10

Other Inventories and Questionnaires

I shall conclude my account of approaches to studying in campus-based and distance education by describing some other inventories and questionnaires that have been employed to investigate differences in how students go about learning in higher education. These are of less importance than the SPQ and the ASI and have been used less often in distance education than the DESP inventory or the ILS. However, the findings of research using these inventories and questionnaires serve to highlight some of the themes discussed in previous chapters and provide further examples of the problems in developing quantitative instruments for use in this field.

The Inventory of Learning Processes

In Chapter 2, I mentioned the notion of ‘levels of processing’ in human memory put forward by Craik and Lockhart (1972). They claimed in particular that the use of ‘deeper’ or more abstract levels of processing during learning would lead to better retention than the use of shallow levels of processing. Craik and Tulving (1975) set out to test the latter claim in a series of experiments in which participants made judgements about each of a series of individual words. For example, they might be asked to say whether a word was printed in upper-case letters, whether it rhymed with a particular word, whether it was the name of a particular kind of thing or whether it fitted into a particular sentence in a sensible way. These tasks were assumed to induce progressively deeper levels of processing. Subsequently, the participants received an unexpected test on their memory for the words that had been presented and Craik and Tulving found that the level of performance varied directly with the depth of processing that they assumed had been induced.

There were, however, some anomalous results that could not be readily handled by assumed variations in depth of processing. First, words that had yielded positive responses were better remembered than words that had
yielded negative responses, particularly with deeper levels of processing, though presumably they had been processed at the same level. Second, there were considerable variations in retention even among tasks that involved only semantic processing. In particular, performance was better when the participants judged whether words fitted into more complicated sentences than when they judged whether words fitted into simple sentences. Finally, the more abstract tasks led to better performance even when the participants had been forewarned of the retention test and were free to try to remember the words just as they wished. To handle these findings, Craik and Tulving argued that learning depended partly on the depth of processing but more on the ‘spread’ or degree of elaboration of the processing at any level.

In fact, the notion that the elaboration of material to be learned would enhance its subsequent memorability was widely accepted in theories of human memory at that time. Schmeck et al. (1977) set out to construct a new questionnaire, the Inventory of Learning Processes (ILP), on the basis of these and other ideas in experimental psychology concerning human learning and remembering. They compiled a list of 121 statements concerning different aspects of studying and asked 503 students at a campus-based university in the US to say whether each statement was true or false as a description of how they generally learned (rather than how they learned within any particular course or academic discipline). In the light of the results of a factor analysis of their responses, Schmeck et al. reduced the list to 62 items in four scales, which can be found in Tables 1 and 2 of their paper. The four scales were defined as follows:

- Synthesis–Analysis (assessing deep, as opposed to superficial, information processing);
- Elaborative Processing (assessing elaborative, as opposed to verbatim, information processing);
- Fact Retention (assessing attention to details and specifics as opposed to generalities); and
- Study Methods (assessing repetitive, drill-and-practice habits of processing information).

(Schmeck and Grove 1979: 43)

To reduce the influence of response bias, for some items the response ‘true’ was in accordance with the meaning of the scale in question, whereas for other items the response ‘false’ was in accordance with the meaning of the scale. A respondent’s score on each scale was calculated simply by counting the number of responses that were in accordance with the meaning of the scale. Schmeck et al. found that the scales had satisfactory internal consistency and test–retest reliability, and that there were no significant differences between the scores obtained by male and female students. Schmeck and Grove (1979) found that the current academic performance of campus-based students (measured by their grade point averages) was
significantly correlated with their scores on the synthesis–analysis, elaborative processing and fact retention scales, but not with their scores on study methods. Fact retention and elaborative processing seemed to have direct effects on academic performance, whereas the effects of synthesis–analysis were exerted mainly through students’ entrance qualifications. Schmeck (1980) subsequently showed that the latter scale was also related to students’ scores on vocabulary and comprehension tests.

Of course, the development of the ILP was essentially contemporaneous with a number of other developments in research into student learning in higher education: the work of Marton and his colleagues in Sweden on approaches to studying (see Chapter 2), the development of the SPQ by Biggs in Australia and Canada (Chapter 5) and the development of the ASI by Entwistle and his co-workers in the UK (Chapter 6). Moreover, Ribich and Schmeck (1979) found that there was a modest amount of overlap between the dimensions measured by the ILP and those measured by Biggs’s (1970a) SBQ. In the light of these developments, Schmeck (1983) relabelled the first two of the scales in the ILP ‘deep processing’ and ‘methodical study’, respectively.

Schmeck claimed that his notion of deep processing was distinct from that which had been put forward by Marton and Säljö (1976a) because it was concerned not with a student’s intention or approach to a specific task but rather with the underlying cognitive processes. As explained in Chapters 2 and 5, it is true that Marton and Säljö used the expression ‘levels of processing’ to refer to students’ strategies when performing the specific task of reading isolated passages of text. There is, however, a far closer relationship between Schmeck’s notion of deep processing and Marton and Säljö’s notion of a ‘deep approach’, which referred to a general way in which students’ might set about their academic studies. Subsequently, indeed, Schmeck and Geisler-Brenstein (1989) acknowledged the parallels between the following pairs of constructs:

- deep processing versus shallow processing (Schmeck 1983)
- deep approach versus surface approach (Marton and Säljö 1976a)
- internalizing versus utilizing (Biggs 1979; see also Chapter 5)
- meaning orientation versus reproducing orientation (Entwistle et al. 1979; see also Chapter 6).

This seemed to demonstrate the convergent validity of the different research methods by which these constructs had been measured. However, it should be pointed out that in interview-based research the first two pairs are mutually exclusive: individual students cannot employ both deep processing and shallow processing, and they cannot exhibit both a deep approach and a surface approach. However, as I pointed out in Chapters 5 and 7, in questionnaire-based research, the latter pairs of constructs turn out to be essentially independent, so that individual students can score high or low on both internalizing and utilizing in the SPQ and on both meaning orientation
and reproducing orientation in the ASI (Trigwell and Prosser 1991a). Finally, Schmeck and Geisler-Brenstein noted that a distinction between deep and elaborative processing had also been made by Vermunt (see Chapter 9).

The ILP was extensively validated in research carried out at several campus-based institutions of higher education across the US (Schmeck 1983, 1988) but it has had only a very limited use outside that country and apparently none at all in distance education. Watkins and Hattie (1981a) used the ILP in a comparison of students in Australia and the Philippines. Some of the scales did not generate satisfactory measures of internal consistency and factor analyses failed to replicate the ILP's intended constituent structure, although the students' scores on the different scales of the ILP were significantly related to their cumulative grade point averages (see also Watkins and Hattie 1981b; Watkins et al. 1983). Schmeck (1983) suggested that the discrepancies between his own results and those obtained by Watkins and Hattie (1981a) could be ascribed to cultural and linguistic differences between the various student populations.

Subsequently, however, Schmeck (1988) described an unpublished study where 269 students at a campus-based university in the US had completed the ASI in addition to the ILP (see chapter 6). Entwistle (1988) provided a table of correlation coefficients between their scores on the scales of the ILP and their scores on the 16 subscales of the ASI. As Entwistle commented, there was very little overlap between the deep processing scale of the ILP and any of the ASI subscales defining meaning orientation. In fact, high scores on deep processing were associated with low scores on the subscales relating to fear of failure, improvidence and surface approach; and the ASI subscales defining meaning orientation were most closely related to the elaborative processing scale. Even so, the strength of these relationships was relatively modest: perhaps unsurprisingly, the strongest relationship was that between the scores on methodical study and disorganized study methods ($r = -0.49$).

A factor analysis carried out on the students' responses to the total set of 126 items was said to have yielded 'six clearly interpretable factors and a "hint" of two additional factors' (Schmeck 1988: 178). I mentioned in Chapter 6 that these factors were based on groupings of items that corresponded to the scales in the ILP rather than to the subscales in the ASI. There were indeed factors corresponding to deep processing and elaborative processing, respectively, but a third factor subsumed items that were associated with both fact retention and methodical study. The five remaining factors represented different aspects of confidence or self-efficacy (see also Schmeck and Geisler-Brenstein 1989). However, two other studies that had used short versions of the ASI and the ILP at campus-based institutions in Scotland (Entwistle and Waterston 1988; see also Entwistle 1988) and the US (Speth and Brown 1988) yielded factor solutions that failed to represent the constituent scales of either instrument. This tends to suggest that the anomalous findings obtained by Watkins and Hattie (1981a) might have been due
to inherent problems with the ILP itself, rather than to cultural or linguistic factors.

Schmeck et al. (1991) supplemented the four original scales of the ILP with seven other scales that examined broader aspects of self-concept and personality, to produce a Revised Inventory of Learning Processes (ILP-R). In completing this instrument, respondents indicated the extent of their agreement or disagreement with each of 160 statements along a six-point scale between 'strongly agree', scoring 5, and 'strongly disagree', scoring 0. Schmeck and Geisler-Brenstein (1989: 100) had mentioned an unpublished study in which an earlier version of the ILP-R had been administered to students at campus-based institutions in both the US and China. They commented that 'a similar factor structure was obtained for both groups', but they did not present any quantitative results. Nevertheless, it is clear from their brief discursive account that the factor solutions failed to replicate the scale structure of the original ILP.

Subsequently, Schmeck et al. (1991) reported the correlation coefficients among the different scale scores that had been produced by the American students in this study. While the scales of the original ILP had earlier been stated to be 'relatively independent' (Schmeck et al. 1977: 420), their scores on the scales of the ILP-R exhibited substantial associations among the deep processing, fact retention and elaborative processing scales. Henson and Schmeck (1993) used the original ILP in an investigation involving 89 students at two campus-based institutions in the US and obtained similar results. It would therefore appear that the scales of the ILP and the ILP-R do not provide distinctive, independent and homogeneous measures of the constructs on which these instruments were originally based and, accordingly, neither can be recommended as a useful instrument for investigating student learning in higher education.

It should be also be pointed out that doubts had already been raised about Craik and Lockhart's (1972) idea of 'levels of processing' even before the ILP had been developed (Baddeley 1976: 167–8) and more extended critiques were published shortly thereafter (Baddeley 1978; Eysenck 1978). Some problems were basically factual. For instance, I mentioned in Chapter 5 that under certain circumstances shallow processing could lead to relatively good retention; in fact, Biggs (1978b) himself obtained results along these lines. Again, people with memory problems due to brain damage often do not show any impairment of semantic processing (Baddeley 1976: 167). However, other problems were conceptual. In particular, Baddeley (1978) and Eysenck (1978) claimed that no satisfactory independent criterion of either the 'depth' or the 'elaboration' of processing had been devised, in which case they argued that explanations in terms of levels of processing were at best post hoc and at worst entirely circular. Schmeck (1983) claimed to have taken such issues into consideration when developing the ILP but, even so, Craik and Lockhart's account is no longer regarded by psychologists as a useful framework for understanding human memory (see, for example, Eysenck and Keane 1990: 148–55; Baddeley 1996: 115–24).
The Learning and Study Strategies Inventory

In Chapter 5, I mentioned Biggs’s (1970a) early finding that differences in how students went about their learning consisted of a number of distinct components that could not be reduced to one single dimension of ‘good’ versus ‘bad’ in their implications for subsequent attainment. Biggs himself took this finding as evidence against the idea that there was a determinate set of ‘study skills’ that somehow guaranteed effective academic performance. Weinstein et al. (1988) arrived at a similar conclusion by comparing several inventories published in the US that claimed to measure study skills. In Chapters 3 and 7, I mentioned research carried out by Martin and Ramsden (1987) and by Ramsden et al. (1986, 1987), which suggested that students’ participation in study skills programmes did not give rise to any improvement in their approaches to studying or their conceptions of learning.

As a result, in the UK at least, the attitudes of teachers and researchers concerning the value of study skills programmes are nowadays very sceptical (see, for example, Ford 1980; Gibbs et al. 1980; Gibbs 1981: chapter 4; Cowan 1989; Entwistle 1992). This may be somewhat unfair: a meta-analytic review of previous studies carried out by Hattie et al. (1996) found that study skills programmes typically produce clear improvements in measures of performance and in students’ self-concepts and attitudes though only very slight improvements in students’ reported study skills. Nevertheless, in the case of studies conducted with university students, Hattie et al. found that the effects upon academic performance were much more modest. At this educational level, Hattie et al. concluded that study skills interventions were relatively ineffective, except in improving students’ attitudes to learning or in reducing their levels of experienced anxiety.

Despite this conclusion, it is clear from research in experimental psychology on human learning and memory that people differ from each other in the kinds of strategies that they use when they engage in learning tasks, and that this strategic variation has implications for the quality of their subsequent performance (Weinstein et al. 1979). On the basis of this research, Weinstein et al. (1988) devised the Learning and Study Strategies Inventory (LASSI), which in its final version contained 77 items within ten scales that were related to different aspects of studying. Brief definitions of the ten scales are given in Box 10.1. For each item, respondents are asked to indicate the extent of their agreement or their disagreement with a particular statement along a five-point scale, and the responses are then coded so that higher scores reflect more desirable patterns of studying. (In particular, the ‘anxiety’ scale as a whole is coded so that higher scores represent lower levels of anxiety.) Using data from 96 students taking an introductory course in educational psychology at a campus-based university in the US, it was found that the ten scales achieved satisfactory levels of both internal consistency and test-retest reliability.

Cano-Garcia and Justicia-Justicia (1994) administered a battery of questionnaires including the ASI, the ILP and the LASSI to 991 students who
<table>
<thead>
<tr>
<th>Scale</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiety</td>
<td>Often worries about school; may worry so much that it is hard to concentrate; easily discouraged about grades; tense about school and studying for tests; nervous even when well prepared (coded so that higher scores represent lower levels of anxiety)</td>
</tr>
<tr>
<td>Attitude</td>
<td>Attitude about and interest in college</td>
</tr>
<tr>
<td>Concentration</td>
<td>Ability to concentrate, pay close attention, listen carefully and think about what is being said; not easily distracted</td>
</tr>
<tr>
<td>Information processing</td>
<td>Uses imaginal and verbal elaboration; thinks about how new information fits with what is already known; interrelates new information; creates comparisons; thinks about the meaning of what is read and heard; translates information into one’s own words; uses logic</td>
</tr>
<tr>
<td>Motivation</td>
<td>Willingness to work hard; level of motivation for college; has a considerable degree of incentive; diligent; stays ‘on top of’ work; self-disciplined</td>
</tr>
<tr>
<td>Selecting main ideas</td>
<td>Seems to be able to pick out key ideas and critical points in information read or heard; focuses on important points in what has been read</td>
</tr>
<tr>
<td>Self-testing</td>
<td>Reviews information learned; reviews regularly; prepares for classes and learning</td>
</tr>
<tr>
<td>Study aids</td>
<td>Makes use of a broad approach to learning, makes good use of aids to help learning; supplements learning with helpful techniques; makes good use of key words, practice exercises, sample problems, examples, headings, diagrams, etc., to help learning</td>
</tr>
<tr>
<td>Test strategies</td>
<td>Approach toward taking tests and exams; generally prepares appropriately; reviews right materials; ties materials together well; flexible when necessary</td>
</tr>
<tr>
<td>Time management</td>
<td>Uses time well, is well organized, systematic in planning the use of time; productive in using time to the best advantage</td>
</tr>
</tbody>
</table>

were taking courses in ten academic subjects at a campus-based university in Spain. A factor analysis was carried out on their scores on the ten scales of the LASSI, and this produced just two factors. One reflected the students’ scores on test strategies, concentration, selecting main ideas, (low) anxiety and attitude, which seemed to relate to the specific goal of succeeding at university. The other factor reflected the students’ scores on self-testing, motivation, time management and study aids, which all seemed to relate more to a constant, responsible and organized approach to studying. A second factor analysis was carried out on the students’ scores on the 16 subscales of the ASI, the four scales of the ILP and the ten scales of the LASSI; this produced just three factors:

- one factor reflected the students’ scores on information processing (LASSI), interrelating ideas (ASI), elaborative processing (ILP), deep approach (ASI) and use of evidence (ASI)
- a second factor reflected the students’ scores on surface approach (ASI), fear of failure (ASI) and improvidence (ASI) and was negatively related to their scores on deep processing (ILP), (low) anxiety (LASSI) and test strategies (LASSI)
- the third factor reflected the students’ scores on time management (LASSI), motivation (LASSI), concentration (LASSI) and methodical study (ILP) and was negatively related to their scores on disorganized study methods (ASI).

These results seem to support the convergent validity of the ASI, the ILP and the LASSI at the level of their constituent scales and subscales. The first two factors appear to correspond quite closely to Ramsden and Entwistle’s (1981) concepts of meaning and reproducing orientations. However, the third factor is concerned specifically with an organized approach to studying.

Köymen (1992) compared 375 distance-learning students and 329 campus-based students who were taking courses in economics and business administration at two institutions in Turkey in their responses to a Turkish translation of the LASSI, which had been modified to ensure that all of the items were appropriate for both groups of students. Table 10.1 shows the mean scores obtained by the two groups of students on the ten scales of the LASSI. It will be noted that the distance-learning students obtained higher scores than the campus-based students on all but one of the scales. Köymen described the differences as ‘only moderate’ and finally concluded that there was ‘no important difference’ between the two groups in terms of their learning and study strategies (Köymen 1992: 111, 116). However, these assertions need to be properly assessed by means of appropriate statistical analyses.

From the information provided in Köymen’s article, I compared the two sets of mean scores by means of two-tailed Student’s t tests. The comparisons that are statistically significant at the 5 per cent level are indicated in Table 10.1 by asterisks. However, in Chapters 6 and 9 I noted the argument that one should use a more stringent significance level in order to guard
Table 10.1 Mean scores obtained on the Learning and Study Strategies Inventory by distance-learning and campus-based students

<table>
<thead>
<tr>
<th>Scales</th>
<th>Distance learning</th>
<th>Campus based</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiety</td>
<td>24.429</td>
<td>22.027</td>
<td>+0.35**</td>
</tr>
<tr>
<td>Attitude</td>
<td>31.753</td>
<td>29.388</td>
<td>+0.49**</td>
</tr>
<tr>
<td>Concentration</td>
<td>24.385</td>
<td>23.297</td>
<td>+0.16*</td>
</tr>
<tr>
<td>Information processing</td>
<td>30.391</td>
<td>30.603</td>
<td>-0.04</td>
</tr>
<tr>
<td>Motivation</td>
<td>27.057</td>
<td>25.559</td>
<td>+0.29**</td>
</tr>
<tr>
<td>Selecting main ideas</td>
<td>19.260</td>
<td>19.142</td>
<td>+0.03</td>
</tr>
<tr>
<td>Self testing</td>
<td>28.287</td>
<td>25.881</td>
<td>+0.45**</td>
</tr>
<tr>
<td>Study aids</td>
<td>28.123</td>
<td>26.419</td>
<td>+0.35**</td>
</tr>
<tr>
<td>Test strategies</td>
<td>27.968</td>
<td>27.293</td>
<td>+0.11</td>
</tr>
<tr>
<td>Time management</td>
<td>28.194</td>
<td>24.778</td>
<td>+0.56**</td>
</tr>
</tbody>
</table>

Source: Köymen 1992: 112
*p < 0.05, two-tailed test; **p < 0.005, two-tailed test

against an increased likelihood of making Type I errors (that is, of obtaining spuriously significant results) when carrying out large numbers of independent statistical tests (see also Harper and Kember 1986). Therefore, those comparisons that are significant using the more stringent criterion of 0.5 per cent are indicated in Table 10.1 by two asterisks. There are significant differences between the two groups of students on seven of the ten scales and six of these differences remain significant even at this more stringent level.

However, Köymen's claim that the differences are 'only moderate' and unimportant needs to be assessed against measures of the size of the relevant effects (see Richardson 1996). This matter was discussed in detail in Chapter 9 in connection with the data obtained by Vermunt (1998) in campus-based and distance-learning students using the ILS. The index of effect size used was the difference between the mean scores of the two groups, standardized against the pooled within-group standard deviation; thus, an effect size of 0.5 means that the two groups differ on average by an amount equal to one-half of their pooled standard deviation. Table 10.1 shows the standardized mean differences for the ten scales of the LASSI, corrected for sampling bias. Once again, effect sizes that are smaller than 0.2 in their absolute magnitude might be dismissed as being of little practical importance (Cohen 1969), but all the differences that are significant at the 0.5 per cent level are associated with effect sizes larger than this.

In other words, Köymen was in error in concluding that there was no important difference between the learning and study strategies of the two groups of students. On the contrary, differences that were statistically highly significant and of potential practical importance were obtained on six of the ten scales in the LASSI: the distance-learning students produced higher scores than the campus-based students in terms of their attitude, their motivation,
their self-testing, their use of study aids and their time management; they also exhibited lower anxiety than the campus-based students. Köymen was disposed to ascribe these differences to the nature of the distance-learning system itself but acknowledged that the two groups of students differed in terms of their educational experience and in their previous qualifications. (They were, however, matched in terms of their academic disciplines, as they were all taking courses in economics and business administration.)

Perhaps more important, Köymen did not mention whether these two groups of students were similar or different in the distributions of their ages. In many national systems, students taking courses by distance learning are on average older than students at campus-based institutions, and the pattern of findings obtained in this study would certainly be consistent with the general trend noted in Chapters 5, 6 and 7 for older students to demonstrate more desirable approaches to studying than younger students (see also Richardson 1994b). It is worth adding that Trueeman and Hartley (1996) obtained responses to a time-management inventory from 293 students at a campus-based university in the UK. They found that students over the age of 25 years gave higher ratings of their time-management skills than did younger students. This is scarcely surprising, because many of the older students will have been successfully juggling their various domestic and occupational responsibilities for several years before embarking on their university course. In the absence of further information, however, it is simply not possible to arrive at an unambiguous interpretation of Köymen’s findings.

Learning style inventories

From time to time in previous chapters I have used the expression ‘learning style’. This may be used to describe students’ preferences for particular kinds of learning activities and perhaps also for particular kinds of teaching or instruction (see, for instance, Jonassen and Grabowski 1993: 5). However, it is used in a wide variety of other ways to describe individual differences in the way that people learn. There is also an even broader literature concerned with ‘cognitive styles’, which was well reviewed by Riding and Rayner (1998). In a widely cited paper, Curry (1983) tried to make sense of the ways in which ‘learning style’ had been used by grouping them under three headings. The different notions varied in terms of the extent to which they could be directly observed and modified as a result of environmental influences and, as a metaphor to capture this idea, Curry likened them to progressively deeper layers of an onion:

- Instructional Preference. ‘This is the individual’s choice of environment in which to learn. We would expect this choice to be modulated by all person–environment interactions. Examples would be a preference for attending lectures versus small group learning situations . . . As this is the layer that interacts most directly with learning environments, learner
expectations, teacher expectations and other external features, we would expect instructional preference to be the least stable, the most easily influenced level of measurement in the learning styles arena.'

- Information Processing Style. 'This is the individual's intellectual approach to assimilating information... An example would be whether better retention occurred in an individual given one or other approach to hierarchies among concepts (i.e. processing generalizations followed by details, or detailed examples followed by generalized principles)... Because this processing does not directly involve the environment we would expect that measures of this Information Processing Style would be a good deal more stable than Instructional Preference, and yet still be modifiable by learning strategies.'

- Cognitive Personality Style. 'This is defined as the individual's approach to adapting and assimilating information, but this adaptation does not interact directly with the environment, rather this is an underlying and relatively permanent personality level dimension that becomes manifest only indirectly and by looking for universals within an individual's behavior across many learning instances.'

(Curry 1983: 3, 8–9)

This is a useful categorization for many practical purposes. Nevertheless, it might be noted that Vermunt's notion of learning style, which was discussed in Chapters 4 and 9, serves more as an overarching construct that appears to straddle all three 'layers' or categories in Curry's scheme. In Vermunt's account, students' learning orientations and their conceptions (or mental models) of learning are assumed to be relatively stable, and so these probably belong to the third layer. In contrast, students' choice of processing and regulation strategies is contextually determined, and so these belong to the first or second layers (see Vermunt and van Rijswijk 1988; Vermunt 1996, 1998; Vermetten et al. 1999a).

The Learning Style Inventory

In noneducational contexts, the concept of learning styles is most commonly associated with the ideas of Kolb (see, for instance, Kolb et al. 1971; Kolb and Fry, 1975; Kolb 1984). Kolb devised the Learning Style Inventory (LSI) to evaluate a person's orientation towards each of four learning modes that were believed to represent successive stages in experiential learning:

- concrete experience (CE)
- reflective observation (RO)
- abstract conceptualization (AC)
- active experimentation (AE).

In its original version, the LSI consisted of nine sets of four words and respondents were asked to rank the four words in each set as they described their own preferred mode of learning (with the best scored as 4), inserting
the appropriate number beside the corresponding word. One set of adjectives, for instance (and the one that perhaps best captures the gist of the four different learning modes), was

_____feeling _____thinking _____watching _____doing

(Kolb and Fry 1975: 37)

Six words had been chosen to represent each of the four learning modes, whilst the remaining words were distractors and were ignored in the scoring procedure. The score on each mode was obtained by summing the ranks assigned to the six words relevant to that mode. The difference between the scores obtained on AC and CE was taken to reflect different ways of perceiving the world (that is, emphasizing abstractness over concreteness), whereas the difference between the scores obtained on AE and RO was taken to reflect different ways of processing information (that is, emphasizing active experimentation over reflection). Comparisons between these two difference scores were taken to be diagnostic of four basic learning styles:

- 'The convergent learning style relies primarily on the dominant learning abilities of abstract conceptualization and active experimentation. The greatest strength of this approach lies in problem solving, decision making, and the practical application of ideas.'

- 'The divergent learning style has the opposite learning strengths from convergence, emphasizing concrete experience and reflective observation. The greatest strength of this orientation lies in imaginative ability and awareness of meaning and values.'

- 'In assimilation, the dominant learning abilities are abstract conceptualization and reflective observation. The greatest strength of this orientation lies in inductive reasoning and the ability to create theoretical models, in assimilating disparate observations into an integrated explanation.'

- 'The accommodative learning style has the opposite strengths from assimilation, emphasizing concrete experience and active experimentation. The greatest strength of this orientation lies in doing things, in carrying out plans and tasks and getting involved in new experiences'.

(Kolb 1984: 77–8).

The instructions, the items and the scoring procedure for the LSI can be most conveniently found in the book by Kolb et al. (1991: 56–7).

I have not located any published research in which the original version of the LSI was given to students taking courses by distance learning and so the following account is based on evidence from campus-based students. Freedman and Stumpf (1978) collected responses on the LSI from 1179 students taking courses in business administration. A factor analysis of their responses to the individual items identified two factors that corresponded to the (AC–CE) and (AE–RO) dimensions. Similar results were obtained by Certo and Lamb (1980) with students of business studies, by Merritt and Marshall
(1984) with nursing students and by Katz (1986) with students in nine different disciplines. In contrast, Ferrell (1983) tested students from both high schools and community colleges and obtained four factors that were claimed to match the four learning modes in the LSI, while Newstead (1992) failed to obtain a coherent factor solution in students of psychology. Freedman and Stumpf also found that the two factors that they had identified explained only 20.6 per cent of the variance in responses to the LSI and that the scores on the four scales showed poor internal consistency and test–retest reliability. They argued that these results cast doubt on the practical usefulness of the LSI in measuring students' learning styles. The LSI's poor measurement qualities were confirmed by Geller (1979) and Sims et al. (1986).

Because the original LSI used a response procedure based upon forced ranking, Freedman and Stumpf noted that the respondents' scores on the four scales were not statistically independent of one another and that this artefact might itself generate data that tended to confirm the LSI's intended scale structure. In agreement with this proposal, Lamb and Certo (1978) showed that the strongest relationships among the words that defined the two scales reflecting each of the two major dimensions of the LSI (that is, AC and CE, on the one hand, and AE and RO, on the other hand) tended to occur between pairs of words that had to be ranked against each other within the same set. To eliminate these mathematical constraints, Lamb and Certo constructed a modified version of the LSI in which respondents rated themselves on each of the 36 adjectives separately, using a seven-point rating scale. A factor analysis of the responses produced by 450 students of business studies on this modified LSI yielded very little support for the existence of two bipolar dimensions, and Lamb and Certo therefore concluded that the results obtained with Kolb's original version mainly reflected instrument bias (see also Certo and Lamb 1980).

Certo and Lamb (1979) explored this idea further by generating artificial data representing 1000 sets of totally random responses to the original version of the LSI. The correlation coefficients among the scale scores computed from these artificial data varied between −0.11 and −0.35. In the data obtained from real students by Lamb and Certo (1978), the corresponding correlation coefficients varied between +0.05 and −0.55. Certo and Lamb argued that the predominantly negative direction of these correlation coefficients resulted from the forced ranking procedure and not the underlying nature of the students' learning styles. In addition, when they examined the correlation coefficients among the responses given to the individual words, they found the same pattern that they had identified in the responses given by genuine students: the strongest relationships tended to occur between pairs of words that had to be ranked against each other within the same set. All of this evidence implied considerable instrument bias within the LSI.

Certo and Lamb demonstrated that the ranking procedure would ensure that even random data were in certain respects in accordance with Kolb's
model. However, they did not demonstrate that the randomly generated data would give rise to the appropriate factor solution for the LSI. To test this idea, I generated an artificial data set equivalent to Certo and Lamb's. The pattern of correlations among the scales and the individual items was very similar to that described in the previous paragraph. However, a factor analysis on the ratings given to the 24 critical words showed that there were 15 factors with eigenvalues greater than one, and the scree test (Cattell 1966) confirmed that this number of factors should be extracted from the data set. Inspection of these factors showed that they reflected all possible statistically orthogonal (or independent) comparisons among the words in each of the different sets. Let me spell this out in more detail:

- In five of the nine sets of words there were two distractors and only two words contributed to the scale scores on the LSI. There is only one way of comparing ratings to two words and these comparisons were represented in five of the factors obtained in this analysis.
- In two of the nine sets of words there was one distractor and three words contributed to the scale scores on the LSI. There are two possible statistically orthogonal comparisons that can be made among ratings to three words and these comparisons were represented in another four of the factors obtained in this analysis.
- In the two remaining sets of words there were no distractors and all four words contributed to the scale scores on the LSI. There are three possible statistically orthogonal comparisons that can be made among ratings to four words and these comparisons were represented in the six remaining factors obtained in this analysis.

In short, the factor analysis faithfully retrieved the underlying structure of this artificial data set. However, when the analysis was constrained to extract just two factors, these seemed to reflect the ratings given to an arbitrary group of just eight words and they certainly did not represent the two main dimensions that the LSI is assumed to measure. In short, although Kolb's ranking procedure can explain some of the properties of data obtained using the LSI, it is not possible to extract two factors representing the (AC-CE) and (AE-RO) dimensions from artificial data sets. It follows that the identification of two such factors in ratings obtained from real students (see Freedman and Stumpf 1978; Certo and Lamb 1980; Merritt and Marshall 1984; Katz 1986) cannot be dismissed as an artefact due to the use of a forced ranking procedure.

Earlier in this chapter, I referred briefly to a study by Ribich and Schmeck (1979) in which the ILP and Biggs's (1970a) SBQ were administered to students taking an introductory psychology course at a campus-based university in the US. In fact, the students also completed the LSI. Ribich and Schmeck carried out canonical correlation analyses among all possible pairs of these questionnaires to assess the degree of overlap in the constructs they were measuring. As I mentioned earlier, there was a modest amount of overlap between the ILP and the SBQ, but there was much less
overlap between the LSI and either the ILP or the SBQ. In both the latter questionnaires, certain subscales appeared to be positively related to abstract conceptualization and (to a lesser extent) negatively related to reflective observation and concrete experience. Ribich and Schmeck concluded that the LSI was sensitive to variations in depth of processing. In general, however, the LSI does not seem to be a measure of differences in approaches to studying, as reflected in the ILP and the SBQ.

A similar approach was adopted by Newstead (1992), who obtained responses from 188 psychology students at a campus-based university in the UK on both the LSI and the 18-item version of the ASI. The correlations between their scale scores on the two instruments were uniformly weak, the largest coefficients being those between AC and meaning orientation (+0.20), between CE and meaning orientation (+0.23), and between (AC–CE) and meaning orientation (+0.23). This pattern is somewhat strange, given that there was a strong negative relationship between AC and CE (–0.52). However, as Newstead himself commented, in terms of the underlying theory it is not clear why those students with a meaning orientation should tend to score more highly on the concrete experience scale. In this case, the poor overlap between the LSI and the 18-item version of the ASI could be ascribed to the basic methodological problems with the latter instrument that were discussed in Chapter 7.

The revised version of the LSI

Because of the poor psychometric properties of the original version of the LSI, Smith and Kolb (1986) produced a revised version. This consists of 12 incomplete sentences, for each of which four alternative endings are given. Respondents are asked to rank the endings for each sentence as they describe their own preferred mode of learning. Sims et al. (1986) obtained responses to the revised version of the LSI from students taking undergraduate and postgraduate courses in business studies at a campus-based university in the US. They found that the internal consistency of the revised version of the LSI was better than that of its predecessor but that its test–retest reliability over a 5-week period was just as poor. Even with only a 9-day follow-up, Atkinson (1988) found that the test–retest reliability of this instrument was far from satisfactory.

In discussing their results, Sims et al. suggested that their students 'may have been measured at a time when they were in the process of developing a particular learning style and therefore this study's finding may be due to a lack of a fully developed approach to problem solving' (Sims et al. 1986: 757–9). Indeed, in response to criticisms based on the poor test–retest reliability of the original LSI, Kolb (1981) argued that learning styles were sensitive to situational influences, in which case scores on the LSI might be expected to change over a period of time. Curry (1983) endorsed this view by placing the LSI in the second 'layer' of her account of learning styles. To
test this idea, Veres et al. (1987) gave the revised version of the LSI to 230 people employed in industry, whose learning styles were expected to be more stable than those of the students in the study by Sims et al. In fact, the test–retest reliability of their scale scores remained poor.

A criticism of the revised version of the LSI that was acknowledged by Smith and Kolb (1986: 10) themselves is that it is vulnerable to response bias. To simplify the scoring of responses, the sentence endings that reflect a particular learning style are arranged in the same column on the response sheet. As a consequence, 'The first available choice of responses to each question has to do with feelings. The second deals with watching and listening, the third focuses on thinking and logic, while the fourth emphasizes working hard, being active and getting things done' (Sims et al. 1986: 759). This could lead to a response bias in two different ways. On the one hand, the participants might stereotype themselves as people who prefer one kind of response over another. On the other hand, they might acquire a response 'set' simply to assign the same ranks to the sentence endings in particular columns regardless of their actual content.

In both cases, this kind of response bias would artificially increase the internal consistency of this instrument, in spite of its poor test–retest reliability (see Sims et al. 1986; Veres et al. 1987; Atkinson 1988). Ruble and Stout (1990, 1991) tested this possibility by devising a scrambled form of the revised LSI in which the four alternative endings for each sentence were arbitrarily reordered. The scrambled form and the standard form were given to students taking courses in business studies at several campus-based institutions across the US. The internal consistency of the scrambled form was lower than that of the standard form but its test–retest reliability was actually higher. Similar results were found by Veres et al. (1991) with a sample of campus-based students and employees. This indicates that the revised LSI is vulnerable to a response set that inflates its internal consistency but suppresses its test–retest reliability.

As I mentioned earlier, the factor structure of the original version of the LSI seemed to be fairly robust, and this could not be attributed to the mathematical constraints that are inherent in the use of a forced ranking procedure. In the revised version of the LSI, these constraints become particularly acute because there are no distractor items and all of the rankings contribute to the resulting scale scores. This means that the revised version of the LSI is an 'ipsative' instrument: that is, one in which the total score is the same for all respondents. In this case, the total score for each sentence will be \((1 + 2 + 3 + 4)\), or 10, and the total score on the instrument as a whole will be \((1 + 2 + 3 + 4) \times 12\), or 120. This gives rise to problems in the interpretation of factor analyses, as bipolar factors can emerge simply as an artefact (Dunlap and Cornwell 1994).

Ruble and Stout (1990) carried out factor analyses on the ratings given to individual items in both the standard and the scrambled form of the revised LSI. Both two-factor and four-factor solutions were obtained for comparison. Analysis of the standard form generated a two-factor solution in which
one factor reflected the AC scale and the other factor reflected the CE scale. A four-factor solution contained factors reflecting the AC, AE and RO scales, respectively, but the CE items were spread across two different factors. Analysis of the scrambled form led to a two-factor solution in which one factor compared AE with CE and the other factor compared RO with CE. The four-factor solution was similar to but less distinct than that obtained with the standard form and this was taken as further evidence that the latter was vulnerable to response bias. In short, there was very little support for the intended structure of either instrument.

Subsequent investigations using the standard form of the revised LSI reached essentially the same conclusion: there was, indeed, a consistent trend for factors to emerge that contrasted AC with AE and CE with RO (Cornwell et al. 1991; Geiger et al. 1992, 1993; Loo 1996). This may be an interesting pattern but it is not in accord with Kolb's theoretical model. Loo (1999) reanalysed the ratings that he had obtained from 200 students at a campus-based university in Canada. Instead of using exploratory factor analysis on the instrument as a whole, he used confirmatory factor analysis on the items defining each of the four scales and the two underlying dimensions (AC–CE) and (AE–RO). In each case, the data were a very poor fit to these scales and dimensions on a number of different statistical criteria.

To eliminate the ipsative property of the instrument, Geiger et al. (1993) devised an alternative, 'normative' form of the revised LSI in which respondents rated 48 randomized sentences along a seven-point scale from 'not like me', scoring 1, to 'very much like me', scoring 7. This was given to students taking courses in business administration at two campus-based universities in the US, along with the standard, 'ipsative' form of the instrument. Both the normative and the ipsative forms demonstrated satisfactory internal consistency and there were moderate correlations between their corresponding scales. A factor analysis carried out on the students' ratings of the 48 sentences produced four factors that could be readily identified with the four scales, but a two-factor solution did not support the intended relationships among these scales. Geiger et al. concluded that these were worth exploring as distinct learning abilities.

Geiger and Pinto (1991) gave the revised version of the LSI to 40 students of business studies at a campus-based university in the US during the fall semesters of their sophomore, junior and senior years (in other words, in Years 2–4 of the undergraduate programme). The test–retest reliabilities of the four learning modes were all greater than 0.50, except for those on concrete experience, which were very low. There were small changes in the students' mean scores on the four scales over the 3 years, which were not in themselves statistically significant, but there was a significant change in the classification of their learning styles (assimilator, accommodator, converger and diverger). In particular, 17 of the 40 students were assigned to different learning styles in their senior year compared with their sophomore year. This is consistent with the idea that the experience of higher education
(and, presumably, other environmental influences, as well) can lead to changes in students’ learning styles.

Dille and Mezack (1991) administered the revised version of the LSI to 151 students who were embarking on distance-learning courses delivered by television programmes with support from study guides, textbooks and other printed materials. They then compared the students who were academically successful (in that they completed their course with a grade of C or better during their first semester of enrolment) with those who were unsuccessful (in that they withdrew from their course or failed to obtain a grade of C or better) in their scale scores on the revised version of the LSI. The unsuccessful students tended to have higher scores on concrete experience (CE) and consequently lower scores on (AC–CE) than the successful students. Dille and Mezack argued that people with high scores on CE had a greater need to relate to other people, and that this was frustrated by the social isolation that was inherent in distance education.

**Canfield’s Learning Style Inventory**

Canfield (1980) devised a similar instrument to Kolb’s LSI that was also named the ‘Learning Style Inventory’ (CLSI). This consisted of 30 sets of four words and in each case respondents once again ranked the words to describe their preferred mode of learning. This generated scores on 21 measures: eight scales represented ‘conditions’ or different dynamics of learning, four scales represented ‘content’ or major areas of interest, four scales represented ‘mode’ or the general preferred modality for learning and the remaining scales represented ‘expectation’ or the anticipated level of academic performance (Canfield 1980: 22).

Alsagoff (1985, 1986) obtained responses to the CLSI from a total of 629 students taking one of four different degree programmes by distance learning in Malaysia. There were statistically significant differences on some of the scales across the four degree programmes and there were significant correlations between the students’ scores on certain of the scales and their current academic performance. In particular, performance was higher in students who affiliated with their peers rather than their instructors, higher in students who preferred to learn by reading rather than by listening and higher in students who expected to achieve good performance.

Coggins (1988; Gibson and Graft 1992) found that scores on the CLSI could be used to predict course completion in 153 students on distance-learning courses at a university in the US. A discriminant analysis showed that course completion versus non-completion could be correctly predicted in 70 per cent of these students on the basis of their scores on the scales of the CLSI alone. In particular, course completion was higher in students who expected to obtain outstanding results and lower in students who expected to obtain average or satisfactory results. Completion was also higher in students who preferred to learn about people rather than about inanimate
things. There were no statistically significant differences associated with conditions of learning or with modalities of learning, although satisfactory course completion tended to be associated with a lower preference for peer affiliation.

**The Learning Style Questionnaire**

Because of possible problems due to mathematical constraints in the forced ranking procedure employed in Kolb’s LSI, Marshall and Merritt (1985, 1986) developed another instrument, the Learning Style Questionnaire (LSQ-MM), which incorporated three different modifications to the LSI. First, the respondents were asked to rate themselves on individual words along a five-point scale. Second, the ends of the scale were labelled with pairs of words that were polar opposites (for instance, accepting—questioning). Third, the items intended to measure the four learning modes were reconstructed afresh by selecting 40 pairs of words from a sample of 100 pairs.

Earlier in this chapter, I mentioned an investigation by Cano-Garcia and Justicia-Justicia (1994) in which a battery of questionnaires was administered to students at a campus-based university in Spain. The battery included the ASI, the ILP and the LASSI and a factor analysis indicated that there was considerable overlap in the constructs being measured by these three instruments. In fact, the battery also included the LSQ-MM and Cano-Garcia and Justicia-Justicia also carried out a factor analysis on the scores on all four instruments. Scores on the four scales in the LSQ-MM defined two factors that reflected the (AC–CE) and (AE–RO) dimensions, respectively.

This provides confirmation of the underlying structure of this instrument in terms of Kolb’s experiential learning theory. However, the factors that reflected the dimensions underlying the LSQ-MM were related only marginally to the students’ scores on the other three instruments. In addition, their scores on the LSQ-MM showed no significant loadings on the factors that were defined by the other three instruments. These results are wholly consistent with the implication of the study carried out by Ribich and Schmeck (1979), that whatever the LSI measures is not related to individual differences in approaches to studying.

**The Learning Styles Questionnaire**

Honey and Mumford (1982) adopted a different approach to measuring four learning styles that were intended to be broadly equivalent to the stages in Kolb’s account of experiential learning. Their Learning Styles Questionnaire (LSQ-HM) was mainly intended as a pedagogical tool for use by staff developers working in the area of management training. It consists of 80 statements describing learners’ beliefs, preferences and behaviour
and, in each case, respondents are asked to indicate whether they agree or disagree with the relevant statement. There are 20 statements relevant to each of four learning styles:

- 'Activists will learn best from activities where they can engross themselves in short here-and-now activities such as business games and competitive teamwork tasks.'
- 'Reflectors learn best from activities where they are able to stand back from events and listen and observe.'
- 'Theorists learn best from activities where what is being offered is part of a system, model, concept or theory.'
- 'Pragmatists learn best from activities where there is an obvious link between the subject matter and the problem or opportunity on the job.' (Mumford 1993: 4)

The respondents' score on each of the four learning styles is obtained by counting the number of 'agree' responses. Honey and Mumford presented results from 29 participants who had completed both the LSI and the LSQ-HM. These showed that there were moderately strong associations between the corresponding scales of the two instruments, except that the concrete experience scale of the LSI was only weakly correlated with the activist scale of the LSQ-HM.

Allinson and Hayes (1988) presented results from 95 managers from developing countries who were taking a Master of Business Administration (MBA) programme at a British campus-based university and from 127 managers from British industry. In both samples, a factor analysis of the participants' scale scores on the LSQ-HM identified two factors: an 'analysis' factor that measured scores on theorist and pragmatist, and an 'action' factor that contrasted scores on activist and reflector. Allinson and Hayes therefore calculated scores on these two dimensions and found that they had satisfactory internal consistency and test–retest reliability. They argued that the LSQ-HM might be preferable to the LSI as a measure of learning style.

Allinson and Hayes (1990) obtained a similar factor structure in the scores obtained by 138 students taking courses in management at a campus-based university in the UK. In a second study, involving 21 MBA students from developing countries, they found no difference in the scores obtained by those who preferred to be taught in a conventional lecture format and those who preferred a 1-day simulation exercise. However, in a third study involving 29 male managers from a British-based multinational corporation, they found that participants who had been identified by their employers as 'high fliers' obtained higher scores on both the 'analysis' factor and the 'action' factor. This was taken to support the construct validity of the LSQ-HM.

Sims et al. (1989) obtained responses to both the LSQ-HM and the revised version of the LSI from 279 students who were taking courses in business studies at a campus-based university in the US. The scales of the LSQ-HM showed levels of internal consistency that were satisfactory, although
inferior to those of scales of the LSI. Nevertheless, the associations between the corresponding scales of the two instruments were relatively weak, the highest correlation coefficient being 0.28. Sims et al. concluded that there was little evidence of any convergence between the two instruments, although this might be partly because there were problems with the ways that both instruments attempted to operationalize the construct of learning styles.

Goldstein and Bokoros (1992) argued that comparisons of this sort were inappropriate because in Kolb's LSI learning styles (that is, converger, diverger, assimilator and accommodator) were captured by particular combinations of scores on the scales rather than by the scales themselves. They obtained responses on both the original version and the revised version of the LSI and on the LSQ-HM from 44 campus-based students in the US. When these participants were classified in terms of their learning styles on the three different instruments, only 41 per cent were assigned to equivalent styles on the original LSI and the LSQ-HM, and only 30 per cent were assigned to equivalent styles on the revised LSI and the LSQ-HM. Goldstein and Bokoros concluded that the degree of coincidence between the different classifications was only modest.

Sadler-Smith (1997) administered a battery of questionnaires to 245 business studies students at a campus-based university in the UK. The battery included the LSQ-HM and the RASI (see Chapter 7). Sadler-Smith found several statistically significant correlations between the students' scale scores on these two instruments:

- activists tended to produce lower scores on strategic approach
- reflectors tended to produce higher scores on deep, surface and strategic approach
- theorists tended to produce higher scores on deep and strategic approach and lower scores on lack of direction
- pragmatists tended to produce higher scores on deep approach, strategic approach and academic self-confidence.

In addition, the students' scores on both of these instruments showed some relationships with their overall academic performance: specifically, good performance was associated with higher scores on the theorist scale from the LSQ-HM and on the deep approach, strategic approach and self-confidence scales from the RASI. Sadler-Smith concluded that there might be some overlap between the LSQ-HM and the RASI, and he speculated that this included 'constructs such as motivation, learning process and degree of learning activity' (Sadler-Smith 1997: 61). Provided that this pattern can be replicated, it would seem that in this respect the LSQ-HM is different from other measures of learning styles, such as the LSI and the LSQ-MM, which tend to show little overlap with instruments intended to measure approaches to studying (Ribich and Schmeck 1979; Cano-Garcia and Justicia-Justicia 1994; see also Murray-Harvey 1994).
Concluding summary

- The ILP was developed on the basis of prevalent notions in psychological research into human memory during the 1970s. However, its psychometric properties are questionable, and several studies have failed to replicate its intended structure. The 'levels of processing' framework on which it was based is no longer regarded as useful for understanding human memory and so the ILP has no advantage by virtue of its supposed theoretical underpinning.

- Many writers are sceptical about the idea of study skills and the value of study skills courses. Nevertheless, the LASSI seems to provide a useful way of assessing learning strategies and shows convergence with both the ASI and the ILP. Distance-learning students have better learning and study skills than campus-based students taking similar courses but this may be due to confounded differences in age, educational experience or previous qualifications.

- The LSI was designed to measure a person's orientation towards four learning modes. It is subject to several methodological criticisms, despite having undergone revision in the 1980s. The revised LSI and the CLSI may be useful in predicting outcomes in distance education. However, with the possible exception of the LSQ-HM, learning style inventories show little overlap with questionnaires designed to measure approaches to studying in higher education.