reporting, financing, organization, marketing, macro-economics, micro-economics, public finances, international economic affairs and behavioural and social sciences.

The entire DS KST comprises 154 items divided between the various subject areas or subdomains. These categories correspond to an equal number of basic disciplines in economics. Each category contains different types of items with different codes: items related to general economics or business studies and items related to quantitative economics; items at two different difficulty levels.
In addition, the construction of the test will not be complete unless caution is devoted to the distribution of true and false items. Here too, chance successes must be avoided. Earlier studies (Ebel, 1972; Grosse and Wright, 1985; Verwijnen, Imbos, Van Hessen and Wijnen, 1987) have demonstrated that items with the answer key 'true' are answered correctly more often than items with the answer key 'false'. This is of importance for the comparison of parallel tests. A significant difference in the distribution of true and false items will result in an built-in bias to one direction or the other (Imbos, 1989). There is no significant difference in the distribution of both sorts of items in our test.

If test analysis is to be quick and the method of answering the items is to be simple, multiple-choice or true/false items must be used. This test consists of multiple-choice questions which can be answered with true/false or ?. The ?-alternative is taken as a third alternative in order to prevent guessing.
In addition, it is recommended that the extent of the test be kept to a minimum. Only the extensive domain-specific knowledge state test would present problems in this connection, if it were to be generally introduced. Although there are sufficient good arguments to justify the length of this test, practical considerations may require it to be shortened. There are, however, sufficient possibilities for doing so (Imbos, 1989; Wijnen, 1984). On the other hand, striving towards a minimum test length means a reduction of reliability.

Domain-referenced achievement testing

Our approach in constructing and using a DS KST is based on what is known in psychometrics as domain-referenced testing. In a final working paper, Hively et al.(1973) stated: "Domains of test items are structured and built up through the specification of stimulus and response properties which are thought to be important in shaping the behaviour of individuals who are in the process of learning to be experts. These properties may be thought of as stratifying large domains into smaller domains or subsets.... This provides the foundation for precise diagnosis of the performance of individuals over the domain and its subsets. In addition clear specification of the properties used to structure the domain makes possible inductive generalization beyond the domain to situations which share those properties."

Domain-referenced (DR) testing can be defined as the assessment of an individual's performance with respect to a well-defined level or body of knowledge. Items are organized into clusters with each cluster serving as a representative of a clearly defined content domain. In this study, the term a "domain-specific (DS) knowledge state test" is used to refer to the DR test we developed and more generally to DR tests constructed specifically to assess primarily the prior knowledge of students.

Underlying notions

First, underlying the concept of domain-specific testing is the notion of a continuum of knowledge acquisition ranging from no proficiency in that domain to a perfect prior knowledge state.
Second, test scores obtain primarily two kinds of information. One is the degree to which an individual has attained full prior knowledge in a certain domain (i.e. his position in the continuum). The other is the degree of prior knowledge of individuals relative to one another. DS testing will only focus on the first type of information, thus to assess an individual's status with respect to a certain domain.

Functions

Scores of the domain-specific knowledge state test may be used (a) to describe student performance for determining intake level, for diagnosing student learning deficiencies or for monitoring student progress, (b) to make mastery - nonmastery decisions, and (c) to evaluate programme effectiveness.
A domain-referenced measure provides considerable information for making decisions concerning student advancement (Cox and Graham, 1966). In our view, the DS KST will serve in the future as both a placement and a diagnostic tool.

3 The use of PKS tests in our investigations and psychometric qualities

In the coming investigations, discussed in chapter 8, the following tests are used: the SO KST1, the OR KST, the ME KST, the MA KST, and the SO KST2. During test-development, special attention was paid to content-validity. This was done by involving content experts (SO KST, OR KST) or by using valid item banks (ME KST, MA KST). To assess the reliability of the tests, the alpha-coefficient was calculated.

Table 1: Reliability of tests (1.0 versions)

	α
SO KST1	.40
OR KST	.68
ME KST	.43

MA KST	.82
SO KST2	.23

Item-test correlation for all tests reveals that there are no items with negative or low correlations to the total score. Nevertheless, reliability is to be considered as rather low, mainly due to the limited number of items in the tests. It is generally accepted that a test needs 40 items to reach a reliability of .80. It should be said that the micro-economics KST is an excerpt of a domain-referenced item bank, in which items are located covering the whole domain, thus not measuring a homogeneous set of sub-aspects. In relation to the MA KST it should be noted that the original Antwerp test (Dyck, 1976) had a reliability of .85.

In the succeeding investigation, discussed in chapter 9, the following tests (1.1 versions) have been used:

- A SO KST1 (1.1): this test includes 34 multiple-choice items (4 alternatives) and is related to learning units 14 and 15 of the "Economics and Money" course. Validity of the test was checked by content experts who assessed whether the items were representative of the subject-matter.
- An OR KST (1.1): this test consists of 8 items (open-ended or multiple-choice), each of them representing a set of sub-items (17 in total). Construction of this test was based on the opinion of economics experts who identified the optimal requisite knowledge for the execution of the learning task.
- A MA KST (1.1): this mathematics test, cutting across the domain, contains 28 items. The test is based on the already mentioned self-evaluation test of the Antwerp university.
- A SO KST2 (post-test (1.1)): this test, consisting of 34 items is a parallel test-version of SO KST1.

In order to bring the reliability of all 1.0 versions of the tests up to .80 or more, we used the general Spearman-Brown prophecy formula. To estimate what the reliability of a test would be if it were made longer the prophecy equation is

$$r_{xx} = \frac{K\,x\,r}{1 + (K-1)\,r}$$

where

r_{xx} = predicted reliability of a test K times as long as test version 1.0

r = reliability of test version 1.0

K = ratio of number of items in test version 1.1 to number of items in test version 1.0

As a consequence, the 1.1 versions of the PKST are approximately three times as long as the 1.0 versions (see table 2, after correction), with the exception for the MA KST which was already sufficiently reliable.

During test development, special attention was paid to content validity. This was achieved by involving content experts (SO KST1 & 2, OR KST) or by using valid item banks (MA KST).

To assess the reliability of the 1.1 versions of the tests (after enlargement based on the Spearman - Brown formula), the alpha-coefficient was calculated (see table 2, before correction). Item-test correlation (r_{it}) for all tests revealed negative or low correlations for some items. After omission of these items, the alpha coefficient was calculated for the corrected PKST versions (see table 2, after correction).

Table 2: Alpha-coefficients of the different PKST-versions 1.1 before and after correction for low r_{it} values

	α before r_{it} correction	α after r_{it} correction
SO KST1	.51	.63
OR KST	.94	.96
MA KST	.92	.93
SO KST2	.84	.85

Domain-specific Knowledge State Test

As described above, also a domain-specific prior knowledge state test will be set to the research population. The particular characteristics of this test already suggest that the determination of its psychometric qualities might be a problem. There is no problem in relation to validity since the test clearly represents - to a very large extent - the domain and has been developed by a team of domain experts. The psychometric quality problem is especially in game when determining the reliability of the test. If we calculate the alpha-coefficient, the test can be considered as very reliable: alpha =.93. This high reliability is probably caused by the fact that

the test is long (154 items), which gives a high alpha-coefficient. Moreover, the test is not necessarily homogeneous (see chapter 10) which implies that basic assumptions on which the calculations of the alpha-coefficient are based could have been violated.

Table 3: Alpha-coefficients for the parameters of the economics subdomains and curriculum accent dimension and mean alpha-coefficients

Parameters	α-coefficient	N items	αm
Economics subdomains dimension			
Reporting	.57	18	
Finance	.64	18	
Organisation	.69	18	
Marketing	.63	18	
Macro-economy	.71	25	.63
Micro-economy	.74	25	
Public	.51	11	
International Economics	.55	11	
Behavioural & Social Sciences	.63	10	
Curriculum accent dimension			
General economics	.93	139	.69
Quantitative economics	.45	15	

One solution to this problem might be to check the reliability of sub-parts of the test, making use of the knowledge profile dimensions, which will be further explained and elaborated in chapter 10. Calculation of alpha was repeated for two of these dimensions (economics sub-domains and curriculum accent), in order to be able to present a mean reliability score. When reorganizing the test into more homogeneous sub-parts, the alpha-coefficient and a mean alpha-coefficient were calculated. The results of this procedure are summarized in table 3. To enable the figures to be judged in a better perspective, the number of items within each sub-group of items is also given. Mean alpha seems to be >.63. This reliability score is - taking into account the restricted number of items in certain sub-parts of the test - acceptable for our research purposes.

"Show me your knowledge profile and I will tell you how to reach study success efficiently".