DOCUMENT RESUME

ED 103 207 SE 017 067

AUTHOR	Tisher, R. P., Ed.
TITLE	Science Education: Research 1973.
INSTITUTION	Australian Science Education Research Association.
PUB DATE	May 73
NOTE	170p.; A publication containing papers presented at the Annual Conference of the Australian Science Education Research Association (4th, Brisbane, May 19-21, 1973). For the 1971 and 1972 editions, see ED 065 325 and ED 076 365
AVAILABLE FROM	Dr. R. P. Tisher, Professor of Experimental Education, Monash University, Clayton, Victoria, Australia 3168 (no price quoted)
EDRS PRICE	MF-$0.76 HC-$8.24 PLUS POSTAGE
DESCRIPTORS	Attitudes; Curriculum; Educational Research; International Education; Research Methodology; *Research Reviews (Publications); Science Education; Secondary School Science; Teacher Behavior
IDENTIFIERS	Australia

ABSTRACT

This publication contains papers organized into four groupings: National and International Issues in Science Education; Teaching Strategies, Attitude and Enquiry Measures, and Surveys; Analyses of Curricula; and Research Techniques. (BR)

Richard Tisher

SCIENCE EDUCATION : RESEARCH 1973

Edited by

R. P. TISHER

A publication containing papers presented at the Fourth Annual Conference of the Australian Science Education Research Association, held in Brisbane, 19th-21st May, 1973.

Australian Science Education Research Association

Published by Australian Science Education Research Association 1973.
Printed by the University of Queensland Printery.

Enquiries relating to *Science Education : Research 1973* should be addressed to:

The Editor,
Science Education : Research 1973,
Faculty of Education,
University of Queensland,
BRISBANE, QUEENSLAND. 4067

CONTENTS

Page

PREFACE

Each year, for the past four years, the Australian Science Education Research Association, has been able to claim "a first". In 1970, the Association held its first annual conference, in 1971 its first publication appeared; in 1972, the proceedings of the annual conference appeared in a new format; and, this year we had our first international visitors who contributed to our annual conference. It is probably significant, too, that the annual meeting decided to change the name of the present publication.

Science Education : Research 1973 represents another significant step forward for the Australian Science Education Research Association. Not only is the publication a larger one than the previous issues of *Research 1971* and *Research 1972* indicating an increase in research activity but, it contains reports covering a wider range of issues in science education, and a special section dealing with research techniques. This last mentioned section was included in the conference with the express objective to develop our own understandings of some techniques which can be of great assistance to the science education researcher.

For convenience of presentation the papers have been grouped under headings, or themes. Although the responsibility for the groupings must again rest with the editor, it did seem that clusters fell, quite naturally, under the several headings. One of these was the Analyses of Curricula and the four papers related to this theme indicate a new area of research activity for Australian science educationists.

Science Education : Research 1973, though still a modest publication, is ample evidence of an increase in the number and quality of research projects in Australia. It demonstrates too, that we have international interests and ties. If the nature and quality of the papers contained in this publication are any guide, the future for science education research in Australia, looks brighter, if more challenging, than ever before.

R. P. Tisher
University of Queensland

PART I

NATIONAL AND INTERNATIONAL ISSUES IN SCIENCE EDUCATION

A NATIONAL APPROACH TO SCIENCE EDUCATION RESEARCH

Malcolm J. Rosier

My aim in this paper is to develop a framework within which to consider educational research, with particular reference to science education in Australia. An examination of this framework may enable us to identify areas where more research is indicated. This approach may assist us to assign priorities and to rationalise our limited resources and may lead to the development of cooperative national projects.

In advance may I say that it is not easy for me to distinguish peculiar characteristics of science education separate from education generally. We are constantly aware that education is a complex matter, and the problems and activities of science education are caught up in this complexity.

May I also say that the definition of an educational research worker implied in my remarks is a broad one. It includes anyone using statistical procedures to make some sense out of educational data. For example, it includes people in Education Departments analysing enrolments, curriculum developers analysing results gathered during trial testing, etc.

I suppose that, as part of these introductory remarks, I should also define education, but will largely abdicate this task. Suffice it to say that education is concerned with both the cognitive and non-cognitive development of people. I will be limiting my comments to formal education in primary and secondary schools, leaving the tertiary sector to solve its own problems.

The needs of students

I decided to develop a framework for science education research based on needs. The basic needs are those of the students of science.

I regard it as salutary and non-trivial to affirm that teachers, schools, Education Departments and even science education researchers do not exist or work for their own sakes, but in a service role to help to meet the legitimate needs of the students. That is why I choose to regard the needs of the students as basic to my framework.

Currently, the main needs of the students are perceived to be cognitive. Politicians recognise this since education systems are funded to increase the knowledge (and employability) of the students. Parents recognise this — they ask their children "What did you learn today" not "What new attitude was inculcated in you today". Sometimes, even the students recognise that they are at school to learn!

I will therefore concentrate mainly on students cognitive needs, not forgetting that there are interactions with many non-cognitive factors as well.

The main questions I pose, and attempt to answer are

> 'What are the students' cognitive needs, especially in science?'
> 'How are they being met at present?'
> 'What research is indicated to improve the situation?'

Let us assume that students accept that they are studying science at school in order to learn. We therefore assume that they thirst after science knowledge and understanding. Their basic need is to have their thirst adequately quenched. They wish to have satisfying learning experiences which are relevant to their intellectual development and to their life in the larger society beyond the school.

The need of students to know what they know

One need of the students is to know what they know. They need to develop a frame of reference for their learning experiences. If they have been studying a concept they need to know the extent to which they have learnt it.

They need to know what they know in both relative and absolute terms. In relative terms we are dealing with normative evaluation. The science student needs to know how he compares with his colleagues. This is especially important where selection procedures exist which are based on achievement; for example, matriculation results as a basis for entry to tertiary studies, or school-awarded certificates as a basis for employment.

At this point I will not discuss the desirability of this competitive thrust in the school and society, but note that it exists and that one of the students perceived needs is to be equipped to play the game according to current rules.

Even in a cooperative atmosphere, such as individualized mathematics and reading programs at primary school, some students still act competitively and often enjoy doing so. Learning theory tells us that mild anxiety facilitates learning.

Students need some normative evaluation. This indicates a need for some standardised tests with normative data. I suggest that the range of achievement tests in science, properly constructed and normed for Australia, could be increased. Much of the construction could be done at local levels, with perhaps national coordination to avoid unnecessary duplication of effort. If national norms are needed, this requires a national project.

Of more importance than normative evaluation is meeting the student's need to know what he knows in absolute terms. This leads us to the idea of mastery learning. In essence the learning theory behind the idea is that most students can learn most topics given adequate instruction and adequate time. These ideas are summarized in Block (1971) and expounded in Bloom, Hastings and Madaus (1971).

The best simple statement I know about the value of mastery learning is implied by the analogy between achievement and pregnancy given by Coffman (1969), and which I paraphrase: achievement is a tangible product, the achiever has some choice in the activity, the achiever has immediate knowledge of results, the evaluation is made by the achiever, and the judgment is absolute not relative.

The evaluation of mastery can be self-administered and corrected, or teacher-administered and corrected, or a mixture of both types. The corresponding research need is for the development of many curriculum units and associated mastery tests based on mastery learning ideas.

At the same time there is a need for mastery tests for use within current school contexts. These tests will usually have diagnostic and remedial aspects also.

I think that it would be difficult to organise or direct this research nationally, since I would expect many spontaneous efforts, often from unexpected places. It is desirable that information about these units and tests, at the planning, development and production stages, be widely disseminated. It would be valuable if the persons responsible for the information dissemination could also conduct a survey into the nature of effective introduction of these innovative ideas.

The need of students to know what they know, especially in the mastery sense, is also very important for their non-cognitive development. A student's image of himself or herself as a person (self-esteem, self-concept) depends in part on achievement. "The research findings reveal a clear, perhaps causal, relationship between a student's academic performance and progress and both his self-concept and his mental health." (Block, 1971, p.95) Given the complexity of today's and tomorrow's society, the school has an obligation to make as large a contribution as it can towards developing a positive self-image in students so that they can cope adequately with these complexities.

The need of students to learn efficiently

The next student need is to learn efficiently. Teaching is both an art and a science. I hesitate to comment on the artistry of teaching. The science of teaching is a different matter. I see it as a correlate of his professionalism that a teacher should be a master of what is known about the science or technology of teaching. "It is obvious that the effectiveness of teaching is directly related to the state of knowledge about the learning process and the teacher's effectiveness in applying it." (Rowlands, 1967, p.15) Or simply: "Many science teachers teach poorly because of their limited conception of how learning takes place." (Balson, 1973, p.5)

Learning theory is complex but there are certain general ideas for which there is enough empirical evidence to regard the ideas as laws of learning.

On this matter a recent paper by Bloom (1972) is relevant. He says that many educators complain that nothing is really known about the educative process and boast about their innocence.

Bloom maintains that there have been advances in our understanding of education and related phenomena. These new insights represent loss of innocence: ". . . the burden of responsibility for appropriate actions and practices rests with the professionals in the field once new ideas are adequately communicated." (Bloom, 1972, p.349)

So my suggestions for research in this area focus on the need to implement effectvely in the science classrooms of Australia what is already known of the science of teaching. As in my previous section, I press for the study of strategies of change in attitudes, hange in teacher behaviour, etc.

There are two areas where I see that more information is needed to help the tudent to learn more efficiently.

Without wishing to espouse any particular model, it seems obvious that the more we know about learning hierarchies, and the more actual hierarchies we investigate and construct within a science topic, then the more efficiently we can help students to learn he topic.

This research level lends itself to national coordination linked to a great deal of ocal activity. In particular, at the national level key conceptual ideas and models could be worked out, complete with literature surveys, etc. Basic research methodologies could be leveloped at the national level and published as in a technical report handbook form, which could be used by an alert science classroom teacher to prepare and test a hierarchy in a limited area of interest.

The other area for work on learning efficiency is ATI (aptitude-treatment interaction). The term is borrowed from Cronbach and Snow (1969). In other words, which treatments (teaching methods, etc.) are best for which students for which curriculum topics under which conditions. Under treatment I include the range of audio-visual media also.

I think we need to tackle much more systematically the question of how to assist the student to accomplish effectively and satisfactorily the components of his curriculum.

I have been upset for several years by the sight of thousands of teachers re-inventing wheels, and the energy thus dissipated which could be spent more usefully. By this I mean that a teacher often prepares his own set of notes, examples, tests, assignments, teaching aids, etc. in isolation from the efforts of his colleagues in the same school and elsewhere.

I am not advocating that the pendulum should swing the other way with all teachers using standardised packages. Nor would I wish to restrict the professional responsibility of the teacher to tailor his offerring to the perceived learning needs of his students. But if a carefully prepared set of materials is shown to be effective in helping students to learn a given topic it seems that the onus lies with the teacher to demonstrate that the students' learning needs are met better by the teacher's idiosyncratic approach than by the prepared curriculum materials.

If each science teacher tries to be 'all thing to all men' then his expertise will be spread thinly. It seems to me that he would be better advised to concentrate his expertise on a particular topic, contribute the results to a 'pool', and draw freely on the rest of the 'pool' for his other topic requirements. There are also important attitude changes for the teacher associated with this 'contributory openness' as compared to 'idiosyncratic self-sufficiency'.

Under this section I should include large scale curriculum materials development projects such as ASEP (Australian Science Education Project). The work of ASEP covers, in a more formal way, the ideas I have been stating about cooperative development of learning materials.

A thorough evaluation of ASEP is needed, if only as an accountability exercise, because large sums of public money have been spent in Australia on this innovative activity.

Without defining what could be covered by the evaluation of ASEP, it seems important to coordinate the activities nationally while encouraging local research studies.

The need of students to have their questions answered

Another need of a science student is to have his questions adequately answered by his science teacher.

At one level this means that the science teacher must be competent in his own knowledge and understanding of science. Here is a research question: 'How much competence in science is enough for which students?'

The student's questions may arise from within a formal curriculum, from within a general studies informal problem-orientated curriculum, or from within a student's out-of-school experience.

It is a need (or reasonable expectation) of the student that the teacher should be able to answer such questions. The answer may be a direct factual answer, as if the student had asked the right scientist or opened the right page in a book.

However the teacher is employed as a teacher not as an encyclopedia. The answering of the question may be related more to the managerial role of the teacher.

I regard a science teacher primarily as a manager of learning experiences in science, not as a scientist. I have not heard of any research in Australia on task analyses of science teachers. Here is another set of research questions: 'What types of tasks fill the science teacher's day (and the working part of his night?) What are the relative proportions of these tasks? Which of these tasks are associated in which ways with effective and satisfactory student learning experiences?' 'How should the teaching behaviour of the science teachers be modified to improve student learning?'

May I digress to explain what I mean by teacher behaviour modification, based on a visit I made to an individualized mathematics instruction project in Pittsburgh (IPI: Individually Prescribed Instruction). The teachers in the scheme were free to deal with the problems of individual students while the rest of the class proceeded with their own work. The researchers had noted that the teachers tended to spend several minutes with each enquiring individual, which meant that only a few students met the teacher each day. As a corollary there were some timid students who hardly ever spoke with the teacher.

The researchers were in the process of training the teachers *not* to spend much time with each student. The teacher was being trained to lead the student to the next step only, throwing the onus on the student to pick up the learning from there. The teachers were also trained to make some work-related contact with each student in the class each period.

Let me also mention some student behaviour modification from the same project! After each section the students did a test and took it to the clerical assistant to be marked, which took some time. So the smarter students took tests prematurely, and submitted them to marking. Of course the tests were returned as unsatisfactory. The re-test, re-mark/ re-test, re-mark/ etc. cycle was a long one with plenty of rest periods for the students!

This brief comment on teacher-student contact provides an introduction to the important field of teacher-student interactions. The key research questions are: 'What are the relationships between personality and behaviour characteristics of science teachers and effective student learning?' 'How may teachers' behaviour be modified, either preservice or inservice, to maximise the development of the desirable characteristics and to eliminate undesirable ones?' The survey by Rosenshine (1971) is a valuable introduction to this area of study.

I will turn to a particular aspect of teacher-student interaction. I said earlier that a need of the science student is that the teacher should answer his questions.

You will be aware that questions may be asked on several levels. So far I have dealt with questions taken at face value. The student thirsts for knowledge and the teacher meets the need.

Often a question has a deeper meaning. The words of a question of a primary student may say: 'Why does electricity turn a motor?' The real question beyond the words may ask: 'Are there strange forces which really run this world?' Or it may ask: 'Are you a mature person with whom I, as a growing person, can establish a meaningful relationship that will help me to grow up and understand this complex world?'

A young adolescent may ask: 'What are the effects of smoking cigarettes?' The latent question may ask: 'How do I make decisions about my life-style which are both rational and satisfying?'

The science teacher has two responsibilities with respect to these questions. Firstly he has a responsibility as a teacher to be sensitive enough to his students to hear the real question, which may be the latent one. Secondly he has a responsibility as a teacher to attempt an answer. As I said earlier, one of the legitimate needs of the students is to obtain answers to their questions.

To my mind this area of teacher sensitivity is very important. Research in the area will also require great sensitivity. The research questions are: 'How can we identify, categorise and measure teacher-student sensitivity situations?' 'How can we select teachers with the ability to deal adequately with these situations?' 'How can we modify the behaviour of all teachers to ensure that they have a minimum level of adequacy in sensitivity situations?' 'What is meant by dealing adequately with a sensitivity situation?'

The need of students for career guidance

I will briefly comment on another area of student needs.

Students express their need to obtain a 'good job' at the end of their schooling. They and their parents expect an appropriate certificate when the students leave school. This is one of their needs.

Unfortunately there is often dissonance between student, parent and teacher attitudes on this future career aspect of the work of the school.

At ACER I am working on a follow-up study, in which we collected information in 1972 from the 16-year-olds who were members of our 1970 sample of 14-year-olds (Population II).

The students were asked how important it was to their teachers, their parents and themselves that they got good results in tests and exams.

The data from our pilot study indicate considerable dissonance.

Table 1 Importance of good results in tests and exams

	Not at all important %	Slightly important %	Fairly important %	Very important %
To your teachers	19	26	36	19
To your parents	1	1	22	76
To yourself	1	1	12	86

(Number of students = 194)

A similar question inquired into the press upon students to make career decisions.

Table 2 Importance of deciding future occupation by the end of the year

	Not	Slight	Fair	Very
To your teachers	47	29	14	10
To your parents	11	11	31	47
To yourself	11	5	22	62

(Number of students = 190)

For further insight into this need, let us recognise the changing nature of the secondary school population. The retentivity has increased markedly. Students now see secondary education as a right. Only a few decades ago it was a privilege. The incr[illegible]ing number of secondary and tertiary students leads to the idea of the 'mass elite' (Bowman, 1970).

I will quantify the changing retentivity picture by supplying some data we collected as part of the IEA Science Project. I define retentivity as the proportion of the age cohort who are at secondary school.

Table 3 Retentivity of Australian students in 1950, 1960, 1970

		NSW	Vic	Qld	SA	WA	Tas	Aus
		%	%	%	%	%	%	%
15-year-olds	1950	37	33	28	34	30	49	35
	1960	54	60	57	62	52	71	57
	1970	81	86	71	84	71	83	81
16-year-olds	1950	15	16	13	14	18	17	16
	1960	26	32	29	34	26	30	29
	1970	52	62	36	57	36	43	51
17-year-olds	1950	6	6	5	5	5	7	6
	1960	9	12	13	11	13	11	11
	1970	31	33	18	25	18	20	28

Source: IEA Age-grade tables incorporating various estimates.

It is an important research problem that we investigate how to change the attitudes and behaviour of all the parties in the educational enterprise teachers, students and parents in a way that rationalises their expectations and avoids the dissonance which currently interferes with the effectiveness of the enterprise.

You will notice in my last few statements I have consciously advocated changing teacher behaviour. I believe this to be consistent with my earlier statements that there is a science or technology of teaching, albeit immature, of which the professional science teacher is obliged to be a master.

This leads me to comment that those with a responsibility help train teachers, that is, to change teacher behaviour, also have a responsibility to carry out follow-up analyses of their trainees. This research procedure will be informal for most teacher educators, and will become an important research exercise for some. The nett result will be the development of feedback-type models where training techniques, etc. are modified as a result of observing the behaviour of the trainees.

Up until now I have mainly dealt with achievement, and also mentioned various non-cognitive learning outcomes. It may be possible to deal with aspects such as sensitivity or self-esteem as isolated factors, for example, in preparing measuring instruments. It is not possible to identify useful relationships at this simple level.

As I keep saying, education is very complex, and valid simple relationships do not occur. Most studies of factors effecting achievement will be complex, and this implies careful samples and multivariate analyses. The recent book by Keeves (1972) provides an example of the statistical treatment of the kind of complex educational relationships I am talking about.

The needs of science teachers

You will recall that I decided to adopt a framework of needs to identify possible science education research activities.

Let me turn, in less detail than for student needs, to the needs of science teachers.

As shown by Australian results from the IEA Science Project, most science teachers feel the need for refresher courses in science. (Rosier, 1973)

They feel the need both for courses in science itself and in science teaching methods. There are indications that they could show more initiative and exert more effort themselves on their professional development.

I would like to see some research on strategies to inspire teachers to improve their professional competence. Is it a mixture of certification carrot and salary stick? If so, what is the recipe?

One approach I have seen is that developed by the US National Science Teachers Association (1970). They prepared ASIST *Annual self-inventory for science teachers.* This in turn followed from work in setting up a statement: *Conditions for good science teaching in secondary schools.* (NSTA, 1970)

Members of NSTA wishing to participate in the exercise complete their ASIST form. Each respondent's data is incorporated into summary data. The respondent receives a copy of his own results and the summary results. He can compare himself against his peers, and presumably act on areas where his behaviour is different from the behaviour of many colleagues.

A similar idea is being developed for dentists (Educational Testing Service, 1973).

Each participating dentist subscribes (for a $40 annual fee) to a series of self-administered exams at 3-month intervals. Test results are confidential, and are sent only to the individual dentist. He also receives summary information to indicate his standing relative to his peers.

Teachers express a need for good conditions. These conditions are not all needed to an equal extent. The key conditions are those needed for effective student learning. There are other conditions which are necessary (or desirable?) for teacher morale. As with poverty, I suspect that definitions of adequacy are relative, but there is need for careful research to establish a statement about adequate conditions for science teaching to replace arbitrary political-type statements of class size, room size, number of teaching periods, etc.

The needs of the system overall

I have defined the final set of needs as system needs. By system I refer mainly to a national or state system of schools, but some comments apply to the individual school as a system.

When I say the system has needs I am really saying that the politicians who supply the funds for education have certain needs which must be met if they are to obtain and distribute funds effectively. Their basic need is to know how effectively the money they provide is being used to purchase efficient learning experiences for the students in the system.

One approach is to count the number of teachers and students, and say that because there are X teachers confronting Y students daily, therefore learning must be occurring.

A more empirical approach is to measure the extent to which the aims of the system are met. This is difficult since Australian education systems do not articulate their aims. If the basic aim is student achievement, as I inferred at the beginning of my paper, the system should be evalaated on the basis of the achievement of the students.

The achievement measured should not be an 'absolute' achievement but one adjusted for characteristics of the students in the system, especially home background characteristics.

System evaluation requires large scale studies, although there are dangers such as those pointed out by Orlans (1972). He states that other researchers can challenge the findings of a small study, by replicating it. A large study cannot be replicated; the data are unique, and can only be challenged on the basis of internal weaknesses which are not or cannot be concealed.

One large scale system evaluation study is NAEP (National Assessment of Educational Progress).

The NAEP science project started by preparing a set of aims in science for 9-year-olds, 13-year-olds, 17-year-olds and young adults. (National Assessment of Educational Progress, 1969). They then prepared 'exercises' (that is, items) to test these aims.

The only scores presented are the percentages of students correctly answering exercises. The exercise scores are not combined into test scores as we know them.

Their first results, now published (National Assessment of Educational Progress, 1970) show the percentages of students achieving mastery of various objectives. Sub-groups are also investigated for example, divided by sex, race and geographical location.

NAEP is a type of large scale mastery learning evaluation at the system level.

If the NAEP results are regarded as norms, an individual group such as a school board can test its own students against the NAEP norms to evaluate its cognitive performance as an educational system.

Should Australia carry out a national evaluation of science education based on the NAEP model?

The other type of system evaluation focusses more on the reasons for achievement. This is the tradition of the Coleman Report (Coleman and others, 1966) the Plowden Report (Central Advisory Council for Education, 1967) and the IEA Mathematics and Science Reports (Husen, 1967; Comber and Keeves, 1973).

These large scale studies offer the opportunity for analyses between schools as well as between students, as demonstrated in Rosier (1973).

These studies of a type which I describe as 'achievement and its correlates' studies also lend themselves to a normative use.

Any school would be able to administer the test instruments to a sample of students, teachers and parents in its own school. Its results could then be compared with national or regional norms. The school could compare its achievement with other schools in absolute terms. Preferably it would compare adjusted achievement scores; that is, achievement scores adjusted for various student, teacher and parent characteristics.

Since the analyses are complex, any self-auditing school would have to send its test and questionnaire data to the 'project headquarters' for processing, and return to the school.

This means that the project headquarters would be able to continually update and expand its data base.

Another 'spin-off' advantage is that it would be possible to identify deviant schools schools whose adjusted achievement scores are rather higher or lower than expected. A special research study of these schools could produce very useful findings.

Large studies can also offer "piggy-back' facilities. An isolated researcher can add his question or test to the main testing program. He does not have to worry about the organisation of sampling or data processing, and yet he will have access to all the other test and questionnaire data collected by the main survey to link with his own questions.

It is very difficult to estimate levels of national achievement from a single survey. Repeat surveys, or the 'rolling' survey I have just mentioned, are necessary to gauge whether national achievement levels have changed or are changing.

Related to repeat surveys are longitudinal surveys, in which the same students are followed for several years. A classical example is given in Husén and Boalt (1967) where students were first tested in 1938 and were subsequently contacted and retested on various occasions.

The advantage of a longitudinal study is that persons can be studied over time to trace their cognitive and non-cognitive development, and their career paths.

This is an attractive idea for a national science education research project since many aspects of science education could be tested and their consequences investigated and evaluated over a longer term than is usually possible.

Conclusion

In conclusion may I summarise. There are certain large scale science education research projects which require a team of researchers. The size of the team will be relatively large by previous Australian educational research standards. These studies of the national science achievement evaluation type require special expertise in sampling and data handling.

On the other hand there are many science education research activities which lend themselves to decentralised data collection and analyses. The great need at the national level is for coordination. This coordination may involve more than simply swapping information about projects planned, under way or completed.

In many cases the coordinating person or centre will be responsible for conceptualising the problems so that a useful framework can be developed within which information transfer can operate effectively.

I would like the coordinator to take the coordination function further and prepare handbooks to asssist the cooperating researchers. I assume the person or centre would be adequately funded both to devote time to these activities, and also to make field visits to discuss local activities with the cooperating researchers.

And I suppose this leads to a coordinator of coordinators of science education research activities, so that all interested persons can be kept well informed.

I realise, at the end of my paper, that I have not mentioned priorities except to maintain that student needs should be our central concern.

Any list we prepare will change over time. I will leave it to the Australian Science Education Research Association to start to prepare its list of priorities.

REFERENCES

Balson, M. (1973). Improving the science excursion: an educational technologist's view. *The Australian Science Teachers Journal,* 19, 1, 5-11.

Block, J.H. (ed) (1971). *Mastery learning: theory and practice.* New York: Holt, Rinehart and Winston, 1971.

Bloom, B.S. (1972). Innocence in education. *School Review,* 80, 3, 333-352. (May 1972).

Bloom, B.S., Hastings, T. and Madaus, G.F. (1971). *Formative and summative evaluation of student learning.* New York: McGraw-Hill, 1971.

Bowman, M.J. (1970). Mass elites on the threshold of the 1970's. *Comparative Education,* 1970, 6, 3, 141-160.

Central Advisory Council for Education (1967). *Children and their primary schools (Plowden Report).* London: HMSO. (2 vols.)

Coffman, W.E. (1969). Concepts of achievement and proficiency. *Proceedings of the 1969 Invitational Conference on Testing Problems (Educational Testing Service),* 3-11.

Coleman, J.S. and other (1966). *Equality of educational opportunity.* Washington: US Government Printing Office. (2 vols.)

Comber, L.C. and Keeves, J.P. (1973). *Science education in nineteen countries: an empirical study.* Stockholm: Almqvist & Wiksell.

Cronbach, L.J. and Snow, R.E. (1969). *Individual differences in learning ability as a function of instructional variables.* Standord: Stanford School of Education, 1969.

Educational Testing Service (1973). Dentists assess their competence with self-administered exams. *ETS Developments,* 20, 1, 1. (Winter 1973).

Husén, T. (ed.) (1967). *International study of achievement in mathematics.* Stockholm: Almqvist & Wiksell/New York: Wiley. (2 vols.)

Husén, T. and Boalt, G. (1967). *Educational research and educational change: the case of Sweden.* New York: Wiley.

Keeves, J.P. (1972). *Educational environment and student achievement.* Hawthorn: ACER.

National Assessment of Educational Progress (1969). *Science objectives.* Ann Arbor, Michigan: NAEP.

National Assessment of Educational Progress (1970). *Report 1. 1969-1970 science: national results and illustrations of group comparisons.* Ann Arbor, Michigan: NAEP.

National Science Teachers Association (1970). *Annual self-inventory for science teachers.* Washington: NSTA.

National Science Teachers Association (1970). *Conditions for good science teaching in secondary schools.* Washington: NSTA.

Orlans, H. (1972). Criteria of choice in social science research. *Minerva,* 10, 4, 571-602.

Rosenshine, B. (1971). *Teaching behaviours and student achievement.* London: NFER.

Rosier, M.J. (1973). *A statistical picture of the Australian science teacher: some results from the IEA Science Project.* Address to CONASTA. 15 May 1973.

Rosier, M.J. (1973). Home background is major factor in school results. *School Bell*, 28, 4, 10-11. (May 1973).

Rowlands, E. (1967). How pupils learn. *Lab-talk*, 11, 9, 15-16. (October 1967).

THE INTERDISCIPLINARY APPROACHES TO CHEMISTRY (IAC) PROGRAM AND RELATED RESEARCH

Marjorie Gardner

The stage has been set nicely for my presentation by earlier presentations which point out in their content analysis of CHEM Study that it represented a strong pendulum swing to the research-based theoretical and quantative sides of chemistry and away from the application side that had become so prominent in the two decades from 1940 to 1960. Now, early in the 1970's, chemistry teachers in the United States have recognized that while CHEM Study represents an important advance in chemical education, it was designed for the limited number of students who clearly have science or engineering career goals in mind. The much larger audience of students who need and want to study chemistry for general education purposes are not satisfied by CHEM Study with their chief complaint being that it is not relevant to their interests and needs. Enrolments in high school chemistry are declining. Student and teacher alike seem to be searching for an alternative programme that will appeal to the large student population interested in studying chemistry for scientific literacy purposes.

The I.A.C. Instructional Materials

Hence, the advent of IAC, the Interdisciplinary approaches to Chemistry programme developed by high school teachers and chemistry professors who have worked closely together at the University of Maryland[1]. The IAC programme presently consists of a series of seven interchangeable instructional modules: *Reactions and Reason*, an Introductory Module; *Form and Function*, an Organic Chemistry module; *Diversity and Periodicity*, an Inorganic Chemistry module; *The Delicate Balance*, an environmental Chemistry module; *The Heart of Matter*, a nuclear Chemistry module; *Communities of Molecules*, a physical Chemistry module; and *Molecules in Living Systems*, a biochemistry module. Four additional modules on consumer chemistry, geochemistry, chemical calculations, and the chemical cycles of the earth are under development. An extensive laboratory programme of mini-experiments and experiments have been integrated directly into the student modules.

Detailed teacher's guides for each module have been designed as a means of in-service education for classroom teachers. These include teaching strategies, help with laboratory and classroom management, a guide to self pacing and independent study, statements of major concepts and performance objectives (knowledge, skill and attitude) and evaluation items keyed to the objectives. In addition a pre-test and post-test that attempt to measure change in the cognitive, psychomotor and affective domains plus dual tests (knowledge and skill) for each module have been constructed.

Instructional Characteristics

Field tested, in experimental editions for two years prior to publication,[2] these

1. For further information:
IAC Director,
Chemistry Department
University of Maryland,
College Park, M.D., U.S.A. 20742.

2. Harper and Row, Publishers, Inc.,
2500 Crawford Avenue,
Evanston, Illinois, U.S.A.

modules were designed to convey the following instructional characteristics: an interdisciplinary nature, relevancy, flexibility, concept and process emphasis, student-centred orientation, an informality of style and fun. It was the intent of the IAC team to return many curriculum decisions to the local school system and the classroom teacher. Such decisions include: (1) which modules to teach, (2) in what order, and (3) whether in conventional classroom group or in a self-paced or independent study mode.

Teaching The IAC Programme

The introductory module takes the longest to teach (approximately eight weeks), with others requiring approximately six weeks in the middle of the school year and cutting back to four weeks near the end of the year regardless of selection of modules. In addition to the common use of four to six of the modules to comprise an 11th grade year-long chemistry course, the IAC modules are finding many other niches in the curriculum. In some cases they are arranged to form a semester course or minicourse of six to nine weeks duration. In others, they are used to enrich a CHEMS course or a biology, physics, or earth science programme in the secondary schools. Other schools, especially those interested in moving toward the teaching of integrated science, are proposing to use the introductory and environmental modules in the 9th grade, organic and biochemistry in 10th, inorganic and physical in the 11th and the nuclear module in the 12th grade. These will be supplemented with instructional materials from biology, physics and the earth sciences to form the four-year secondary school programme. Community colleges find the IAC modules to be attractive for their chemistry programme for non-science students. Part of the excitement of IAC Chemistry is its adaptivity and, for the developers, watching it find its place in the science curriculum.

If we are correct in some of our basic hypotheses about science teaching and learning, IAC Chemistry may well be an important new direction in science education. One of these is that a structured scope and sequence programme (the conventional textbook) is no more effective in the learning of science than a range of interchangeable content modules. Another is that students will learn better and retain useful knowledge, skills and attitudes longer if they enjoy what they are doing, don't feel too threatened, and see relevancy to themselves and the everyday world around them. Third, that it is just as important to teach chemistry for cultural reasons, for scientific literacy, as for career preparation. We believe that this broad, general, interdisciplinary approach to chemistry is an adequate preparation for the more specialized studies a student will encounter if they continue into university education and a far better preparation for the understandings of science and technology that they will need to live in and cope with this complex culture of the 1970's.

Field Testing and Feedback Systems

Now, if a curriculum development group sets the goals outlined earlier in this paper, they must be prepared to evaluate their efforts and to encourage others to evaluate it as well. Formative evaluation must be a hallmark of any new curriculum project. Summative evaluation needs to be carefully designed as well.

After defining the purpose (to popularize the teaching of Chemistry) and the instructional characteristics of IAC Chemistry (cited earlier), the development of instructional materials began. Following general agreement on the content, the experiments and mini-experiments were developed and tested before the narrative text was written to ensure a student-centre programme. The feedback system went into operation immediately as the experiments were run and criticised by an in-service class of about thirty experienced secondary school teachers who met weekly in the evening at the University of Maryland.

During the next summer (1971), chemistry professors teamed with high school teachers to test-teach their modules to students enrolled in summer school classes in nearby secondary schools. Now, direct student feedback was combined with teacher response to revise the student modules. During the summer also, the teacher's guides were brought into first draft form with classroom teachers taking the lead in development. During the 1971-72 school year, the revised experimental editions of the modules were test taught in twenty schools. These schools, representing a range of urban, suburban and rural schools, were located near the university so that teachers could meet with the IAC author team on Thursday evenings in feedback sessions. Each test teacher taught two or three IAC classes while the remainder of his/her schedule was CHEM Study or a conventional chemistry programme.

Subjectively assessed, a high level of enthusiasm for IAC Chemistry was apparent in students and teachers. The all-important student grapevine began to operate in the schools and chemistry enrolments increased for the subsequent year in nineteen of the twenty schools, in some cases dramatically.

Admittedly, some of this must be ascribed to the novelty effect. All of the students under the direction of test teachers, IAC and non-IAC, were given the NSTA-ACS High School Examination. Surprisingly, the IAC students, even though they had experienced a broader, more interdisciplinary, less quantitatively and theoretically rigorous study of chemistry, had a good grasp of the important facts and principles of chemistry as measured by this national examination. Statistically there was no difference in performance on the NSTA-ACS test by IAC and non-IAC students. Since the sample was modest (twenty teachers and approximately 2,000 students), the teachers were atypical (all known as above-average classroom teachers with a committment to IAC) and the sample geographically parochial in location, the results of this first content assessment could only be considered as an indicator that IAC students were not being seriously short-changed with respect to their knowledge of chemistry.

Related Research

The increasing enrolments were evidence of an important attitude change and the knowledge performance was comforting, but it was considered necessary to try to construct instruments that could more precisely measure changes in attitude, laboratory skills, knowledge and the adaptability to self-pacing that we claimed. This led to a very important decision in curriculum development that a series of research studies should be designed and based on IAC Chemistry from its earliest stages.

Four research studies were immediately identified. Additional studies are being initiated now. The original four were designed by four advanced doctoral students working under the direction of the IAC project director (Marjorie Gardner) the Director of the Bureau of Educational Research (James Raths) and the Chairman of the Department of

Measurement and Statistics (John Gibblette). Members of the Chemistry Department faculty and the Science Teaching Center faculty also served on these research committees. Two of the studies have been reported and the other two are still in progress. The four include research on student attitudes (Henry Heikkinen), student laboratory skills (Robert Hearle), student knowledge of fact and concept (Harry Gemberling) and factors influencing student success in independent study (Amado Sandoval). The first two studies have been reported and will be summarized here.

Attitude Study

Heikkinen's study assessed the relative influence of IAC and non-IAC programmes on the measured attitudes of high school chemistry students. The study's purposes were (1) to develop an easily administered and scored attitude-toward-studying-chemistry scale, the twenty-item Likert scale, Student Opinion Survey in Chemistry (SOSC); (2) to compare IAC with conventional chemistry courses in their capacity to differentially affect student attitudes toward studying chemistry; and (3) to investigate specific student variables which might account for the range of student attitudes observed after a year's experience in high school chemistry.

Pearson correlation coefficients and analysis of variance and covariance techniques tested hypotheses related to possible treatment (IAC vs. non-IAC) and sex (male-female) influences on student attitudes as measured in September, January, and April. Stepwise regression checked the ability of selected student variables, either singly or jointly, to reduce the amount of unexplained variance among end-year SOSC attitudes.

Initial attitudes of girls were significantly less favourable toward the studying of chemistry than were the attitudes of boys. No treatment or sex differences were found among either mid-year or end-year SOSC attitudes, when initial attitudes were held constant, however. Girls in both types of courses achieved significantly higher TOUS scores than boys, although no TOUS difference was attributable to treatment.

From 47% (IAC males) to 75% (non-IAC males) of end-year SOSC attitude variance could be explained by the variables studied. Probability of taking more chemistry, enjoyment of previous science courses, initial attitude and expected grade acted jointly to account for most of the end-of-course attitude variance.

Although all sub-groups expressed favourable end-year attitudes toward the study of chemistry, modest declines in the degree of favourability were noted for both treatment and sex groups during the school year.

Laboratory Skills Study

Hearle's study was directed toward measuring attainment of laboratory skills, a badly neglected area of evaluation in the teaching of chemistry. The study's purposes were: (1) to identify skills that are expected learnings from high school laboratory experiments regardless of course; (2) to develop means of assessing both the cognitive and manipulative skills learned primarily through laboratory based experience; (3) to translate this into a multiple choice instrument that could be used nationally to compare performance in the skills areas, and (4) to compare IAC with non-IAC students and boys with girls in their achievement of laboratory skills.

First a grid of cognitive (i.e., observing, interpreting, predicting) skills and manipulative (i.e., weighing, titrating) skills common to the IAC, CHEM Study and Modern Chemistry programmes was developed. A pre-test and post-test that incorporated items believed to measure each of these skills was developed and revised to a satisfactory level of validity and reliability. A teacher's guide of instructions for setting up the practical stations was designed. The instruments were given pilot runs and tested for validity (by Jury technique), reliability and discrimination characteristics (statistical treatments) as well as for logistic problems of preparation, administration and scoring.

In September at the opening of school the revised pre-test of laboratory skills was given to about 1,500 IAC and non-IAC students studying under three types of chemistry teachers: (1) all IAC classes; (2) combination of IAC and non-IAC classes; and (3) no IAC classes. The results were analyzed for both treatment (IAC vs. non-IAC) and sex differences (male vs. female). Pearson correlation coefficients and analysis of variance and covariance techniques were used.

The results indicated that there is no significant difference in laboratory skills related to sex. The girls scored as well as the boys. The IAC students' achievement in the laboratory skills test was significantly higher than the non-IAC students. Correlations between the Laboratory Skills Test and tests of conceptual and factual knowledge of chemistry such as the NSTA-ACS High School Chemistry Test or the IAC-developed concept and fact test were low enough to indicate that the Laboratory Skills Test was in fact measuring a new dimension in a student's chemical education. The students' enthusiastic reception of the Laboratory Skills Tests was a subjective indicator that they felt that their work in the laboratory was finally being recognized and rewarded.

In addition to extending the two research studies already reported and completing the IAC study of students' knowledge and the factors that are predictive of student response to independent study, some other studies that need to be initiated can be cited. It's possible that the Likert-type attitude instruments measure only the more superficial opinion-type attitudes. How can the deeper attitude and science values of students be assessed? How do attitudes of teachers correlate with student attitudes? What role does self-concept play in either teacher or student independence in teaching and in learning? How can the essence, mystic or spirit of innovation in curriculum be effectively transferred from developers to teachers? These are only a few of the research areas that cry out for exploration.

A DESCRIPTION OF THE DEVELOPMENT OF ASSESSMENT PROCEDURES FOR THE SCHOOLS COUNCIL INTEGRATED SCIENCE PROJECT

William C. Hall

Introduction

The Schools Council integrated Science Project is being developed in the UK for 13-16 year old pupils and leads to a double certification GCE 0 level. The first group of students to participate in trials commenced during September, 1970. Evaluation, assessment and measurement have played prominent parts in the preparing and ammending of the scheme; this paper is restricted to a description of the development of procedures for assessing objectives achievement by pupils. Models used by the project are described in the Handbook[1]; the society, science and technology model is influenced by Layton's work[2]; and the model used for producing the list of aims has been described by Hall[3]. A brief outline of some of the decision making involved in the project can also be read[4]. The most important model for the present discussion is the learning model which was inspired by (but is by no means identical to) the ideas suggested by Gagné[5]. Figure 1 illustrates the model.

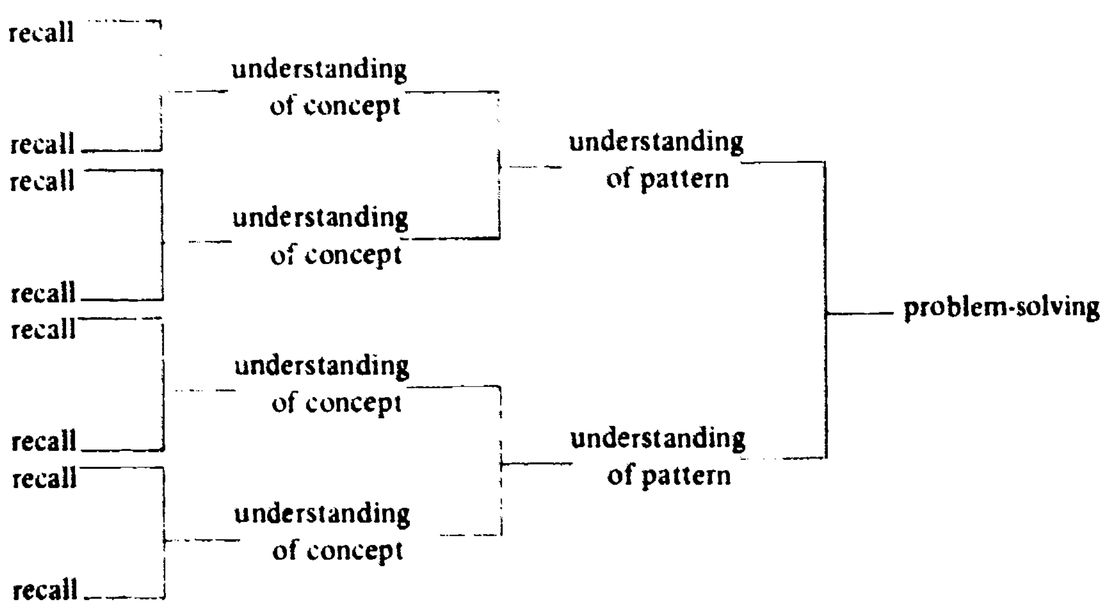


Figure 1: The SCISP Learning Model

The use of this model in teaching allows complete flexibility within a structured learning framework. This is best illustrated by stating that in the examinations, none of the material in any of the pupils' books is tested. These books show just one way of achieving the aims of the scheme, and so long as the 75 patterns with their associated concepts are learnt, then the student should be able to deal with the cognitive part of the assessment.

Figure 2 shows the original list of aims, and the ways these were revised after the 1971 assessment.

	ORIGINAL AIMS (1971)		REVISED AIMS (1972 onwards)
1A	To recall and to understand that information which would enable them to take A level courses in biology, physics, chemistry or physical science, would enable them to follow a job in science or technology, would enable them to read popular scientific reporting and to communicate intelligibly with others on scientific matters, and would enable them to pursue science as a hobby.		Skills Pupils should be able to demonstrate their degree of competence in
		1	(a) recalling and (b) understanding those concepts which would enable them to pursue science (courses in physics, chemistry, biology or physical science) to a higher level or as a hobby.
		2	(a) recalling and (b) understanding those patterns which are of importance to the scientist.
2A	To understand the importance of patterns to the Scientist and to use those patterns in solving problems (both of a laboratory and of an every day type).	3	Making critical appraisal of available information, from whatever source, as an aid to the formulation or extraction of patterns.
		4	using patterns and making critical appraisal of available information in order (a) to solve scientific problems and (b) to make reasoned judgments.
3C	To have the ability to organize and to formulate ideas in order to communicate to others, and as an aid to understanding, critical analysis, etc.	5	organizing and formulating ideas in order to communicate them to others.
3A	To understand the relationship of science to technical, social and economic development, and to be appreciative of the limitations of science.	6	understanding the significance, including the limitations, of science in relation to technical, social and economic development.
1B	To be honest in reporting scientific work.	7	being accurate in the reporting of scientific work.
2C	To be able to use resources (e.g. books, apparatus) at their disposal.	8	designing and performing simple experiments, in the laboratory and elsewhere, to solve specific problems and to show perseverence in these and other learning activities.
1C	To work independently and as part of a group.		Attitudes Pupils should:
		9	be willing to work (a) individually and (b) as part of a group
4B	To be willing to make some decisions on the balance of probability.	10	(a) be sceptical about suggested patterns yet (b) be willing to search for and to test for patterns.
2B	To be concerned for the application of scientific knowledge for the good of the community.	11	be concerned for the application of scientific knowledge within the community.
3B	To have an interest in science and technology and be willing to pursue this interest to higher levels.		

A aims: Knowledge; B aims: Attitudes; C aims: Skills

Figure 2: Original and modified lists of aims

"Problem solving" is seen to be the ultimate cognitive achievement. Merrifield's application[6] of the structure of intellect model[7] to Gagné's eight types of learning was used to hypothesize seven types of SCISP problem. These are listed. (The letters in brackets refer to Guilford factors.)

(a) Interpretation of photographs or scientific drawings.
(NSI)

(b) Detection of "errors" in experiments.
(CMT)

(c) Originality, or creative ability (for example, in devizing an experiment to test an hypothesis, or in suggesting uses for a machine).
(DMT)

(d) Selecting from various pieces of equipment a particular one for a job. (Although this can be problem solving, it is most likely that this activity will be straightforward recall.)
(NMT)

(e) Calculating the consequences of given information (for example, given present day trends, predicting future events).
(CMI)

(f) Giving the best interpretation to a series of facts or observations, when a number of interpretations are possible.
(EMI)

(g) Giving the one and only interpretation to a series of facts or observations.
(NMI)

(Compared with Merrifield this has an additional factor, NSI, but does not include DFT.)

The 1971 Assessment

An attempt was made to answer eight questions as a result of the 1971 assessment:

1. Which aims should be assessed for GCE?
2. How should these aims be assessed?
3. What weightings should be given to the various aims?
4. What form should the GCE assessment take?
5. What will GCE double certification imply?
6. Is problem solving related to high scores in either verbal or non-verbal reasoning?
7. Have pupils undergone any change in their ability to achieve aims during their first year of SCISP?
8. Can different problem types be identified?

A population of about 900 14 year old students drawn from a wide variety of secondary schools in London, Birmingham and N. Ireland were taking part in SCISP trials and these were the pupils who were assessed after their first year of SCISP. The matrix shown in Figure 3 summarises the procedures used and the aims which they were supposed to assess.

ASSESSMENT PROCEDURE	AIM (see figure 2)												
	1A	2A	3A	4A	1B	2B	3B	4B	5B	6B	1C	2C	3C
AH4 intelligence test	✓	✓											
SCISP paper 1	✓												
SCISP paper 2		✓											
SCISP paper 3						✓		✓		✓			✓
Teacher detailed assessment	✓			✓	✓		✓	✓			✓		✓
NFER pupil opinion poll				✓		✓	✓		✓	✓			
Pupil self assessment	✓						✓		✓		✓		
Pupil interviews	✓						✓						✓
Global teacher assessment	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
7 factor test		✓											

Figure 3: Summary of assessment procedures

The AH4 test measures verbal and spatial ability[8]. SCISP paper 1 contained 32 multiple choice items which tested recall and understanding of concepts and patterns, and also pattern finding. The latter questions consisted of presenting students with data and asking them to extract the pattern from the information given. (Sometimes there was no pattern.) Paper 2 was the problem solving paper and consisted of all seven problem types in multiple choice and "essay" form. Paper 3 contained questions like the following.

Question

A new factory producing an important substance and requiring a large number of unskilled people was being planned. Unfortunately, a waste product of the factory which was emitted from a chimney had an unpleasant smell. Which one of the following courses of action would you suggest? (There is no one 'correct' answer.)

A do not build the factory
B build the factory well away from any town
C build the factory close to the workers
D do not build the factory until the smell can be eliminated.

Explain your decision.

Suggest cou 's of action other than A D which might be possible.

Question

Someone who has studied no science says to you 'I know that energy can be conveyed by a flow of water, or of air, or of electricity. What I can't understand is why electricity needs two wires, while water or air needs only one pipe'.

Write several sentences explaining for him the similarities and differences between water, air and electricity in conveying energy.

Question

A scientist says he observed that when liquid helium was poured down the side of a beaker it ran along the bottom of the beaker, up the other side, and out on to the bench!

1 What would you *expect* to happen when liquid helium is poured into a beaker?
2 In whay way(s) does the explanation pattern for liquids not fit in with the observation for helium?
3 Does this mean that the explanation pattern is no longer useful? Explain your answer.

Teachers were given detailed notes for assessing pupils on a five point scale. (The notes had been written after detailed meetings with teachers, and only after general agreement had been reached.)

The pupil opinion poll[9] purports to test five factors: science interest, social implications, learning activities, science teachers, and school.

The detailed notes given to pupils enabled them to assess themselves on a five point scale. For the interviews, a random selection of three separate groups, three pupils were taken and interviewed. They were encouraged to bring exercise books with them. These groups were chosen from each of eight English schools. For four of the schools the process was repeated with a different judge in order to moderate the judges. When When globally assessing, teachers were asked to give an overall percentage mark for the groups of aims A, B and C for each student.

The seven factor test was devised on the basis of examples given by Guilford from tests which are supposed to assess each of the problem types which had been identified. In addition to these assessment procedures, teachers were also asked to weight the thirteen aims according to their estimate of importance; to suggest other ways of assessing aims; to state which aims they believed could not be assessed (and why); and to offer general criticisms of the measuring instruments used.

The usual statistics were computed (means, s.d., frequency distribution, reliability, facility and discrimination indices) and, in addition, a 40 x 40 correlation matrix to help determine the validity of instruments. The seven factor test and problem solving paper were factor analysed (principal components analysis).

As a result of these analyses, and following feedback from teachers, the following results were obtained:

a. The correlations between different levels of the learning hierarchy were not high (on average about 0.3).

b. There was also a low correlation (0.28) between problem solving and the AH4 test.

c. Pupil self assessment generally had insignificant (0.1% level) correlation with other measures for the same aims.

d. Teacher assessment correlated well with other measures.

e. There were moderate (0.43) to fairly high (0.63) correlations between the various factors of the Pupil Opinion Poll.

f. A and C aims (knowledge and skills) correlated highly (0.74).

g. The social implications factor of the Pupil Opinion Poll had low correlation with other measures of this aim.

There was a highly significant difference between the means of Papers 1 and 2 of the pretest and post test populations.

The factor analysis was disappointing. The seven factor test showed expected results, but when combined with the problem solving paper 2, clear factors did not emerge. Questions containing significant physics content were grouped together, and most of these were numerical. There was no distinction between multiple choice and essay questions. Questions based on the same information provided in the stem had similar factor loadings.

As a consequence of these results, and following feedback from teachers, it was decided to drop some of the aims, to refine others and to combine the knowledge and skill aims (see figure 2). Certain instruments were chosen for further investigation in 1972 and others were dropped (e.g. pupils self assessment and the Pupil Opinion Poll). The weightings for the first GCE were also proposed and the 1972 assessment used these specifications.

1972 Assessment

The same population was used for the 1972 studies.

There were seven aspects to the 1972 assessment of pupils' achievement:

1. To refine the chosen measuring instruments following the 1971 assessment.
2. To decide how the GCE SCISP 'A' pass and SCISP 'B' pass should be composed. ('A' and 'B' are the two separate certifications offered by SCISP.)
3. To compare the distribution of 'A' and 'B' passes.
4. To produce agreed norms for the internal assessment of aims.

5. To compare SCISP and non-SCISP pupils of the same age.
6. To perform an item analysis of items which were not pretested.
7. To compare an objective problem solving paper with a short answer problem solving paper containing identical questions.

Five written papers (two multiple choice and three short answer or essay) were administered. In addition, an alternative form of the multiple choice problem solving paper (Paper IV) was produced which contained identical questions to those on the multiple choice paper but instead demanded short written answers.

Specifications agreed for the GCE 0 level examination were followed and these are shown in figure 4. The aims correspond to the modified list given in figure 1.

	AIM ASSESSED															TOTAL
	1	2	3	4 a	4 b	5	6	7	8 a	8 b	9 a	9 b	10 a	10 b	11	
Paper III **Integrated Science A** Recall concept	6															
Recall pattern		4														
Understanding concept	10															60
Understanding patte		16														
Pattern finding			24													
Paper I Cognitive and non-cognitive					4	8	8									20
Paper IV **Integrated Science B** Understanding concept	4															
Understanding pattern		6														31
Problem solving				31												
Paper V Problem-solving				12	4	4			4							24
Paper II Cognitive and non cognitive				1	4	4	12						4			25
Teacher assessment						8		4	4	4	4	4		4	8	40
Total weightings	20	26	24	34	12	24	20	4	8	4	4	4	4	4	8	200

Figure 4: Specification for 1972 examination

Meetings were held with teachers and questionnaires completed by them to develop a set of suggested norms for the internal assessment of aims. (Conventional moderation was impossible: because of the large number of school eventually to be involved visits by moderators was impracticable, and statistical moderation would have been meaningless because teachers were generally assessing different aims from those evaluated by written papers.)

The usual statistics were computed for all papers. To our surprise, nearly perfect normal distributions were obtained for all papers except for the alternative paper IV. Double and triple marking of essay papers showed remarkable consistancy and point

 BEST COPY AVAILABLE

biserial coefficients for the multiple choice items were generally in the 0.3 to 0.4 range. The variances were identical for the two types of problem solving paper and the differences between means highly significant. The correlation between the two papers was 0.45. On analysis, it was revealed that in the short answer version pupils were also achieving aims other than "problem solving" and were also being marked on some of these aims, with the result that the mean for this paper was approximately one half that for the multiple choice version.

There was a correlation of 0.4 or less for the teachers' internal assessment of aims with aims assessed by written examination, thus indicating that teachers were assessing different characteristics. The teachers were also asked to assess student achievement on the written papers (before the papers were attempted!) and the correlations between estimates and actual performances were high (above 0.7). The intercorrelation of marks for the various aims tested in figure 2 showed significant but not high values.

A carefully selected parallel population of non-SCISP pupils was compared with SCISP pupils for each level of the learning model. There were significant or highly significant differences between the means for the achievement of both populations.

A comparison of the various ways of distributing the internal assessment component between SCISP 'A' and SCISP 'B' (to provide the double certification offered by Integrated Science) showed that the internal assessment was best equally divided between the two passes. It was decided to make SCISP 'A' consist of Papers III and I with internal assessment, and SCISP 'B' consist of Papers II, IV and V with internal assessment, thus providing a crude profile of student achievement in GCE O level.

On the basis of the research (albeit necessarily "messy" because of its applied nature) the first GCE examination will be taken by pupils this summer (1973). A report on this examination will be available from the examining board which is administering the examination for the Project on behalf of all GCE boards[10].

REFERENCES

1. Hall, W.C., *Patterns Teachers' Handbook*, Longman/Penguin for the Schools Council, 1973.

2. Layton, D., *Science in General Education, Trends in Education*, no. 25, 1972, HMSO.

3. Hall, W.C., *General Science to O level, School Science Review*, 51, no. 175, John Murray.

4. Hall, W.C., *Case Study in Curriculum Decision Making, Australian Science Teachers' Journal*, 17, no. 3.

5. Gagné, R.M., *The Conditions of Learning*, 1967, Holt, Rinehart and Winston.

6. Merrifield, P.R., in Klausmeyer and Harris, *An Analysis of Concept Learning*, 1966, Academic Press.

7. Guilford, J.P., *The Nature of Human Intelligence*, 1967, McGraw Hill.

8. Hiem, A.W., *AH4 Group Test of General Intelligence*, 1970, NFER.

9. National Foundation for Educational Research, *Pupil Opinion Poll*, 1970, NFER.

10. Associated Examining Board, Wellington Road, Aldershot, Herts., U.K.

PERCEPTIONS OF SCIENCE AND SCIENTISTS BY 13-YEAR-OLD SECONDARY PUPILS

Donald Hutchings

Introduction

Science faculties in universities in Australia, as in other countries, are finding it more difficult to recruit new students. How far this may be attributed to possible shortage of employment opportunities or to other factors is a matter of conjecture. In Britain, it has been postulated that an adverse image of scientists may be one factor which has contributed to the lack of popularity of science as a school subject (Dainton, 1968).

The research reported here forms part of the initial phase of a longitudinal investigation being carried out at Oxford University Department of Educational Studies into the factors influencing secondary school pupils in their choice of school subjects and possible careers

A questionnaire on attitudes to science and scientists was developed in order to obtain a profile of how these are seen by third-year secondary school children. Comparisons were made between boys and girls, science and arts specialists, and orientation of father's occupation as measured on a scale of arts-science bias. The findings indicated that the traditional stereotype of the scientist is changing in several important ways which may be crucial to the choice of science as a subject, and that pupils' understanding of the nature of both the man and the job differs quite considerably along a number of dimensions between the various groups. The main hypothesis is that the concepts of science and scientists form an integral part of the framework within which this subject choice process occurs.

Previous Research

The study attempts to differentiate between the scientist as a man and the nature of his work, and has investigated a more diverse number of groupings than has been the case in earlier work (Hudson, 1966). The aim is to identify individual differences between groups in their perception of the common myth of the scientist as an eccentric out of touch with society at large which do not, of course, emerge when the sample is taken as a whole. The 'halo effect' (Hudson, 1968), for example, can be estimated only by comparing the judgments of interested parties (i.e. those specialists viewing the mythical attributes of their chosen specialism with a certain defensiveness) with those of the total sample of school children.

Selmes (1969), reporting a study of attitudes to science and scientists made by the Association of Science Education (ASE), commented that, although comparatively naive, the concern of a sample of 12-13 year old school children was real and that 'however inadequate their knowledge of scientists and their work, they showed strong feelings about what they believed scientists were doing in the world'. The present research has tended to concentrate on the practical aspects of the job as well as the social implications of scientific activity. The children proved to have very definite ideas on these and on the kind of person who becomes a scientist. How far this image is changing, rather than the extent to which it corresponds to reality, may be the crucial factor as far as the subject choice of the individual pupils is concerned.

The Experimental Procedure

2,000 third-year secondary school pupils, from 17 schools in England and Wales were asked to complete a questionnaire on 'science and scientists'. The results were correlated with other information already collected in the first main field work with this sample. The tests for this phase included:

AH4 group test of general intelligence exemplifying a verbal-numerical and a diagrammatic bias.

HSPQ personality questionnaire derived from Cattell's 16PF.

APU occupational interests guide.

CP1 questionnaire providing background information on socio-economic status of the family, level of parents' education, parents' ambitions for children, arts-science orientation of parents' occupations, pupils' attitudes to school and school subjects and their plans for the future as regards higher education and careers.

The questionnaire was designed under four main headings: (i) science in school (ii) the abilities of scientists (iii) the character of a scientist and (iv) the job of a scientist. The first three were designed as a semantic differential on a four-point scale and the fourth as eighteen true-false statements.

The results were broken down by six main variables: (i) sex (ii) scientific interest as demonstrated by the APU Occupational Interests Guide (iii) interest in science as a school subject and (iv) as a career as expressed in the attitude scale (v) subject group determined by expressed preferences for school subjects and (vi) orientation of father's occupation on a scale of art-science bias. The intention was to identify those groups holding the stereotyped view of scientists, which aspects of it are peculiar to the opinions of science specialists and which are common to all pupils.

Results

(a) *Science in School*

The present data confirms that, even at 13, boys are much more favourable in their attitude towards science at school than are girls. 64 percent of the boys compared with 34 percent of the girls felt that they were doing well in science. The girls particularly complained that science involved 'learning too many facts'; also science teachers were less popular with them than other subject teachers. It is clear from Table 1, which gives responses on a four-point scale to enjoyment of science lessons, that boys seem to get greater pleasure and satisfaction from these lessons. It would appear too that there is a much higher level of enjoyment in Physics and Chemistry than in Biology.

TABLE 1 'I enjoy science lessons'

(a) by subject preference

	% Very true	% Quite true	% Not really	% Not at all
Mathematics	22	39	33	6
Physics, Chemistry	60	36	3	1
Biology	28	51	19	2
Arts, Languages	9	42	40	9
Social Studies	11	41	37	11
N = 1903	445	780	548	130

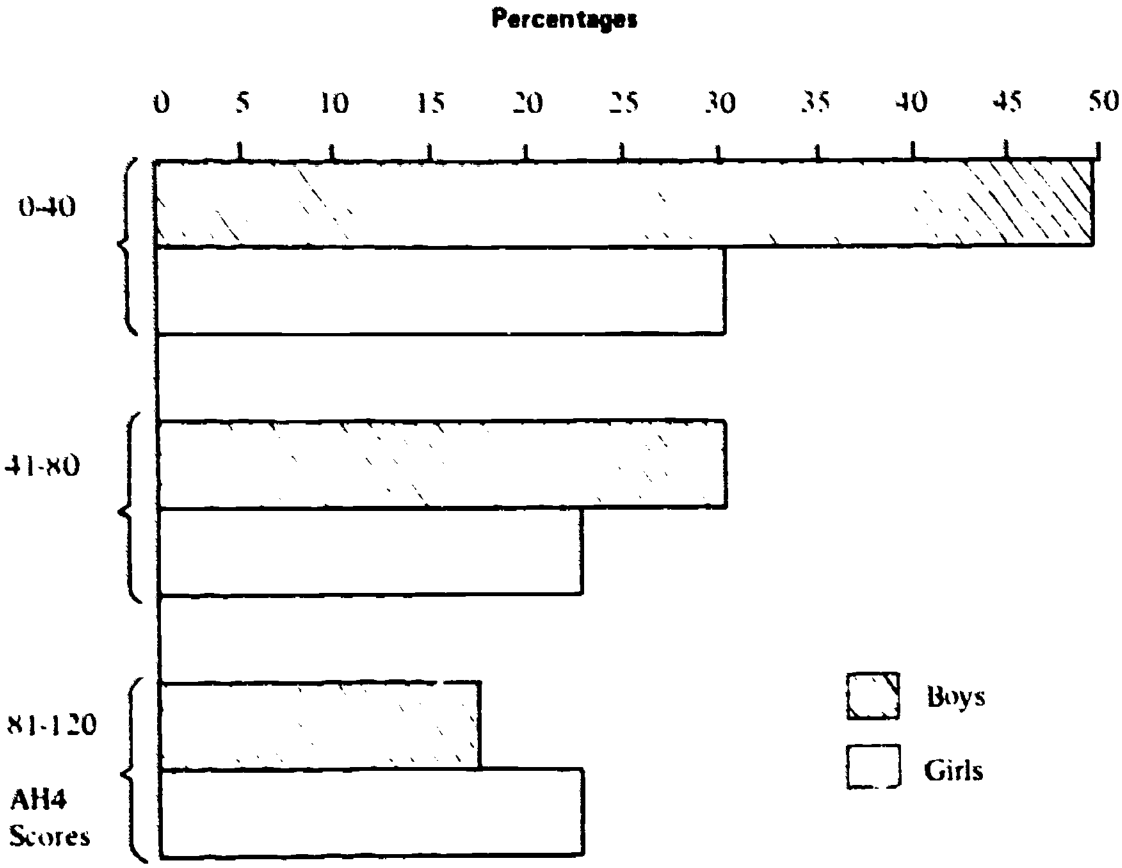
(b) by sex

	% Very true	% Quite true	% Not really	% Not at all
Boys	33	45	18	4
N = 1065	349	483	189	44
Girls	11	35	43	10
N = 838	96	297	359	86

The question arises whether science is more difficult for girls or as Ormerod (1971) has suggested, the emergence of social factors at this stage of pupils' development, is more pronounced in the case of girls. An item in the questionnaire may throw some light on the first of these. It sought to discover how far boys and girls themselves felt that science was in some way more difficult for girls, and it was possible to correlate these responses with the results of the AH4 intelligence test over the whole range of abilities. (See figure 1)

It is clear that the majority of boys and girls felt that both sexes had equal potentialities in science. However, at the lower end of the intelligence scale there was some tendency for the boys to agree with the given statement. The important point is that at this stage the pupils themselves feel that, on the whole, science is no more difficult for girls than for boys.

FIGURE 1 'Science is more difficult for girls than boys'



(b) *The Abilities of a Scientist*

The findings of the present study indicate that there is an overall consensus on the abilities of a scientist. A picture emerges of the scientist as someone who is very clever, not easily put off when things go wrong and very good at explaining things but who is not physically fit and not particularly good at languages. However, there are a number of striking differences between the opinions of science-biased groups. For instance, the 'halo effect' was in evidence in the responses to 'uses his imagination' and 'full of new ideas'. Though both items received a very favourable response (83 percent and 94 percent respectively), the science specialists were between 5 and 6 percent more likely to endow the scientist with these attributes. A rather more surprising result was that the opposite proved to be true of the ability to speek and to be good at public speaking. In both of these cases, the arts-biased pupils tended to give responses more favourable than those of the science-biased. It was interesting too to note that the pupils did not rate particularly highly the scientists's aptitude for making things. This is perhaps surprising when the fictional Dr. Who-type characters of T.V. series are always portrayed as very competent at constructing and repairing extremely complicated pieces of machinery.

TABLE 2 The Abilities of a Scientist

Abilities	Want to be a scientist		Scientific father		Scientific interest		Total
	Yes %	No %	Yes %	No %	High %	Low %	%
Very clever	92	92	91	92	91	91	91
Able to spell	57	69	54	65	60	97	63
Good at languages	23	27	19	24	24	24	25
Able to add numbers	76	66	75	73	75	69	71
Neat and tidy	78	68	74	72	72	73	73
Good at public speaking	35	53	29	43	37	52	44
Able to give orders	43	44	42	47	61	50	44
Full of new ideas	92	96	92	88	92	84	94
Physically fit	31	37	27	33	28	40	34
Uses his imagination	86	80	88	83	87	81	83
Not easily put off	95	91	96	93	96	90	93
Good at making things	53	47	50	49	47	54	50
Puts his work first	72	84	79	75	83	78	78
Good at explaining	91	91	85	90	87	91	91

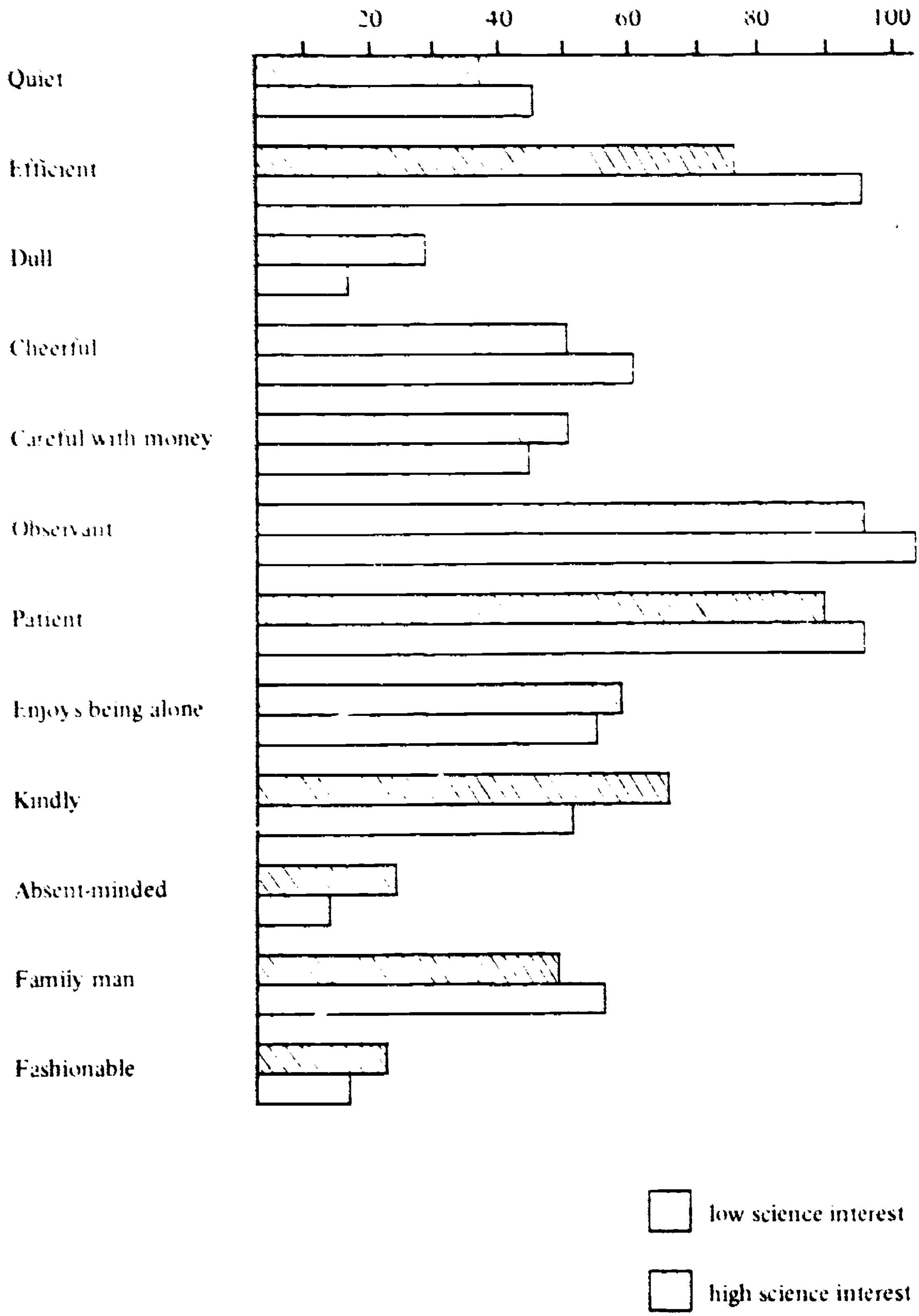
Table 2 shows the different relations existing between three pairs of arts-science biased groups on this section.

The fact that the scientist is seen by almost all as being good at explaining things may perhaps be regarded as an encouraging comment on science teachers. As Selmes (1969) remarked of the ASE sample of school children, they preferred a teacher who make time to explain experiments or practical work, or theory rather than one who seemed 'rushed along' by the syllabus or the book.

(c) *The Character of a Scientist*

This dimension in particular is the factor most generally thought to provide the basis of the scientist myth. Selmes (1969) was left with the impression of 'the mad scientist of horror films and comic papers', and inference drawn from the fact that 21 per-cent of the pupils' comments suggested the image of a scientist leading a narrow devoted life. Similarly, Ashton and Meredith (1969), commenting on an investigation undertaken by the West Midlands Regional Committee of the ASE, state that the popular image of the scientist was of 'a drab, bespectacled, overalled figure, bending over a bunsen burner in a back room' or of a 'dishevelled and wildly excited man dancing around, waving a test-tube and talking to himself'. He was also regarded, unlike the artist or social scientist, to be quite unable to converse socially.

FIGURE 2 The Character of a Scientist by Interest in Science



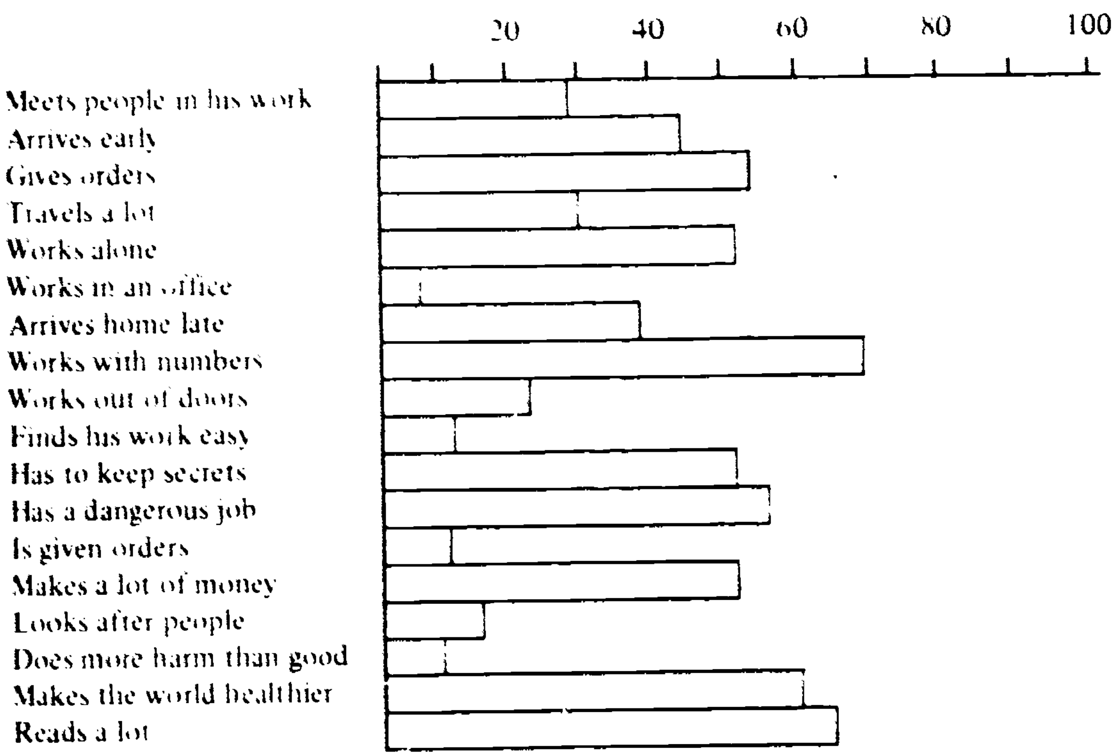
Our findings have tended to present a much more encouraging picture of the scientist as a quietly efficient person who is more in the centre of things than locked away in a back room. He is not the eccentric, solitary figure devoting himself to his work but much more society orientated, working in a team and dealing with the world of politics and big business that we have come to recognise through the media.

The 'absent-minded professor' is certainly becoming outdated. Only 18 percent of the overall sample described the scientist as very or quite absent-minded and only 24 percent of those least interested in science as a subject. Similarly, around 50 percent of the sample described him as enjoying company, being a family man and being kindly and cheerful. He was also described as an interesting person and was not seen as especially dowdy and unfashionable. Figure 2 shows the pattern of responses on these questions divided by science interest, and illustrates the high degree of consensus here.

(d) *The Scientist's Job*

How far subject choice is related to career choice is an open question. However, as far as science is concerned, there is a general assumption that boys and girls who choose science subjects will probably opt for a career in science. If this is so, the image of the scientist's job is likely to be an important factor in the choices being made. Figure 3 illustrates the pupils' responses overall to eighteen true-false statements concerning the scientist's work.

FIGURE 3 The Scientist's Job



These findings certainly present a more favourable picture of science careers than that found in the ASE survey, in which 19 percent of the comments reflected a negative or critical attitude to the work. The responses summarised in Figure 3 give the 13-year-olds' views of the job as being difficult, secret, even dangerous, involving reading and computation, but responsible, well-paid and, on balance, doing more good than harm.

Conclusion

This paper provides data which should throw light on two important questions related to science teaching in the schools. The first is whether 13-year-old pupils tend to have a negative view of scientists and the work they do, and second to what extent boys and girls differ in these attitudes.

We are entering here on an area of opinion rather than fact and conclusions based on responses given by the pupils must be treated with some caution. Even so, the answers suggest that 13-year-old secondary school pupils have a generally favourable impression of the scientist and his work and show that this positive view is shared equally by those boys and girls who enjoy science lessons and those who find them difficult or dull.

Further analysis of these profiles by the HSPQ personality variables will be carried out. It is also intended to extend the work in 1973 in order to investigate the effect of two years' ageing on these images. As Hudson has noted, 'the consensus about certain figures fades over the four years between thirteen and seventeen, while with others it grows sharper'. Many past studies have not indicated who in fact endorses the stereotyped views. Do scientists themselves believe in the image, and if the young science specialists' view of their subject is one peculiar to themselves, it is necessary to know how long they hold this opinion. Do they reject the popular image, and if so, is this before or after adopting the specialism? More specific information is needed over a larger age range, which we are hoping this longitudinal study will afford.

The present paper has examined pupils who have some experience in science but who have not yet chosen or rejected it as a subject for specialised study. By this age, interest in science has crystallised, according to Butcher (1969). Also the third year is the year of decision for or against science options for the majority of pupils in Britain and it is normally only up to the end of the third year that a common science course is followed by all. The data obtained in the next phase of field work, when the pupils in the sample will be 15 years old, should provide information relating to the various choices made by the pupils. How far a change in attitude towards science and scientists by young people today may be the result of a more favourable presentation of scientists by TV and other media, or to what extent this may be ascribed to the introduction of different teaching methods are interesting questions.

These, and no doubt others, would seem to be issues to which the information obtained in this study may be relevant.

Acknowledgements

The study was supported by a grant from the Social Science Research Council. Assistance was also received from the Atlas Computer Laboratory, Chilton, and we are particularly indebted to Mr. J.I. Hailstone and Miss I. Vollmer for their expert advice and help. The work has been facilitated also by the support of the C.E.I. and the C.S.T.I. Thanks are due also to the headmasters, teachers and pupils who generously gave their time to take part in this study.

REFERENCES

Ashton, B.G. and Meredith, H.M. (1969). The attitudes of sixth formers to science and scientists. *School Science Review*, vol. 51, no. 174.

Butcher, H.J. (1969). An investigation of the swing from science. *Research in Education*, vol. 1.

Dainton, F.S. (1968). Enquiry into the flow of candidates in science and technology into higher education. H.M.S.O.

Hudson, L. (1966). *Contrary Imaginations*.

Hudson, L. (1968). *Frames of Mind*.

Hutchings, D.W. (1963). *Technology and the Sixth Form Boy*.

Hutchings and Clowsley, J.M. (1971). Why do girls settle for less? *Further Education*.

Omerod, M.B. (1971). The social implications factor in attitudes to science. *B.J.E.P.*, vol. 41, no. 3.

Selmes, C. (1969). The attitudes of 12/13 year-old pupils to science and scientists. *School Science Review*, vol. 51, no. 174.

PART II

TEACHING STRATEGIES, ATTITUDE AND INQUIRY MEASURES AND SURVEYS

ASEP IN THE CLASSROOM – some issues for research

Scott McKenzie

I have been asked to outline some of the problems associated with the use of ASEP materials, that might give rise to educational research programmes. These problems have been gathered from reports of meetings of trials teachers, discussions with trials teachers, my observations of ASEP classes in action, and from my experience in using ASEP materials with Grade 8 classes.

Many science teachers today are products of an era in education in which the teacher's role was to impart information, and the student's role to receive this information. We left school with this view of the teaching-learning process, and (despite our pre-service training), entered the profession with much the same view. Twelve years of schooling in which we were dominantly teacher-directed, gave us a certainty that such a teaching strategy would be successful for us also. There was no hard evidence to support this view merely our empirical observation over a twelve year period.

We are now confronted with an ASEP view of science education, and the resultant units reflecting this view, which tell us that:

- a major source of learning is the activity of the child,
- children learn by social interaction,
- children should have considerable control over their own learning,
- children should work at their own rate.

Many, if not most of us, are quite prepared to accept these principles per se. It is in the classroom practice that our previous certainties are shattered. We need hard evidence to be convinced of the educational worth of ASEP materials. Most of us *feel* that ASEP science is excellent but our uncertainties run deep.

The problems have been categorised in terms of

(i) the context in which learning takes place,

(ii) the ASEP materials themselves,

(iii) the learner,

(iv) the teacher.

Context

The ASEP view of science education, and its implicit view of the whole junior secondary education, as being directed towards the personal and social development of the child, poses some interesting questions. The materials prepared by the Project seem to have been written for school situations in which rank ordering and grading of students does not occur. The unit structure with a short core followed by many optional core development activities, and the absence of achievement tests give support to this idea.

- What influence do environmental pressures have on ASEP learning, in terms of

  (i) demand for conformity in a school, as against the encouragement of individual expression;

  (ii) extrinsic rewards for students, as against intrinsic rewards;

  (iii) rank ordering and grading of students.

- What influence does the 'climate' of the school have on ASEP learning? Are ASEP materials equally effective in authoritative and supportive 'climates'.

- Is there an optimum context in which ASEP learning will flourish?

Materials

ASEP materials have been designed to tap the interest of students, to cater for individual differences in ability, to be readable by the vast majority of students, to be activity-oriented, and to develop a wide range of abilities, skills, and attitudes. There seem to be quite a number of underlying assumptions that might be subjected to test.

- What are the effects on learning of the ASEP strategy of

  (i) tapping the interest of the student,

  (ii) giving students some choice in what they study,

  (iii) self-pacing,

  (iv) child-centred instruction, rather than teacher-controlled instruction?

Conducting an ASEP class is hard work for science teachers. What we want to know is this — is our effort to be rewarded?

- What learning gains are achieved when ASEP materials are used? What cognitive abilities are developed, attitudes implanted or changed, manipulative skills enhanced, and interests evoked?

- What learning gains result from the use of open-ended, or unstructured activities?

One of the Project's major aims concerns the development of an understanding of the nature, scope, and limitations of science.

- To what extent do the materials attempt to achieve these aims?

- How successful are the materials in achieving these aims?

A similar set of remarks might be made about "attitudes characteristic of scientific investigations".

Some teachers report that ASEP materials are suitable for middle and low ability students only.

- Is this in fact so?

  Or is such a report an indictment of the way the materials were used with high ability classes, rather than of the materials themselves?

ASEP materials consist of forty units with few essential links between units.

- What is the effect of sequencing units so that definite links are made use of?
- What effect on learning will a series of non-linked units have?

The readability of materials has been a major concern of the Project. An American readability model has been used.

- Are the materials readable to the extent that the American model indicates?

Most ASEP units have prepared and highly structured Record Books into which students write their observations and answers to questions posed in other parts of the unit materials. These Record Books are clearly very beneficial in many circumstances.

- What is the effect, if any, of such a confined and definite space on the quality of student responses?
- Does the use of such Record Books inhibit the tendency of students to pursue unstructured activities?

The Project has developed an inquiry model containing a spectrum of the types of inquiry possible. The types of inquiry depend on the amount of guidance given, ranging from a complet ly programmed sequence to a situation in which the student is given no help at all.

- To what extent have the seven inquriy types been used in the forty units?
- How effective is each of the inquiry types in promoting the development of ASEP science objectives?

Already quite a number of teachers have asked whether students are likely to become bored with a three year ASEP course of study.

- Is this so?

- Is there an optimum number of ASEP units taken in sequence, for the various ability groups?

Our main concern with the materials centres on their overall effectiveness in achieving their stated objectives. Thus it is imperative that an extensive summative evaluation of the final ASEP product be conducted.

The Teacher

An ASEP teacher has quite a different role in the classroom. He is responsible for:

(i) the organisation of the facilities for learning;
(books, equipment, room)

(ii) preparing students for learning;
(pre-requisite background, motivation)

(iii) effecting learning: guiding, advising and encouraging students as they undertake their investigations;

(iv) evaluating the effectiveness of learning.

Such a role takes the ASEP teacher out of the lime-light, out of the centre of the stage. Teachers sometimes find it difficult to relinquish the initiative, passing on to students more of the responsibility for learning.

- Is such a role change a real problem for science teachers? To what kind of science teacher? How can this difficulty be minimised or removed?
- Does such a role change reduce job satisfaction?
- How does an ASEP teacher spend his time in class?

Since the role of the ASEP teacher is different, it may be that a successful ASEP teacher differs from a successful traditional teacher.

- How can we identify a successful ASEP teacher?
- What attitudes and other personality components characterise a successful ASEP teacher?

Using eight ASEP units for a one year science course with a class presents a teacher with an incredible burden of reading preparation. Each unit averages about 100 pages. A teacher using ASEP materials with Grade 8, 9, 10 classes faces the prospect of reading 2,400 pages of student text just to become familiar with them.

- How is it possible for teachers to become so familiar with 2,400 pages of ASEP material, that they can promote effective learning in the classroom?

Many teachers who have used ASEP materials express uncertainty as to the quality and quantity of the learning that occurs in the course of a unit.

- How can such teachers be convinced of the effectiveness of learning when ASEP materials are used?

The Learners

ASEP materials have been designed to take account of the wide variation in abilities and interests and needs of students. I have referred elsewhere to a number of studies which might be conducted, centreing on how effective the materials are in catering for individual differences between students.

Many students when asked at the completion of a unit what they now know, that previously they did not, respond with some difficulty. Some even maintain that they know nothing more.

- How can we promote in students, the realisation of what changes they have undergone, in the course of a unit?
- Do students realise the intent of the activities they undertake?
- What effect on learning and motivation does the regular marking/correcting of Record Books have?

Most ASEP materials have been written at a reading level two Grades below its intended users.

- What effect does this low reading level have on the learning and motivation of high ability students?

Most students seem to cope quite well in ASEP classes. Only the very dull and the very dispirited (through years of failure), students seem to achieve nothing.

- Is this in fact so?

The Piagetian stages of intellectual development were undoubtedly useful to the Project writers in the development of the materials.

- What use should teachers make of this concept?
- How are teachers going to be able to determine the stage of development of their students?

The unit MALES AND FEMALES has been designed for 12 year old students. The core of the unit is so frank that it will invoke controversy wherever the unit is used.

- Is the unit really suitable for Grade 7 or 8 students?

ASEP materials cater for high ability students in two ways.

(i) by providing a large number of Options,

(ii) by the inclusion of unstructured activities at the end of Options

- What is the pattern of preference among high ability students? Do they prefer to do more Options, or do they research more deeply into an Option?

ASEP materials normally contain at least one diagnostic test at the end of the Core. No achievement tests have been written — indeed one gains the impression that the Project staff consider the summative evaluation of students for the assignment of grades to be contrary to their view of science education.

In a typical unit students in a class may complete from one to six Options in the time allocated.

- What techniques might be employed in summative testing to allow for the variation in numbers of Options completed by students?

- What effect does testing in order to assign grades have on the students' approach to learning, and on the learning that occurs?

Conclusion

It seems to me that science education in this nation will reflect the influence of ASEP for some considerable time. It may be that the Project's influence will be felt in many other subjects in the secondary school as well.

Whatever the case there is an urgent need for research into the learning that occurs in ASEP classrooms. We need hard evidence to support the widely held view of science teachers that ASEP materials and the ASEP view of science education represent a large step forward in secondary science education.

TEACHER VALUES AND THEIR ASSOCIATION WITH TEACHING STRATEGIES IN A.S.E.P. CLASSES

C.N. Power and R.P. Tisher

The last fifteen years have been characterised by a great deal of interest in the teaching-learning process and many researchers have directed their attention to the classroom to study what goes on, and what is happening to pupils. The increased research activity has been reviewed by several writers (e.g. Flanders and Simon, 1969; Nuthall, 1968, and Rosenshine and Furst, 1971) and recently a special issue of the *International Review of Education* (Vol. 8, No. 4, 1972) was devoted to works dealing with the classroom behaviour of teachers. But even though there has been a great deal of study of teaching some writers believe that there is little "pay-off" for practice. Heath and Nielson (1973), for example, paint a very gloomy picture of the value of, and benefits to be gained from the researchers on teaching. While it is the case that there are some difficulties and deficiencies in some studies, it is not the case, we believe, that the research has been of no value. For example, research on classroom procedures has produced a greater knowledge of what is actually happening in lessons. Also, research workers have developed and fostered a vocabulary which can be used to describe and categorize classroom activities (and hopefully give teachers greater control over what they do). Furthermore, the research on teaching has stimulated innovations such as micro-teaching and mini-courses.

Be that as it may, there are several criticisms of the previous research which cannot be ignored. For example, there is the failure to relate the criterion measures to the content covered in lessons and to the purposes of each lesson (Cohen, 1972; Rosenshine and Furst, 1971). In many studies standardised tests have been used as the criterion measures. Whereas these instruments do measure some of the outcomes of instruction it is not usually the case that they are the most valid instruments for studies, especially short term ones, which attempt to examine the association between classroom activities related to specific content and pupil growth. A more appropriate procedure in research is, according to some critics, to limit the nature of some studies of teaching to the examination of the effects of teaching strategies for specific curricular or specific curriculum units. The criterion tests would be limited to the material covered in the curriculum units or in the actual lessons. Some recent studies (Hughes, 1973) have taken great care to develop valid criterion tests and in two major projects* in Brisbane pre and post-tests are based on the material covered in class.

These last mentioned projects are designed to study the effect of various teaching strategies in lessons where the pupils are following curriculum units prepared by the Australian Science Education Project. In one study a highly structured unit, "Light Forms Images", was used in the experimental classes, and pre and post-tests were designed to measure pupil's level of mastery of the material in the core section of the unit. This study was also designed to take account of the influence of teachers' beliefs about education.

That teachers' beliefs about teaching and/or particular curricular innovations can

* These projects have been funded by the Australian Advisory Committee for Research and Development in Education.

affect that nature of instruction with the curriculum is generally accepted. Unfortunately too few studies take account of the modifying effect of the teacher's values. Certainly, there is a need, as a number of writers (e.g. Grobman, 1968) point out to take account of these beliefs in evaluation of curricula and in studies on the effects of strategies of teaching. This point of view is associated with another, namely the expressed need to take into account the effects of person-environment interactions on pupils' achievement in class (Mitchell, 1969, Power, 1971). Mitchell (1969) indicates, quite cogently, that the determinants of behaviour need to be sought more often in the characteristics of the environmental context and in the interaction of these characteristics with individual traits and abilities. He stresses, too, that social forces and environmental contexts may in certain instances be prepotent over individual traits. In some cases these contexts may have such an immense impact that individual behaviour can only be fully understood when their effect is considered. Recently, Cohen (1972) has argued along similar lines.

The Brisbane Studies

As indicated above, the Brisbane studies attempt to take account, among other things, of the influence of teachers' beliefs and of the actual content covered in lessons. The major aims of the projects were to examine the effects of teaching strategies in various teaching situations including classrooms taught by experienced teachers. This report deals with only one small aspect of the studies namely the association between teachers' values and teaching strategies. Accounts of other aspects of the researches appear elsewhere (e.g. Tisher and Power, 1973) and a detailed report of a completed project is available from the authors. However, it is appropriate to describe some of the procedures followed before discussing in somewhat greater detail the findings which relate to values and teaching strategies.

Initially a Teacher Opinions Scale (based on the A.S.E.P. Opinions Scale) and a Teaching Practices Questionnaire were administered to a representative selection of science teachers (N = 68), diploma in education students (N = 73) and A.S.E.P. staff (N = 16). The data obtained were analysed by using a principal components factor analysis. Varimax scores were obtained for all persons and ten factors extracted in the analysis. The varimax scores were used in an Hierarchical Grouping Analysis (H Group) (Veldman, 1967) to establish clusters of individuals and a Multiple Discriminant Analysis was employed to determine the manner in which the clusters established by H Group actually differed from each other. The characteristics of the clusters are described more fully elsewhere (Tisher and Power, 1973). Twenty-one persons, i.e. science teachers (N = 6) and diplomates (N = 15), representative of the clusters were then selected as the teachers to be used in the study. These twenty-one teachers based five lessons on the material in A.S.E.P. unit Light Forms Images, and three of the lessons were video recorded by the research team. The experienced teachers remained with their usual grade 9 class and used the A.S.E.P. unit during June 1972. The Dip.Eds. each remained with ten to twelve randomly selected grade 9 pupils who were in their assigned practice teaching classes. They used the A.S.E.P. materials during their second practice period in September 1972. A pre-test (readiness test) was administered before teaching on the unit began and the post-tests were administered during the sixth lesson. One of the post-tests was the Classroom Activities Questionnaire (Steele, House and Kerins, 1971) which is designed to measure pupils' perceptions of the learning environment. One was an Attitude to Science Scale and another, the criterion achievement test.

The video-tapes were analysed using a classroom behaviour coding scheme designed especially for this study. Twenty-eight of the categories are listed in Table 2. The scheme provides data on the number of times certain persons and materials acted as the source of interactions, the frequency with which certain interaction occurred (e.g. pupil to pupil, pupil to teacher), and the frequency of types of behaviour (e.g. fact-stating, describing, etc.). Some categories are not listed in Table 2 as the interactions they measured were not significantly associated with any teacher values.

The data from the Classroom Activities Questionnaire was subjected to factor analysis and after varimax rotation eight factors were extracted. These components accounted for 55% of the variance. The names which were given to the components appear in Table 1. Detailed definitions of the various categories are given in the detailed report of the study. However, the descriptions which follow will indicate the meanings attached to a number of these categories.

Correlations were calculated between the teachers varimax scores on the factors obtained from the analysis of the Practices Questionnaire and Opinions Scale and

(a) classroom process measures and

(b) criterion attitude and achievement measures.

The classroom process measures included the mean scores (from the three videotaped lessons) for the interaction measures shown in Table 2 and the class mean scores on the eight factors extracted in the component analysis of the Classroom Activities Questionnaire. The criterion measures were the class means and standard deviations on the achievement post-test and the attitude post-test. The results of the calculations are shown in Tables 1 and 2.

Discussion of Results

The data in Table 1 indicate that there are some associations between teachers' values and pupils' perceptions of the learning environment. For example, teachers who believe that science courses should be "wide ranging" including a great deal and variety of scientific information and information from the social sciences (i.e. teachers with a "breadth orientation") tend to have classes where the pupils perceive the learning environment as one which devalues rote memorization, and expects pupils to go beyond the information given. In these classes, as might be expected, pupils do not perform as well on a specific, curriculum unit-related achievement test as do pupils whose teacher emphasize and restrict their discussions to specific items of subject matter. Also, those teachers who value divergence (unstructured) within lessons, and who encourage original ideas and use a variety of sources for information tend to have classes where the pupils perceive the learning environment as one which values ideas more than grades, fosters invention and discovery (independent inquiry), devalues memorization (cramming) and where the teacher does not dominate discussions.

A more complete description of science classes can be obtained by combining the data in Tables 1 and 2. It is appropriate to note that only the statistically significant relations are reported in these tables and that the information applies to classes when they use the A.S.E.P. materials. However, the experimental teachers were representative of types of teachers with differing educational values and the A.S.E.P. materials are a type of self-paced programme. Consequently, the descriptions may be taken as indicators of the range and variety which could exist in Brisbane's metropolitan school, were they to adopt A.S.E.P. materials or another similar self-paced science course.

Some of our interpretations of the data are as follows.

1. Teachers who value an experimental, problem-centred approach to science teaching and who believe in groupwork tend to have classes in which

   (a) reading is a prevalent activity, pupils frequently consult and/or interact with curriculum materials and description and discussions of difficulties are not frequent activities; and

   (b) pupils perceive the learning environment as one which permits a great deal of participation and involvement and which presents enjoyable ideas for study.

2. Teachers who place little value on an understanding of the nature of inter-personal relationships in classes and who believe that school courses should consist of the separate rather than integrated sciences tend to have classes in which pupils perceive the learning environment as one which requires them to search for implications, trends and consequences in the material presented.

3. Teachers who place an emphasis on rules to guide pupils conduct, and who believe they should discipline and control and make effective use of punishment tend to have classes in which

   (a) relatively few questions, and particularly follow up questions, are asked; discussion seems to centre around the giving of directions with less frequent dissemination of information; the teacher most frequently initiates interactions with individual pupils although pupil groups rather than the teacher are the major source of interactions; and

   (b) pupils perceive the learning environment as one which requires them frequently to explain, summarise and restate ideas in their own words.

4. Teachers who believe that there is a fundamental core of knowledge which all pupils should possess and that science courses should convey the structure of sciences to pupils tend to have classes in which

   (a) criticism, description, brief question-answer interactions and teacher use of and interaction with lesson materials are frequent activities while reading, the giving of directions and observation by the teacher are not; and

   (b) pupils vary greatly in their attitude to science and perceive the learning environment as one which places little emphasis on (and consequently infrequently requires them to engage in) analysis, logical reasoning, recall, recognition, application and extrapolation.

TABLE 1

Pupils perceptions and outcomes

vs Teachers values

Teachers Values / Pupils Perceptions and Outcomes	Problem Centred Approach	Pure Science Orientation	Emphasis on Teacher Rules	Emphasis on Conceptual Schemes	Emphasis on Lecture-demonstration	Emphasis on Intrinsic Motivation	Subject Matter Emphasis	Teacher Centred Approach	Unstructured Divergence (with Curricula)	An Orientation to Breadth (within Curricula)
A. Perceptions of Instructional Climate										
I. Independent Inquiry							(36)		43	
II. Enthusiastic Discussion (R)	-41									
III. Humour					(-35)			-51		
IV. Cramming (R)									39	45
V. Interpretation (R)		-41		(35)	(-34)					-43
VI. Participation										
VII. Translation (R)			-37		-41			(-33)	(31)	
VIII. Information Processing				-53						
B. Outcome Measures										
Achievement (Mean)										(-34)
Achievement (S.D.)										
Attitude (Mean)							(30)			
Attitude (S.D.)	(34)			43	(33)		46			(-35)

r $\geq$ 0.37, significant at the 0.05 level

r $\geq$ 0.50, significant at the 0.01 level

Note: R in parenthesis indicates the reverse of the factor, thus factor IV could be more accurately, if somewhat clumsily, designated as "De-cramming".

5. Teachers who believe that pupils gain little from carrying out their own investigations and that lectures and demonstrations should take precedence over pupils' laboratory work tend to have classes in which

   (a) there are frequently periods of inactivity or task irrelevant activities, and more frequent discussion of purposes, and

   (b) pupils perceive the learning environment as one which requires explanations, summaries and statements of ideas in their own words.

6. Teachers who value intrinsic motivation and do not believe in the use of extrinsic rewards for pupil accomplishments or excessive question answer interchanges tend to have classes in which pupils are often the initiators of interactions but pupil pupil interactions are infrequent, as are discussions of difficulties.

7. Teachers who value the subject matter of the separate sciences tend to have classes in which

   (a) the teacher is frequently the "target" of interactions, factual content is often discussed but reading is a rare activity, and

   (b) there is considerable variation among the pupils in their attitude to science.

8. Teachers who believe that they should be the major disseminators of scientific information and assessors of the accuracy of pupils' understandings tend to have classes in which

   (a) groups rather than the teacher are the most frequent initiators of interactions, periods of inactivity and task-irrelevant activities frequently occur, observation is a prevalent behaviour whereas explanation is not, and

   (b) pupils perceive the learning environment as one which makes frequent judgments about them and rarely contains humour, e.g. laughing and joking.

9. Teachers who believe science courses should be such that pupils are encouraged to use their own ideas, and who do not believe curriculum materials are the only sources of information, tend to have classes in which

   (a) praise and encouragement are frequently given but writing is a rare activity, and

   (b) pupils perceive the learning environment as one which places little emphasis on memorization or on grades, but encourages independent exploration and discovery and values pupils' ideas.

TABLE 2 (PART A)

Classroom Interaction Measure vs Teachers Values

(for 21 classes)

Interaction Measure / Teacher Values	Problem Centred Approach	Pure Science Orientation	Emphasis on Teacher Rules	Emphasis on Conceptual Schemes	Emphasis on Lecture-demonstration	Emphasis on Intrinsic Motivation	Emphasis on Subject Matter	Teacher Centred Approach	Unstructured Divergence	An Orientation to Breadth (within Curricula)
1. Pupil (P) as source						38				
2. Group (G) as source			38					40		
3. Teacher (T) as source			-54					-54		
4. Teacher (T) as target							42			
5. Materials (M) as target	40									
6. P → M interaction	37									
7. P → T interaction										
8. P → P interaction						-46				
9. T → M interaction				39						
10. T → P interaction			41							
11. No (0) activity					45			53		41
12. Questioning (Q)			-43							
13. Follow up Q			-59							
14. Information Giving			-37							

<u>Note</u>: $r \geq 0.37$, significant at the 0.05 level.

$r \geq 0.50$, significant at the 0.01 level.

TABLE 2 (PART B)

Classroom Interaction Measure vs Teachers Values

(for 21 classes)

Interaction Measure / Teacher Values	Problem Centred Approach	Pure Science Orientation	Emphasis on Teacher Rules	Emphasis on Conceptual Schemes	Emphasis on Lecture demonstration	Emphasis on Intrinsic Motivation	Emphasis Subject Matter	Teacher Centred Approach	Unstructured Divergence	An Orientation to Breadth (within Curricula)
15. Direction giving			49	-37						
16. Writing									-60	
17. Observing				-38				41		
18. Reading	43			-38			-43			
19. Q answers				39						-41
20. Fact stating							47			
21. Describing	-42		-54	38						42
22. Interpreting										
23. Explaining								-41		
24. Purposes					39					
25. Difficulties	-47					-41				
26. Nothing/Task Irrelevant					41			45		67
27. + ve Affect									37	
28. - ve Affect				43						

10. Teachers who believe that science courses should be "wide ranging" combing many scientific ideas as well as relevant findings from the behavioural sciences tend to have classes in which

    (a) there is frequent inactivity, task-irrelvant activities and description often occur whereas question-answer interactions do not, and,

    (b) pupils perceive the learning environment as one which places little emphasis on memorization or on grades, contains little talk by the teacher, but which encourages them to draw implications and search for consequences and trends.

Concluding Comment

The research described here is part of a larger study on the effects of strategies of teaching in A.S.E.P. classes. In one sense the preceding discussion is incomplete for it must be related to other findings in the large project. Nevertheless, the details discussed here provide much information on the teaching-learning process, and, in particular, on the perception pupils have of their learning environments. It is left to readers to draw what implications they wish from the data — we prefer to be cautious and non-committal.

We believe that many more studies, similar to this one, are required for they, and other researches on teaching, can provide the data we need on the effect teaching has on pupils' development, e.g., pupils' achievement. In order to discover what teaching strategies influence learning it is necessary to study classroom procedures and their effects. It is not sufficient to infer teaching strategies from learning theories. This project is a contribution to the large task of providing information on effective classroom procedures and of developing our understanding of the teaching-learning process.

REFERENCES

Cohen, E. "Sociology and the Classroom: Setting the Conditions for Teacher-Student Interaction", *Review of Educational Research*, 1972, *42*, 441-452.

Grobman, H. *Evaluation Activities of Curriculum Projects*, A.E.R.A. Monographs on Curriculum Evaluation, Rand McNally, Chicago, 1968.

Flanders, N. and Simon, A. "Teacher Effectiveness", *Classroom Interaction Newsletter*, 1969, 5, No. 1, 18-37.

Heath R. and Nielson, M. "The Myth of Performance-Based Teacher Education". Paper presented at the annual meeting of the American Educational Research Association, New Orleans, 1973.

Hughes, D. "An Experimental Investigation of the Effects of Pupil Responding and Teacher Reacting on Pupil Achievement", *American Educational Research Journal*, 1973, *10*, 21-37.

Mitchell, J. "Education's Challenge to Psychology – The Prediction of Behavior from Person-Environment Interactions", *Review of Educational Research*, 1969, *39*, 695-721.

Nuthall, G. "Types of Research on Teaching", *New Zealand Journal of Educational Studies*, 1968, *3*, 125-147.

Power, C.N. *The Effects of Communication Patterns on Student Sociometric Status, Attitudes and Achievement in Science*, Ph.D. Thesis, University of Queensland, Brisbane, 1971.

Rosenshine, B. and Furst, N. "Research on Teacher Performance Criteria" in Smith, B (Ed.), *Research in Teacher Education*, Prentice Hall, Englewood Cliffs, 1971.

Steele, J., House, E. and Kerins, T. "An Instrument for Assessing Instructional Climate through Low-Inference Student Judgments", *American Educational Research Journal*, 1971, *7*, 447-466.

Tisher, R.P. and Power, C.N. "Congruency with A.S.E.P.: A Study of practices and values of a group of teachers and Dip.Eds.", *Australian Science Teachers' Journal*, 1973, *19*, (in press).

COMPUTERISED AUSUBEL – SUPPORT FOR AUSUBEL'S THEORY FROM COMPUTER SIMULATION STUDIES

Leo West

1. Introduction

In a post hoc analysis of some sixty studies of achievement predictors in senior secondary or early undergraduate sciences, Novak, Ring and Tamir conclude that

> "... general ability, general scholastic aptitudes interests, personality factors, biodate and possibly other variables alone or combined together account for no more than 50% of the variance regarding academic achievement. It seems highly probably that prior knowledge will account for the major portion of the remaining 50%." Novak et al. (1971) p.509.

If the relative size of the effect of prior knowledge on subsequent cognitive learning in science is anywhere near that suggested by Novak et al. then an understanding of the role that prior knowledge plays in learning should be a major interest of educational researchers and should have very important implications for the practice of science education.

The desire for a theoretical base to undergird the relationship between the learner's existing knowledge and his subsequent learning has led some science educators (e.g. Novak (1971), Kuhn (1972)) to see Ausubel's subsumption theory as providing such a base. Ausubel's theory has considerable intuitive appeal but it has not, as yet, been completely supported by empirical evidence.

The research relating to Ausubel's theory has tended to concentrate on his advance organizer notion. The results from this research has produced contradictions and conflicting results, but most of the apparent contradictions can be shown to be consistent with Ausubel's theory (West and Fensham in press). Very few attempts have been made to test empirically his more basic postulate that meaningful learning occurs when the learning can be related in a 'non-arbitrary and substantive' fashion to the learner's existing knowledge (cognitive structure).

Novak et al (1971) in a post hoc analysis of a large number of empirical studies demonstrated that the results could be interpreted in terms of Ausubel's theory, but this is not, as they point out, an adequate test of the theory.

Studies by Ring and Novak (1971), Kahle (1971) and Shovelson (1972) probably exhaust the field of studies that have produced support for Ausubel's theory of subsumption under prior knowledge.

In the absence of further experimental evidence, one can turn to alternative methods of theory testing. In recent years, physical scientists have created simplified models of complex systems in order to develop or test theories. The use of computed simulation models has played a major role in this technique. It is difficult to know how to assess the validity of a theory based on a program which simulates a human cognitive process. It is suggested by Hallworth (1969) that a program that simulates human learning behavior represents an *acceptable* theory of human learning. An acceptable theory is not a necessary theory. In this paper such an acceptable theory obtained from computer simulation of rote and meaningful learning is shown to be similar to the independently proposed theory of Ausubel which is not based on computer simulation. (In fact, Ausubel rejects the possibility of computer models providing support for his theory — Ausubel (1968) p.122). I believe that this 'triangulation' of theories provides support to the proposition that Ausubel's theory represents the process of meaningful learning in humans. The programs to be discussed in this paper are: The Teachable Language Comprehender (TLC) developed by Quillion (1969) which is designed to be capable of being taught to comprehend English text. This program represents a theory of meaningful verbal learning.

The Elementary Perceiver and Memoriser (EPAM) developed by Feigenbaum (1963) which is designed to learn lists of paired nonsense symbols. This program represents a theory of rote verbal learning.

Computer programs are precise. A theory based on a program will be a very detailed theory. A psychological theory (such as that of Ausubel) will not be as detailed. In comparing TLC and EPAM to Ausubel's theory, the comparison will be made at the *level of detail specified in Ausubel's theory*. The further detail in the theory derived from the programs, (which represent an acceptable theory of the precise processes involved in that particular type of learning) will not be discussed in this paper. It is of interest to observe that Collins and Quillion (1969) have produced evidence, from experiments with human subjects, that supports the detailed theory of memory structure derived from the TLC program.

II. TLC and Ausubel's theory of meaningful learning

Comparision between TLC and Ausubel's theory will be made under four headings:

(a) structure of the memory

(b) subsuming new learning

(c) an appropriate cognitive structure

(d) obliterative subsumption

(a) *Structure of the memory*

TLC has a N-dimensional hierarchical network memory, which contains 'units' and 'properties'. A unit represents a concept that TLC understands. Incorporated in each unit is a link to (a) its superset unit (more general concept) and (b) at least one modifying property. A property is an attribute : value pair such as colour : red but can also be a preposition : object pair such as up : hill or a verb : object pair such as paints : pictures. The attribute and value are also linked to other units. Thus a concept is always represented by (1) its superset of which it is a specific instance and (2) its properties stating how that superset must be modified to constitute that particular concept.

Ausubel's view of cognitive structure is that it is hierarchical in nature with concepts or propositions being subsumed under more inclusive concepts or propositions. Thus, in structure TLC's memory and Ausubel's cognitive structure are identical.

(b) *Subsuming new learning*

When TLC comprehends a piece of text, it begins by searching through its memory (in what approximates a parallel search) for an appropriate relationship between concepts it understands in the text. It will use this chosen section of its memory (cognitive structure) to make inferences about the meaning of the text. It may find, for example, in its memory network, intersections of the superset chains of the words in the text that it understands. It can use this network to infer the meaning of the text. This network enables it to create a representation of the input text which is "much richer and less ambiguous than the input text itself" and which is "a highly intraconnected structure with the various concepts mentioned in the text linked in ways supplied from a memory of facts about the world".

Ausubel uses two terms to describe the relation of new learning to existing knowledge in meaningful learning - the relationship must be "non-arbitrary" and "substantive". The non-arbitrariness "enables the use of previously acquired knowledge to internalize and make understandable", and the substantiveness "assimilates the substance of the idea rather than the precise words used in expressing them". These are also the results of TLC's method of relating text to memory. The use of network intersections enables the understanding of the text (and also its inclusion in memory by the creation of new units). The network arrangement leads to the 'comprehension' of the text rather than the internalization of the precise words.

(c) *An appropriate cognitive structure*

TLC, in its search for an appropriate representation of a piece of text, may find more than one such candidate cognitive structure. An example of this is given by Quillion. The text to be comprehended is "John shoots the teacher". TLC has a unit representing John with an associated property 'photographs subjects'. It also has a unit for teachers with an associated property 'teach child(ren)'. In its search TLC will find intersections of each of the superset chains beginning at John, subjects, teacher and children at the concept 'person'. Thus there will be a set of relationships between (i.e. cognitive structures relating) John and teacher, which TLC could use to infer the meaning of the text. Clearly, the choice of an inappropriate one, will lead to a misinterpretation and the creation in memory of a misconception for this particular meaning of 'shoot'. TLC includes tests (called 'form tests') which can be used to check the appropriateness of a particular candidate cognitive structure. As one would expect with the comprehension of English text, these form tests are based on simple syntactic feasibility.

Ausubel also considers the possibility of the use of inappropriate cognitive structures. Part of the role of advance organizers is to alert the learner to the most appropriate cognitive structure to be used to subsume the new learning. At the last A.S.E.R.A. conference, Fensham (1972) proposed that the use of less relevant cognitive structures may lead to misconceptions being formed and Ausubel has agreed that this is predictable from his theory (personal communication to Fensham).

TLC's form tests are not equivalent to Ausubel's advance organizers, but they play an equivalent role. For TLC, form tests ensure that the program uses the most appropriate cognitive structure for comprehension. For Ausubel, advance organizers ensure that the learner uses the most appropriate cognitive structure for meaningful learning.

(d) *Obliterative subsumption*

TLC's progress at this stage has been to comprehend a fairly limited range of simple phrases and sentences. It is is to proceed further, it will need a way of keeping its memory within limits. Quillion recognises this problem and suggests a solution:

> "TLC also makes no effort to get rid of most of the new units created during comprehension of text. Such new units represent known concepts plus things said about them in specific instances, and any adequate learning mechanism must forget such specific instances, while extracting any important generalizations from them and adding these to the more general concepts left in the memory. Some method of achieving such generalization and forgetting must probably be programmed before a significant amount of TLC's memory can actually be built up by reading text." p.473

Quillion is proposing that TLC needs to use what Ausubel has called obliterative subsumption. On this issue TLC fails as a total model of Ausubel's theory, but it is significant that Quillion should consider the process of obliterative subsumption as the direction in which TLC must develop in order to expand its capabilities.

III. EPAM and Ausubel's theory of rote learning

Ausubel views rote learning as the opposite to meaningful learning. Throughout the description of his theory, Ausubel highlights the difference between rote and meaningful learning. Thus he suggests that rotely learned material is related in an ARBITRARY and VERBATIM fashion to cognitive structure in comparison to the NON-ARBITRARY and SUBSTANTIVE fashion that he proposes for meaningful learning. As a consequence of this, Ausubel postulates for rote learning, unlike meaningful learning, that anchorage is not achieved, that variables such as frequence and reinforcement are important for learning; and that interference on the basis of inter-list and intra-list similarity and stimulus and response generalization are important variables for retention. (Ausubel (1968) p.110).

The correctness of this 'theory' of rote learning is important in Ausubel's argued development of his theory of meaningful learning. For instance, on a number of occasions he explains obstacles to the development of his theory by the rote : meaningful comparison (e.g. Ausubel (1968) p.115. 'Alternative theories of retention and forgetting').

EPAM has two learning processes, one to construct the images which form parts of its network memory, the other to build these images into the growing net.

The image building process copies a part of the stimulus symbol so that it is distinguishable from the other images. This will usually be less than (and never more than) the VERBATIM stimulus symbol.

The images are built into a discrimination net (memory) by the discrimination learning process. The net is not organized in any logical or hierarchical sense but simply on the basis of what is sufficient for discrimination at that particular point in its learning. Thus the relationship between parts of the net (i.e. cognitive structure) is completely ARBITRARY.

These two learning processes are equivalent to Ausubel's proposition that rote learning involves an arbitrary and verbatim relationship between new learning and cognitive structure. It is fascinating that the effects of inter- and intra-list similarity; of stimulus and response generalization; and of familiarization and meaningfulness on learning and retention in EPAM and EPAM-III have been shown to agree quantitatively as well as qualitatively with the data from identical experiments with humans (Simon and Feigenbaum (1964)).

IV Concluding Comments

The central argument presented in this paper is.

(1) that there exist computer simulation models of rote and meaningful learning which represent acceptable theories of these cognitive processes;

(2) that these theories are very similar to Ausubel's proposed theories of the same cognitive processes.

(3) that as a result, these computer simulation models support the proposition that Ausubel's theory represents the processes involved in rote and meaningful learning in humans.

The concept of computer simulation models raises a number of interesting ideas that have only been briefly mentioned in this paper.

One is the use of computer simulation models to propose detailed theories of cognitive processes. Feigenbaum (1964) has proposed a detailed 'Information Processing Theory of Memory' based on EPAM. He proposes a three level theory of memory:

"1. An Immediate Memory of very small size, in which information is stored for very brief intervals, which acts as a buffer storage to decouple the input (peripheral encoding) processes from the central processes and as a temporary storage for central processing.

2. An Acquisition Memory, a working memory with the structure of the EPAM discrimination net, in which discrimination learning takes place and in which the internal representations of stimulus objects are built.

3. A Permanent Storage in which the internal representations are organized and stored for long term retrieval."

Fiegenbaum (1964) p.14

EPAM is a precise formulation of the Immediate and Acquisition memories. It would be interesting to postulate a TLC type memory as a precise theory of permanent storage thus completing the model. This detailed model of memory could be used as a basis for designing research studies.

Another interesting idea can be derived from the use of 'form tests' in TLC. The role of these form tests is identical to the role of advance organizers, but they are not advance organizers. Perhaps we should be thinking of a general concept of 'organizational aid' of which advance organizers are a specific example rather than the narrow concept of advance organizers proposed by Ausubel.

REFERENCES

Ausubel, D.P. (1968). "Educational Psychology — A Cognitive View". Holt, Rinehart and Winston.

Collins, A.M. & Quillian, M.R. (1969). "Retrieval Time from Semantic Memory". *J. Verb. Learn. Verb. Behav.*

Feigenbaum, E.A. (1963). "The Simulation of Verbal Learning Behavior" in "Computers and Thought" Ed. Feigenbaum, E.A. & Feldman, J. McGraw-Hill.

Feigenbaum, E.A. (1964). "Elements of an Information Processing Theory of Memory". Paper presented to American Psychological Association, Annual Meeting, Los Angeles, California. September 9, 1964.

Fensham, P.J. (1972). "Prior Knowledge — A Source of Negative Factors". Research 1972. Ed. R.P. Tisher. University of Queensland.

Hallworth, H.J. (1969). "Computer Simulation of Cognitive Processes" *Educational Technology*, Oct. 1969, 60-63.

Kahle, J.B. (1971). "The effect of an advance organizer and the predictive ability of micro learning tasks when utilized with carefully sequenced audio tutorial units". Ph.D. Thesis, Purdue University, 1971.

Kuhn, D.J. (1972). "Science Teaching, Concept Formation and Learning Theory". *Science Education, 56(2)*, 189-196.

Novak, J.D. (1971). "Concept Learning in Science". *Theory into Practice, 10*, 129-133.

Novak, J.D., Ring, D.G., & Tamir, P. (1971). "Interpretation of Research Findings in Terms of Ausubel's Theory and Implications for Science Education". *Science Education, 55(4)*, 483-526.

Quillian, M.R. (1969). "The Teachable Language Comprehender". *Comm. ACM, 12(8)*, 459-476.

Shavelson, R.J. (1972). "Some Aspects of the Correspondence Between Content Structure and Cognitive Structure in Physics Instruction", *J. Educ. Psychol.*, *63*, 225-234.

Simon, H.A. & Feigenbaum, E.A. (1964). "An Information Processing Theory of Some Effects of Similarity, Familiarization and Meaningfulness in Verbal Learning", *J. Verb. Learn. Verb. Behav.* *3*, 385-396.

West, L.H.T. & Fensham, P.J. To be published in "Studies in Science Education, Leeds University.

THE MEASUREMENT OF A SCIENTIFIC ATTITUDE-CURIOSITY

R.B. Flegg and A.A. Hukins

Educators have included the development of scientific attitudes as an aim of science teaching for many years. The writings of Curtis (1932), Noll (1935), Heiss (1958) and Collette (1973) illustrate this fact. During that time, however, assessment of accomplishment of this aim has been generally omitted in the evaluation of student progress. Difficulty in measurement has been at least one factor responsible for this omission. Increasing pressure is now being exerted on the classroom teacher to take the aim seriously. Bollen (1972, p.217) has indicated that the assessment of attitudes has been part of the thinking on examinations which will emerge as a result of the Nuffield Secondary Science Project and the Schools Council Integrated Science Project in England. In New South Wales greater responsibility than hitherto is to be placed upon the teacher in the matter of assessment and the Secondary Schools Board (1972) has proposed that an attitude dimension is to make up 30% of the mark for the subject Science in the award of a School Certificate. There is an urgent need for clearer definition of objectives, and for tests to assist the teacher in his task.

In writings such as those cited above curiosity has generally been included as one of the desirable scientific attitudes. The present paper reports progress in an investigation which seeks to make a contribution to meeting the need for measuring instruments in relation to the scientific attitude curiosity.

Previous attempts to measure curiosity

Apart from some earlier work by Berlyne (1960) most of the reported investigations of exploratory behaviour involving human subjects date from the beginning of the 1960's. Pielstick and Woodruff (1964) and Peterson and Lowery (1968) used observations of children's exploration of objects in a specially arranged environment e.g. a waiting room, to measure curiosity.

Maw and Maw (1961) used a complex procedure involving peer, teacher and self ratings to establish criterion groups of high and low curiosity for subsequent investigations of the relationships of curiosity to certain other variables.

Methods which would be more easily employed by the classroom teacher were adopted by Penny and McCann (1964), Jenkins (1969) and Campbell (1971). Each of these used scales involving opinion statements though none of the instruments was of the generally accepted types of attitude scales.

As a first stage in the present investigation it was decided to develop an idea put forward at an earlier date by Hukins (1963). This used Berlyne's concept of epistemic curiosity. Berlyne (1960, p.274) described this as "the brand of arousal that motivates the quest for knowledge and is relieved when knowledge is procured". This appears to be related to the "thirst for new knowledge" of the cluster 5 obtained by Cohen (1971, p.141) in his analysis of the components of the scientific attitude.

An approach to the measurement of this kind of curiosity might be to produce a scale relating to the person's attitude to the seeking of knowledge.

Procedure

A Likert type scale was produced accordi ıg to the method outlined by Edwards (1957). The Likert scale was selected because it is regarded by some authors, e.g. Triandis (1971, p.45), as at least equal to and possibly superior to, the Thurstone or Guttman type scales for reliability and validity.

In order to obtain statements for the scale a set of questions was drawn up relating to the student's desire to gain knowledge, the strength of this desire as evidenced by his reaction to opposition to his questions, his reaction to new situations involving things, places or people, and his opinions about gaining information from sources such as people, books and the media. The questions were designed in such a way as to elicit extended answers from the statements. 69 from IV students in two high schools responded to these questions and from their answers 60 opinion sta' ments were selected.

The 60-item protoscale was administered to 180 students who had just completed third form in two high schools. The responses were subjected to the reliability maximization procedure of Ray (1972) to produce a smaller scale of high internal consistency. The 25-item "New Knowledge Scale", included as an appendix, was thus produced.

In an attempt to gain evidence about validity it was decided to form criterion groups of students high and low on curiosity, as rated by their teachers. Each teacher was asked to select the five or six most curious and the five or six least curious pupils in his class with the aid of a set of criteria. This has been carried out in a preliminary way only, at this stage, with five teachers producing "high" and "low" groups having a total of 28 in each. The attitude scale was also administered to these classes and the scores for the two groups compared.

It would be helpful to know the relationship, if any, between attitude to knowledge-seeking and other variables such as science achievement and intelligence. Maw and Maw (1970) have reported a correlation of .36 between curiosity and intelligence. A science achievement score was obtained from the schools and I.Q. data is being sought also.

Results

The value for the reliability coefficient obtained by the procedure of Ray was 0.80.

The null hypothesis, there is no difference in scores on the attitude scale between those who are rated high in curiosity by their teachers and those who are rated low, was tested by the Kolmogorov-Smirnov two-sample test (Siegel 1956). With $N = 28$ a value of $K_D = 10$ is required for rejection at the .05 level. The obtained value was $K_D = 7$. Thus, although the difference in the cumulative frequencies was in the direction which might be expected, it was not sufficiently great to reject the null hypothesis.

A correlation coefficient for science achievement and attitude to knowledge-seeking was calculated by the Spearman rank-difference method. The value of rho was found to be 0.20 indicating a very small relationship.

Discussion

The comparison between attitude to knowledge-seeking as measured by the attitude scale and as rated by teachers has not led to a definite conclusion because the numbers of teachers involved was too small. However the result suggests that it would be worthwhile pursuing this line of investigation. This assumes that teachers are able to distinguish between the most and least curious students in their classes on the basis of the criteria given.

An internal-consistency measure of reliability is involved in the coefficient reported above. It would be advantageous to know the test-retest reliability coefficient and this is currently being investigated.

The investigation reported here relates to a self-report technique for measuring a scientific attitude, curiosity. A question to be considered is to what extent such instruments, useful in research situations, are useful to the classroom teacher in arriving at a score to be included in the assessment of his students. They would probably be much less valid if the subjects felt that they were under some pressure because the results were to count towards an award e.g. the School Certificate in New South Wales. Another question which many educators would like to consider further is whether, on ethical grounds, the assessment of attitudes should be taken into account in the granting of an award.

REFERENCES

Berlyne, D.E. *Conflict, Arousal and Curiosity*, New York, McGraw-Hill, 1960.

Bollen, F.A. "Attitude Assessment in Science Teaching." *School Science Review*, 54, 217-235, 1972.

Campbell, J.R. "Cognitive and Affective Process Development and Its Relation to a Teacher's Interaction Ratio". *Journal of Research in Science Teaching*, 8, 317-324, 1971.

Cohen, David. "Can Scientific Attitudes be Evaluated", *Research 1971*, pp.135-143.

Collette, A.T. *Science Teaching in the Secondary School*, Boston, Allyn and Bacon, 1973.

Curtis, F.D. "A Synthesis and Evaluation of Subject Matter Topics in General Science" in *A Program for Teaching Science*, Thirty-First Yearbook of the National Society for the Study of Education, Chicago, 1932.

Edwards, A.L. *Techniques of Attitude Scale Construction*, New York, Appleton-Century-Crofts, 1957.

Heiss, E.D. "Helping Students Develop a Scientific Attitude", *The Science Teacher*, 25, 371-373, 1958.

Hickins, A.A. *A Factorial Investigation of Measures of Achievement of Objectives in Science Teaching*, Unpublished Ph.D. Thesis, University of Alberta, 1963.

Jenkins, J.A. "An Experimental Invesigation of the Effects of Structured Science Experiences on Curiosity Among Fourth Grade Children", *Journal of Research in Science Teaching*, 6, 128-135, 1969.

Maw, W.H. and Maw, E.W. "Establishing Criterion Groups for Evaluation Measures of Curiosity", *Journal of Experimental Education*, 29, 299-305, 1961.

Maw, W.H. and Maw, E.W. "Self-Concepts of High- and Low-Curiosity Boys", *Child Development*, 41, 123-129, 1970.

Noll, V.H. "Measuring the Scientific Attitude", *Journal of Abnormal and Social Psychology*, 30, 145-154, 1935.

Penny R.K. and McCann, B. "The Children's Reactive Curiosity Scale", *Psychological Reports*, 15, 323-334, 1964.

Peterson, R.W. and Lowery, L.F. "A Study of Curiosity Factors in First Grade Children", *Science Education*, 52, 347-352, 1968.

Pielstick, N.L. and Woodruff. "Exploratory Behavior and Curiosity in Two Age and Ability Groups of Children", *Psychological Reports*, 14, 831-838, 1964.

Ray, J.J. "A New Reliability Maximization Procedure for Likert Scales", *Australian Psychologist*, 7, 40-46, 1972.

Secondary Schools Board. *Circular Memorandum No. 41 to School Principals*, Sydney 1972.

Siegel, S. *Nonparametric Statistics*, New York, McGraw-Hill, 1956.

Triandis, H.C. *Attitude and Attitude Change*, New York, Wiley, 1971.

NAME .... CLASS: ....

THE NEW KNOWLEDGE ATTITUDE SCALE

Here are some statements from students about the way they think and act. As you read them you will find yourself agreeing with some, disagreeing with some and perhaps you will be undecided about others. Your job is to say just how much you agree or disagree with each of the statements. Remember that no matter which choice you make there will be many people who think the same way as you do.

HERE'S WHAT TO DO:

Each statement is followed by the numbers: 1, 2, 3, 4, 5.

If you *strongly agree*, put a circle around 1.

If you *agree*, put a circle around 2.

If you are *undecided*, put a circle around 3.

If you *disagree*, put a circle around 4.

If you *strongly disagree*, put a circle around 5.

FOR EXAMPLE:

Supposing there were a statement like the following and you strongly agreed with it. This is how you would indicate your choice:

	Strongly agree	Agree	Undecided	Disagree	Strongly Disagree
88. Too much knowledge can drive a person crazy.	1	2	3	4	5

Give your judgement on EVERY statement.

Work carefully through each item and don't spend too much time on any one of them.

STATEMENTS	Strongly agree	Agree	Undecided	Disagree	Strongly disagree
1. I don't have courage enough to question the teaching of strangers.	1	2	3	4	5
2. I prefer to learn more about familiar things rather than tackle new ones.	1	2	3	4	5
3. Knowledge keeps your mind active.	1	2	3	4	5
4. I try to look as if I understand things, even if I don't.	1	2	3	4	5
5. It doesn't matter how a thing works, what it does is more important.	1	2	3	4	5
6. I'm too lazy to look up answers to my queries in books.	1	2	3	4	5
7. Visiting new places is a joy for me.	1	2	3	4	5
8. If I could get what I want in life without knowing a thing, I would.	1	2	3	4	5
9. I get very enthusiastic when I meet people with new ideas.	1	2	3	4	5
10. Inventions are intriguing.	1	2	3	4	5
11. If people refuse to tell you things, you want to find out all the more.	1	2	3	4	5
12. I like to know what makes things tick.	1	2	3	4	5
13. Knowing too much would make life very boring.	1	2	3	4	5
14. I'd like to get a good look at things out there in space.	1	2	3	4	5
15. Other people's inventions give me ideas.	1	2	3	4	5
16. I won't rest until I find out about something which intrigues me.	1	2	3	4	5
17. It's a waste of time to look things up in a library.	1	2	3	4	5
18. A person with an enquiring mind is a threat to those around him.	1	2	3	4	5
19. I only read the comics in the newspaper.	1	2	3	4	5
20. It is almost criminal to be the least bit intelligent.	1	2	3	4	5
21. Strange places are fascinating.	1	2	3	4	5
22. I enjoy looking at the newspaper headlines to find something that might interest me.	1	2	3	4	5
23. I like to learn things from people who put them forward in a new way.	1	2	3	4	5
24. I like to stay in familiar surroundings.	1	2	3	4	5
25. I like keeping my mind free of problems.	1	2	3	4	5

A NOTE ON A POSSIBLE ALTERNATIVE TO LIKERT SCALES

R.T. White and L.D. Mackay

Criticisms of Likert scales

Likert scales are widely used in tests of attitudes. The test of curiosity developed by Flegg and Hukins (in this issue) is a typical test of this type. Three general criticisms of Likert scales are made below; these criticisms are not meant to apply specifically to Hukins' test, which is constructed in accord with the accepted principles of Likert scales, but are an attack on the accepted principles themselves. A suggestion for an alternative procedure is illustrated by reference to the measure of curiosity.

Criticism 1. Likert scales lack validity because it is often simple for the respondents to perceive the intent of the items, and then those who have a set to score positively (or negatively) on the dimension being measured will slant their responses accordingly. The test then measures a complex combination of the respondent's standing on the dimension being measured, of his ability to perceive the intent of a set of items, and of his set to respond in a particular way.

Criticism 2. Scores on 5-point Likert scales are affected by the dogmatism of the respondent. More dogmatic people will tend to choose extreme values, and so will appear more (or less) curious than someone less dogmatic though of really equal curiosity. The only exception is for people whose total score on the test is equal to the mid-point of the range of possible scores — their dogmatic tendency in one direction is exactly cancelled by their tendency in the other direction.

Criticims 3. Respondents may have a general positive or negative set in answering Likert-scale items. Some people may tend to choose the "favourable" responses of "agree" or "strongly agree", or whatever the phrasing of the responses is, while others may tend to choose the "unfavourable" responses. Their scores on the scale therefore are a composite of their set as well as their standing on curiosity, or whatever dimension the scale is attempting to measure.

To meet these criticisms the alternative procedure must disguise the intent of the test, at least to a greater degree than the Likert scale; and it must be free of the influence of irrelevant attributes such as dogmatism and positive-negative response set. The procedure described below appears to have these properties.

Alternative procedure

Collect a number of statements that describe the writer, in the first person, e.g. Flegg and Hukins' item 9: "I get very enthusiastic when I meet people with new ideas", and which clearly have some direct relevance to curiosity. Other statements of this type are Hukins' items 2, 4, 6, 8, 12, 14, 16, 19, 22, 23, 24, and 25 — though some of these, e.g. 2 and 24, are negatively or anti-curiosity phrased, in a logical attempt by Flegg and Hukins to meet criticism 3.

Collect a larger number of statements of similar form, which do not have relevance to curiosity but which relate to a diverse set of attributes. Examples are:

> "I am generally kind to animals"
> "I am not scared of dogs"
> "I like team games rather than individual sports"
> "I prefer going to the football to going to the pictures"
> "I am too shy to give my opinion to a large group".

Reproduce both sets of statements on cards which contain one statement each. Shuffle the order of the cards. Give the cards to a number of respondents, and ask them to rate the statements in nine ordered categories in regard to the question "Which of these statements would you most like pe ple to say about you?" The procedure of placing the statements in the nine categories is described by Cohen (1971). No mention of curiosity is made, so that the respondents are classifying the statements purely on the degree of their social acceptability. The categories are marked 1 to 9, and the mean value is calculated for each statement. Alternatively, the median category may be found for each statement.

Each curiosity statement is then paired with a general, or "background" statement of equal mean (or median) social acceptability. The test of curiosity then consists of the set of pairs of statements, which are given to the testees with the instruction "Mark the one of each pair that describes you better". The scoring of the test is simply a count of the number of times that the curious statements in the pairs are selected.

The above outline does not go into the details of wording and procedure, which would be crucial in tests of this type. It is presented as a suggestion for constructors of new tests in science education research to conside. The authors are currently attempting to produce a test of this type to measure the congruence between children's and scientists' beliefs of how scientists think and act, which should be a useful instrument as part of a battery for evaluation of science curricula.

REFERENCES

Cohen, D. "Can scientific attitudes be evaluated?" *Research 1971*, R.P. Tisher (ed.), Brisbane: Australian Science Education Research Association, 1972, 135-143.

Flegg, R.B. and Hukins, A.A. "The measurement of a scientific attitude – curiosity". This issue.

A TEST OF ENQUIRY SKILLS

Barry J. Fraser

Purpose of the Test

The worth of any evaluation instrument depends largely upon the purpose for which it is to be used. Before discussing the development of *A Test of Enquiry Skills* (TOES), therefore, the use to which TOES is to be put must be clearly understood.

The main use envisaged for TOES is in the summative evaluation of the Australian Science Education Project (ASEP) curriculum. The specifications for designing TOES, then, need to be consistent with contemporary theoretical considerations in curriculum evaluation.

Curriculum evaluation theory

For the purposes of this article, the sense in which the term summative curriculum evaluation will be used needs to be clarified in two ways. Firstly, discussion here will be restricted to *hard* evaluation, which involves pupils' achievement of desirable goals, and will ignore *soft* evaluation, which includes experts' opinions of ASEP materials, teachers' impressions of classes studying ASEP, etc. (Further discussion of hard and soft evaluation can be found in Fraser, 1973 and Whitfield and Kerr, 1970).

Secondly, curriculum evaluation will *not* be defined simply as the determination of whether or not a particular curriculum is effective (or more effective than an alternative). Instead, the definition adopted here will be the more complex one given by Carroll (1965, p.253) who claims

> "I would define curriculum evaluation as a process . . . . of determining which objectives it can attain, under what conditions, and for what kinds of pupils."

For curriculum evaluation defined in this way, it is possible to judge the attainment of objectives of any curriculum either against some absolute standard or against the attainment under alternative curricula.

A definition of curriculum evaluation similar to that given above is also reflected in Walberg's model (1969) which is expressed by:

$$L_h = f(I_i, A_j, E_k)$$

In this equation, L_h stands for the set of learning outcomes, I_i for the instruction or curriculum, A_j for the aptitudinal and personological attributes of the pupil and E_k for the learning environment. By using such a theoretical model as a basis for curriculum evaluation, it is possible not only to provide more useful and comprehensive information but also to "contribute to basic educational research and to the formulation of verifiable principles and theories of instruction". (Walberg, 1969, p.185).

The acceptance of the above definition of curriculum evaluation has important implications for the design of any instrument of curriculum evaluation. Consequently, for the present purposes of developing a test of enquiry skills, a set of specifications consistent with this definition will now be established.

Specifications for the Test

Multidimensional Criterion

From the above discussion it follows that curriculum evaluation should cover a comprehensive range of as many different learning outcomes as possible. Indeed, as Cronbach (1969, p.45) advocates, "an ideally suitable battery for evaluation purposes will include separate measures of all outcomes . . . . considered important". Similarly, Stake (1967, p.536) has described his commitment to a multidimensional criterion of curricular effectiveness by saying

> "It is a great misfortune that the best trained evaluators have been looking at education with a microscope rather than with a panoramic viewfinder."

For the present purposes of test development, then, it is desirable to construct an instrument which provides separate measures of a wide variety of student outcomes.

Important Pupil Outcomes

For any curriculum evaluation instrument, an overriding consideration is that each outcome measured must be an important objective in science education. Consequently, curriculum evaluation should be based upon the goals that a curriculum *ought* to attain rather than those *stated* for that curriculum. If the goals stated for a curriculum are unimportant, then their achievement is no indication of the efficacy of that curriculum. As Scriven (1967, p.52) points out, "it is obvious that if goals aren't worth achieving then it is uninteresting how well they are achieved". Clearly, if curriculum evaluation is based solely upon curricula's own goals, the curriculum deemed most effective could simply be the one with the least enterprising goals. This opinion is shared by Cronbach (1963) who claims that

> "an ideal evaluation would include measures of all the types of proficiency that might reasonably be desired in the area in question, not just the selected outcomes to which the curriculum directs substantial attention."

In an attempt to obtain a comprehensive statement of all important aims in science education, an extensive literature survey was conducted. This survey included a systematic scrutiny of all articles over the last ten years in *Science Education, The Journal of Research in Science Teaching*, and *The Australian Science Teachers' Journal*, as well as the large number of cross-references found in the bibliographies of these articles. Examples of important sources of aims include Hurd (1969), Hurd (1970), Lombard (1965), Aims Committee of the Newcastle and Districts Association of Science Teachers (1967), Science 5-13 (1971), and the NSSE Yearbooks (1960 and 1947). From this literature survey, it was found that Klopfer's classification of science objectives (1971) was the most comprehensive one and embraced almost all outcomes stated elsewhere. For this reason, Klopfer's scheme was used to classify the outcomes listed in all other aims statements.

For the purposes of test construction, it was decided to restrict attention to Klopfer's categories called the *process of scientific inquiry*. From the literature, ample evidence was collected to support the importance of these goals, which have formed the basis for the development of *A Test of Enquiry Skills*. The scope of the skills measured in TOES is shown in Table 1.

TABLE 1

Scope of **A Test of Enquiry Skills**

PART A REFERENCE MATERIALS

Skill 1 : Library Usage

Skill 2 : Index and Table of Contents

PART B INTERPRETING & PROCESSING INFORMATION

Skill 3 : Scales

Skill 4 : Averages, Percentages and Proportions

Skill 5 : Charts and Tables

Skill 6 : Graphs

PART C CRITICAL THINKING IN SCIENCE

Skill 7 : Comprehesnion of Science Reading

Skill 8 : Design of Experimental Procedures

Skill 9 : Conclusions and Generalizations

Aptitudes for the Enquiry Approach

Reconsideration of Walberg's model, described above, shows that both aptitudinal and environmental variables are of importance in curriculum evaluation. The relevance of some of these variables in the case of ASEP can be inferred from the nature of the ASEP course. In particular it is interesting to note that the ASEP curriculum is based upon the *enquiry approach* (ASEP, 1970). With this approach the pupil is engaged predominantly in independent study using ASEP booklets to provide reading material, instructions for conducting experiments and questions to answer. Furthermore, with any approach involving independent pupil enquiry, it is reasonable to assume that there are a number of attitudes and skills that are prerequisite for success. For example, motivation to work independently is clearly an important attitude in this situation.

In addition to attitudes, however, the use of independent enquiry assumes pupil mastery of a host of cognitive skills. For instance, it is assumed that the student can find relevant reference materials in the library and can interpret science reading material provided. It is assumed, also, that students can read the scales of instruments used in experiments and can record their measurements in tabular or graphical form. Similarly, the independent enquiry approach assumes that students have some ability to set up apparatus, to select appropriate instruments and to draw valid conclusions from data. In other words, the skills listed in Table 1 as criteria of curriculum effectiveness (independent variables) could also serve as aptitudinal (dependent) variables affecting pupil achievement of other goals in science education. For a discussion of these dual roles played by enquiry skills, reference can be made to Newport (1972).

For the above reasons, a test of enquiry skills can be justified in curriculum evaluation, not only as a criterion measure, but also for use as aptitudinal variables. Such aptitudinal variables could not only influence pupil success on a wide variety of outcomes but could also interact differentially with alternative curricula to yield the *aptitude-treatment interactions* discussed by Salomon (1971).

Content-free Outcomes

If one wished to compare the ASEP course with a traditional one, a problem would immediately arise since the content covered in the two courses would be markedly different. Furthermore, even if ASEP was evaluated without such a control group, the problem of ascertaining the content covered by the ASEP sample would still be acute. The reason for this is that it is part of ASEP's philosophy that different ASEP teachers are free to choose different units for their classes whilst different students within a class may choose different options from a given unit. Consequently, in curriculum evaluation, it would be invalid to insist that all students should cover identical ASEP material since this would be alien to the way ASEP materials are intended to be used.

One way of overcoming this problem of determining common content in curriculum evaluation is to focus upon *content-free* outcomes. In curriculum evaluation then, content-free tests are more widely applicable than ones covering specific content. For this reason, an enquiry skills test would be particularly useful in curriculum evaluation because of the content-free nature of such enquiry skills.

Form 1 Level

Because of the relatively slow changes that would be expected in students' enquiry skills (and other content-free outcomes) over time, it was decided to concentrate ASEP evaluation at the Form 1 level. The reason for this is that the ASEP curriculum is likely to have more impact on Form 1 students who would have negligible prior background in science than on pupils who have had some years of experience in other science courses.

Consequently, in writing TOES, care is needed to ensure that items are suitable for the reading level and stage of development of Form 1 students. In practice, the difficulty of vocabulary used in items was controlled using Gardner's list of words commonly used in science (1972).

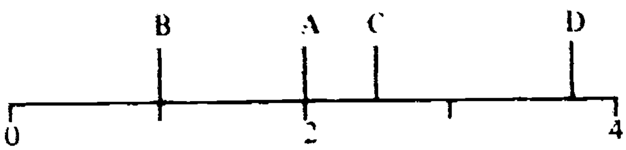
Sensitivity of the Instrument

A defining characteristic of curriculum evaluation has been taken here to be its concern for the differential effectiveness of any curriculum for different students. Consequently, if attention is to be directed to the achievement of pupils of *all* ability levels, it is inappropriate to aim test item difficulties at the "average" child. Such a bias would fail to provide information about the achievement of both high and low ability students. Instead, an instrument with appropriate sensitivity would provide a *wide range of item difficulties* compatible with the ability level of all students in the target population. The importance of this wide range in item difficulty levels has been stressed by Tyler (1966).

Furthermore, in developing TOES, it was assumed that each enquiry skill would be mastered to varying degrees by different pupils. For example, the skill of reading a scale could be tested at progressively more complex levels, as shown in Figure 1. In this figure, reading the scale at point A is relatively easy, at point B a little harder, at point C still more difficult and at point D the most difficult of the four.

During construction of TOES, then, an important specification is that each skill should contain a range of item difficulties compatible with Form 1 students' range of ability in each skill. In practice, this was attempted both by interviewing Form 1 students and by obtaining the subjective judgements of experienced Form 1 teachers.

FIGURE 1



Multiple Choice Format

The definition of curriculum evaluation employed earlier stipulated both the use of a multidimensional criterion and consideration of aptitudinal and environmental variables. The implementation of such an approach to curriculum evaluation, however, incurs considerable expenditure in test administration and in analysis. For this reason, it was decided to use multiple-choice format for the test so as to economize on testing and correction time.

Once the decision to employ multiple-choice tests is made, however, it is important to differentiate clearly between the *analytical mode* and the *constructive mode* described by Schwab (1963, p.473). The analytical mode of testing could require the student to interpret a graph, to select from alternative hypotheses, or to choose the best of alternative experimental set-ups. Clearly, multiple-choice testing is appropriate for measuring all these skills in analytical mode. On the other hand, however, the same skills tested in the constructive mode would require a student to construct a graph, to form his own hypotheses or to describe an appropriate experimental set-up. Clearly, multiple-choice items cannot measure these skills in the constructive mode. Consequently, when interpreting results obtained using TOES, it is important to appreciate that data were obtained for the analytical mode and are not necessarily generalizable to the constructive mode. Nevertheless, it is feasible that performance in the analytical mode could be highly related to that in the constructive; this question, however, requires further investigation.

Development of the trial version

In writing the trial version of TOES, three measures were taken to ensure the validity of items and their suitability at the Form 1 level. Firstly, extensive use was made of the literature both to clarify the meaning and scope of each skill and to be sure that each item measured an objective important in science education.

Secondly, an extensive review of both published and unpublished evaluation instruments was used as a source of ideas for test items. Examples of tests that proved useful for this purpose are those developed by the New South Wales Department of Education (1963), Science Research Associates (1958), Nelson (1967), Research and Curriculum Branch in Queensland (1972), Australian Science Education Project (1972), Gardner (1969), Biological Sciences Curriculum Study (1962), Dressel and Mayhew (1954), Lindquist (1952), Morse and McCune (1964) and Watson and Glaser (1964).

Thirdly, all items written were subjected to the scrutiny of a panel of experts in educational measurement and in science education. By means of these panel sessions, the face validity of items was checked to ensure that each item did measure the skill to which it was allocated, that items were suitable for the Form 1 ability range and that items were well constructed and free of clues or ambiguities.

The 117 items which survived the panel sessions were then assembled into a trial battery and administered to 14 Melbourne Form 1 classes, each from a different school. In each of these classes, a random third of the class sat for Part A of the battery, another random third took Part B and the rest of each class took Part C. Furthermore, in order to prevent fatigue, the test administration was spread over two class periods occurring on different days.

Finally, in order to illustrate the nature of the items included in TOES, an example of one item from each of the nine skills is given in the Appendix.

Statistical analysis of the trial version

Tests as a Whole

Table 2 shows data for the battery as a whole, including the sample size, the number of test items, the mean and standard deviation and the Kuder-Richardson 20 reliability of each part of the test and each individual skill in the battery.

For all tests in the battery, with the exception of Skills 3 and 5, it was found that satisfactory sensitivity had been achieved. That is, except for two skills, each test had a satisfactory spread of item difficulties ranging from very easy to very hard. On the other hand, for both Skills 3 and 5, the tests were found to be insensitive for measurements of high proficiency since 28 of the 29 items in these tests were correctly answered by at least 60 percent of students. The reason for this occurrence was that, during item-writing, the current level of Form 1 pupils' proficiency in these skills had been under-estimated. In order to obtain tests with appropriate sensitivity for these skills, it will be necessary to rewrite the tests adding a number of more difficult items, to administer the new version to a sample of pupils and then to analyze the new data.

TABLE 2

Statistics for Tests as a Whole

Test	Number of Students	Number of Items	Mean	Standard Deviation	$K-R_{20}$ Reliability
Skill 1	151	15	8.6 (57%)	2.4	0.53
Skill 2	151	11	7.8 (71%)	2.4	0.70
All Part A	151	26	16.4 (63%)	4.2	0.74
Skill 3	147	13	9.7 (75%)	2.6	0.74
Skill 4	147	9	4.3 (48%)	2.7	0.83
Skill 5	147	16	13.7 (86%)	2.4	0.73
Skill 6	147	15	9.3 (62%)	3.5	0.79
All Part B	147	53	37.1 (70%)	9.5	0.92
Skill 7	134	15	10.5 (70%)	2.7	0.72
Skill 8	134	10	5.7 (57%)	1.8	0.46
Skill 9	134	13	7.0 (54%)	2.2	0.54
All Part C	134	38	23.1 (61%)	5.3	0.78

Finally, since it is assumed that each test in the battery does measure a unique skill, it is important that each test has reasonable internal consistency, as indicated by its Kuder-Richardson 20 reliability. Nevertheless, it must also be appreciated that the size of the $K-R_{20}$ reliability is restricted by the shortness of each skill test. Since the definition used here for curriculum evaluation stipulates a multidimensional criterion, TOES has been designed to maximize the number of separate skills whilst practical considerations have limited the length of each skill test to 10 to 15 items. Furthermore, since TOES is designed to measure group performance, a much lower $K-R_{20}$ reliability is acceptable than for tests used to discriminate among individuals.

Based on the reasons outlined above, it was decided that values above 0.50 would be accepted for the $K-R_{20}$ reliability of each skill in TOES. Examination of the $K-R_{20}$ data provided in Table 2 shows that a value above 0.50 was obtained for all skills except Skill 8. Nevertheless, upon removal from Skill 8 of the two items that were not significantly correlated with the skill total, the value of the $K-R_{20}$ reliability for this skill did rise above 0.50. Consequently, for the purposes for which TOES was designed, all skills had acceptable $K-R_{20}$ reliability.

Individual Items

Table 3 provides some statistics for each of the nine test items shown in the Appendix. The data include the percentage of students correct, the point-biserial correlation of each item with its test Part together with its significance level, and the correlation of each item with all skills within its Part of the battery. The correlations for each item are unbiassed in that they were calculated after that particular item had been removed. Also, the correlations of items with each skill have been corrected for attenuation to account for the variation in reliability between different skill tests.

TABLE 3

Statistics for Individual Items

Part	Skill	Item	% Correct	Unbiassed Correlation with Part	p	Unbiassed Correlation* with Skills			
						Skill 1	**Skill 2**		
A	1	3	57	0.33	0.0001	0.39	0.36		
A	2	19	82	0.45	0.0001	0.47	0.54		
						Skill 3	**Skill 4**	**Skill 5**	**Skill 6**
B	3	35	65	0.55	0.0001	0.63	0.53	0.49	0.47
B	4	40	46	0.51	0.0001	0.50	0.69	0.39	0.43
B	5	57	89	0.26	0.001	0.21	0.18	0.40	0.25
B	6	71	58	0.39	0.0001	0.35	0.38	0.29	0.43
						Skill 7	**Skill 8**	**Skill 9**	
C	7	87	79	0.46	0.0001	0.65	0.28	0.41	
C	8	101	84	0.28	0.001	0.28	0.43	0.18	
C	9	106	50	0.14	0.05	0.08	-0.06	0.40	

* These correlations have been corrected for attenuation.

The data in Table 3 can be used in three ways to guide the writing of a final version of TOES. Firstly, where a skill test has numerous items of the same difficulty level, the removal of some of these items would reduce test length without adversely affecting the test's sensitivity. Secondly, since each test is intended to measure a unique skill, the construct validity of each test can be improved by removing each item not significantly correlated with both the skill and Part to which it belongs. Thirdly, if an item is located within the correct skill, its correlation with its own skill should be larger than its correlation with any other skills contained in the same test Part. Consequently, the removal of any item that diverges much from this criterion would also strengthen the construct validity of a skill test.

Writing the final version

From the previous discussion, a summary can now be made of suggestions to guide the rewriting of TOES. Firstly, the two skill tests with inappropriate sensitivity will need to be rewritten, administered again and re-analyzed. Secondly, to improve construct validity, it is advisable to remove those items which are either not significantly correlated with tests as a whole or whose correlation is higher with other skills in the battery than with its own skill. Once items are removed from tests, however, it is desirable to ascertain the new statistical characteristics of the battery by re-analyzing trial data after those items have been removed.

For the items that remain after applying the procedures above, only very minor changes should be made in wording, labelling of diagrams or choice of distractors. These minor changes could be based upon the opinions of measurement experts who have scrutinized TOES, information obtained by interviewing the students who trialled TOES, or from the distractor analysis from the administration of TOES.

REFERENCES

Aims Committee of the Newcastle and Districts Association of Science Teachers. "Objectives for Senior Science Courses", *The Australian Science Teachers' Journal*, 38, 45, 1967, 45-53.

Australian Science Education Project (ASEP). "Document 38: Use of the Inquiry Approach", Melbourne: ASEP, 1970.

Australian Science Education Project (ASEP). "Averaging and Reading a Scale Student Questionnaire", Melbourne: ASEP, 1972.

Biological Sciences Curriculum Study (BSCS). *Processes of Science Test*, New York: Psychological Corporation, 1962.

Carroll, John B. "School Learning Over the Long Haul", Ch. 10 of J.D. Krumboltz (ed.), *Learning and the Education Process*, Chicago: Rand McNally, 1965, 249-269.

Cronbach, Lee J. "Evaluation For Course Improvement", *Teachers' College Record*, 64, 8, 1963, 672-683.

Cronbach, Lee J. "Validation of Educational Measures", *Proceedings of the 1968 Invitational Conference on Testing Problems*, New Jersey: Educational Testing Service, 1969, 35-52.

Dressel, Paul L. & Mayhew, Lewis B. "Pervasive Objective 1: Critical Thinking", Ch. 7 of *General Education Explorations in Evaluation*, Washington: American Council on Education, 1954.

Fraser, Barry J. "Curriculum Evaluation Part 1: Hard or Soft?". *The Australian Science Teachers' Journal*, 19, 2, 1973 (In press).

Gardner, Paul L. *Biology Readiness Test Materials*, Melbourne: Australian Council for Educational Research, 1969.

Gardner, Paul L. for ASEP. *Words in Science* Melbourne: ASEP, 1972.

Hurd, Paul Dehart. *New Directions in Teaching Secondary School Science*, Chicago: Rand McNally, 1969.

Hurd, Paul Dehart. *New Curriculum Perspectives for Junior High School Science*, Belmont, California: Wadsworth Publishing Co., 1970.

Klopfer, Leopold E. "Evaluation of Learning in Science", In Benjamin S. Bloom *et al.* (Eds.), *Handbook on Summative and Formative Evaluation of Student Learning*, New York: McGraw-Hill, 1971, 561-641.

Lindquist, E.F. (Ed.). "Test 9: Use of Sources of Information", *The Iowa Test of Educational Development* (ITED), Illinois: Science Research Associates, 1952.

Lombard, John W. & Owen, William B. "Objectives of Science Education", Chicago: Science Research Associates, (ERIC), December, 1965.

Morse, Horace T. & McCune, George H. *Selected Items for the Testing of Study Skills and Critical Thinking*, Washington: National Council for the Social Studies, 1964.

National Society for the Study of Education (NSSE). *Science Education in American Schools*, Forty-sixth Yearbook, Chicago: University of Chicago Press, 1947.

National Society for the Study of Education (NSSE). *Rethinking Science Education*, Fifty-ninth Yearbook, Chicago: University of Chicago Press, 1960.

Nelson, Clarence H. *Improving Objective Tests in Science*, Washington: National Science Teachers' Association, 1967.

Newport, John F. "Process: Ends or Means or Both?". *Science Education*, 56, 2, 1972, 139-141.

New South Wales Department of Education. *N.S.W. Tests of Basic Skills*, Melbourne: Australian Council for Educational Research, 1963.

Research and Curriculum Branch. *Graph Test*, Queensland: Department of Education, 1972.

Salomon, Gavriel. "Heuristic Evaluation Models for the Generation of Aptitude-Treatment Interaction Hypotheses", *Review of Educational Research*, 42, 3, 1971, 327-343.

Schwab, Joseph J. *Biology Teachers' Handbook* New York: John Wiley, 1963.

Science 5 13. "Objectives for Children Learning Science", Bristol: MacDonald & Co., 1971.

Science Research Associates (SRA). "Test 1. Work-Study Skills", *SRA Achievement Series*, 6 9, Illinois Science Research Associates, 1958.

Scriven, Michael. "The Methodology of Evaluation", *Perspectives of Curriculum Evaluation*, AERA Monograph Series on Curriculum Evaluation 1, Chicago: Rand McNally, 1967, 39-83.

Stake, Robert E. "The Countenance of Educational Evaluation", *Teachers' College Record*, 68, 7, 1967, 523-540.

Tyler, Ralph W. "The Development of Instruments for Assessing Educational Progress", *Proceedings of the 1965 Invitational Conference on Testing Problems*, New Jersey Educational Testing Service, 1966, 95-105.

Walberg, Herbert J. "A Model for Research on Instruction". *School Review*, 78, 2, 1969, 185-200.

Watson, Goodwin & Glaser, Edward M. *Watson-Glaser Critical Thinking Appraisal*, New York, Harcourt, Brace & World, 1961.

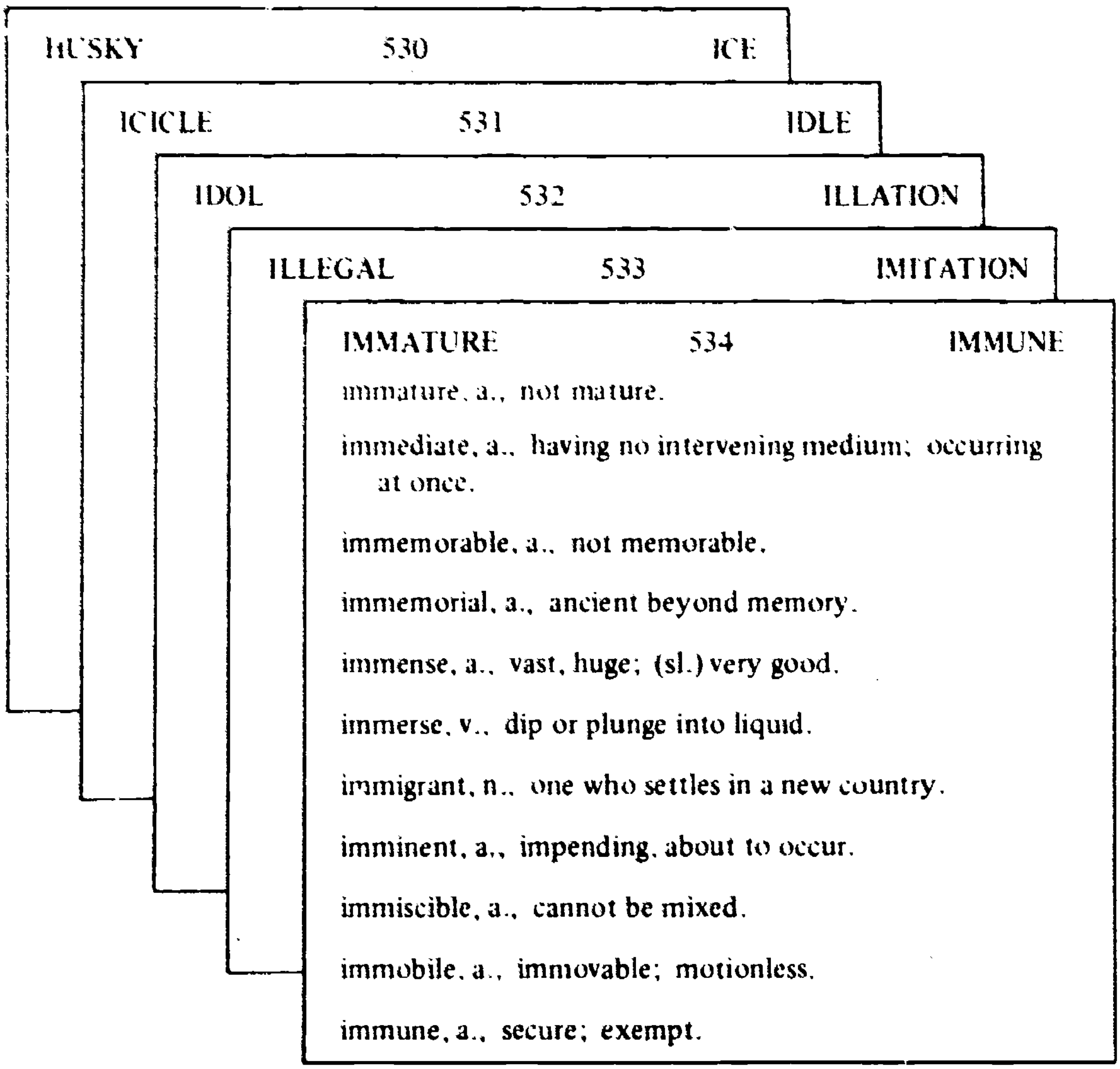
Whitfield, R.C. & Kerr, J.F. "Some Problems in Course Evaluation: A British View". *Teachers' College Record*, 72, 2, 267-274.

APPENDIX – SAMPLE TEST ITEMS

Skill 1 : Library Usage

Figure 1 below shows copies of several pages from a dictionary. Use these to answer Questions 3 – 6.

FIGURE 1



3. On which dictionary page in Figure 1 would you find the word "igloo"?

   A Page 530

   B Page 531

   C Page 532

   D Page 533

   E None of the above

TABLE OF CONTENTS

19. Information about Evonne Goolagong (the well-known tennis player) would be found in

A Chapter 1.

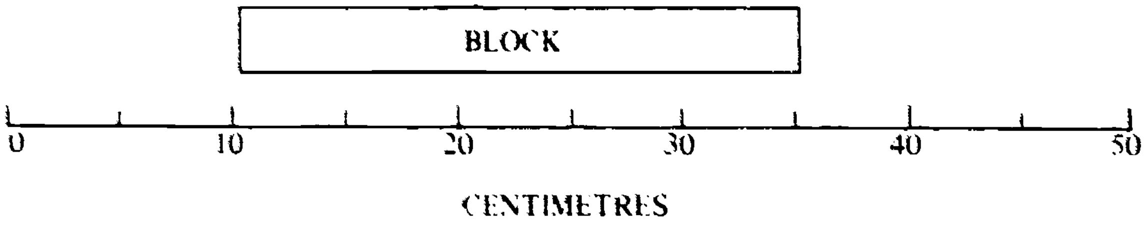
B Chapter 3.

C Chapter 4.

D Chapter 6.

E Chapter 7.

Skill 3 Scales



35. The length of the block of wood in Figure 13 is

A 25 centimetres.

B 20 centimetres.

C 15 centimetres.

D 30 centimetres.

E 35 centimetres.

Skill 4 · Averages, Percentages and Proportions

In a football match the winning team kicked a total of 20 goals. The names of the 4 goal-kickers together with the number of goals kicked by each person are shown below.

Name of Goal-kicker	Number of Goals Kicked
Jim	10
Bill	5
Alex	3
Jack	2
	20

Use this information to answer Questions 40 46 below.

40. The average (mean) number of goals kicked by Jim, Bill, Alex and Jack was

A 4

B 5

C 6

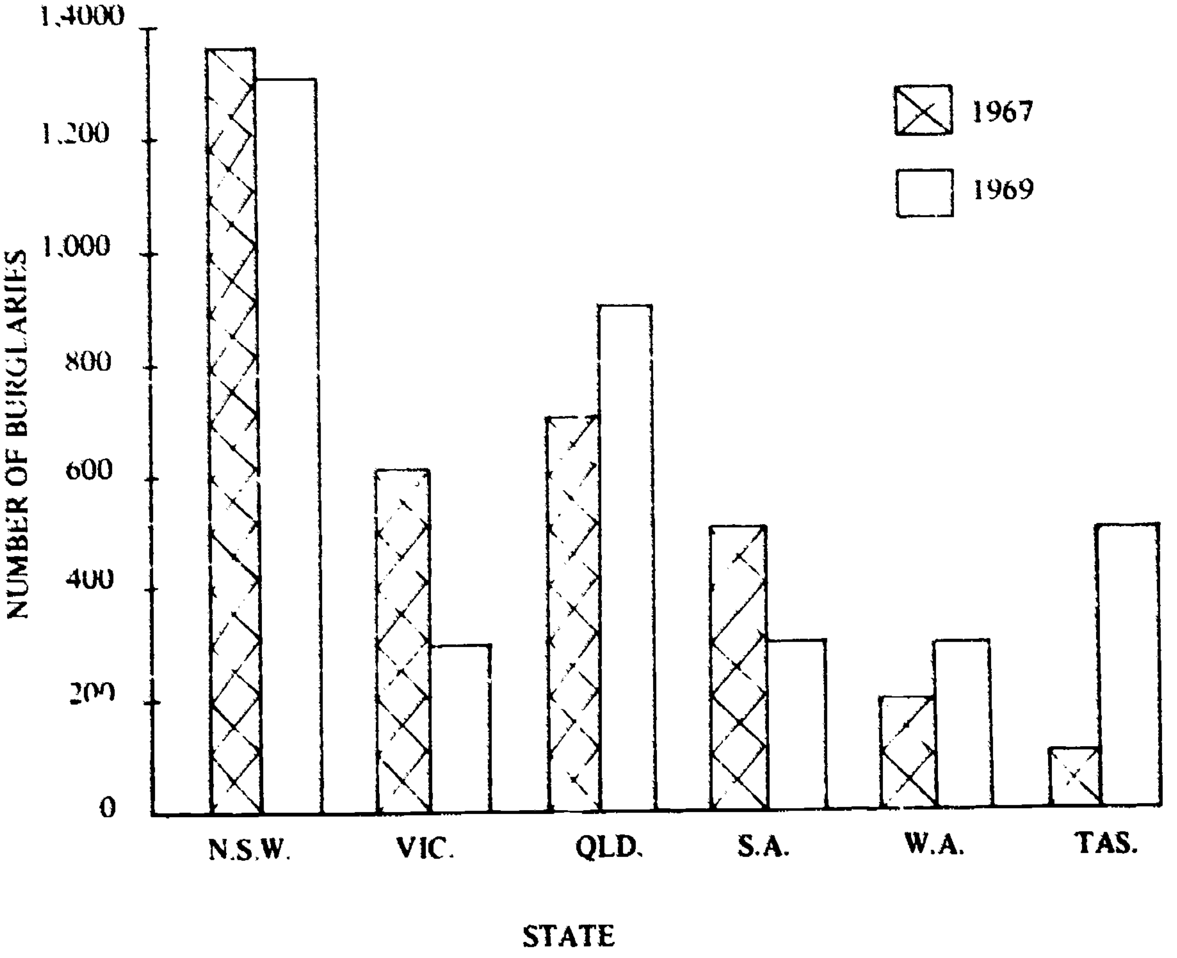
D 10

E 20

Skill 5 : Charts and Tables

CRIME IN AUSTRALIA

Statistics for house burglaries
for years 1967 and 1969

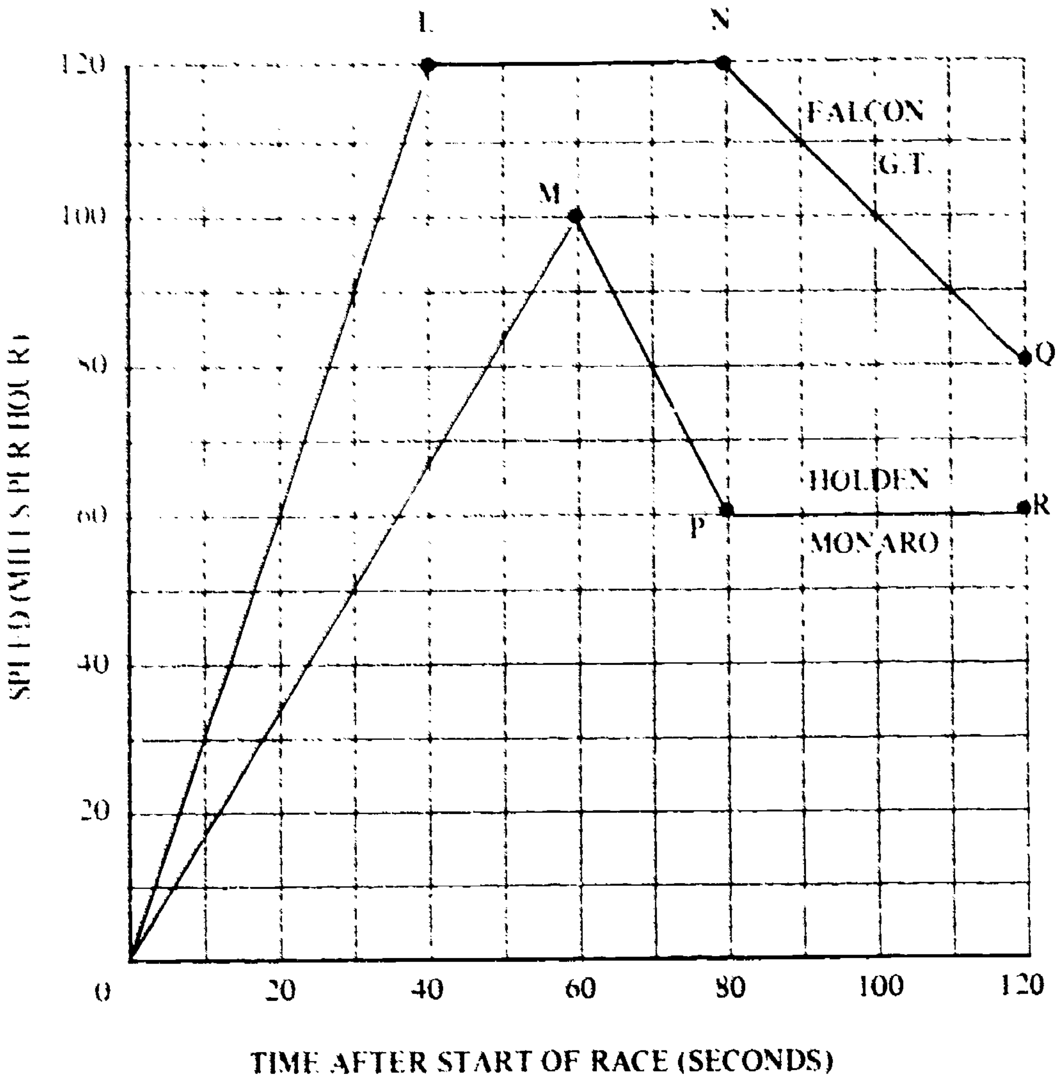


57. Which state had 500 burglaries in 1967?

A Tasmania
B South Australia
C Victoria
D Western Australia
E Queensland

Skill 6 Graphs

The graph below has been drawn from information obtained during a series of car tests at Sandown. It shows the speed of two cars at different times. Use this graph to answer Questions 68 – 75 below.



71. After 60 seconds, the speed of the Falcon G.T.

- **A** is greater than the speed of the Holden Monaro.
- **B** is less than the speed of the Holden Monaro.
- **C** is equal to the speed of the Holden Monaro.
- **D** cannot be predicted from the graph.

Skill 7 : Comprehension of Science Reading

Use the following paragraphs to answer Questions 86 – 88 below.

Par 1 Did you know that many bird songs are not for the enjoyment of man but a means of telling other birds to stay away? Before selecting a female for a mate the male bird stakes out a territory. He does this by fighting with, and chasing out, other birds from the chosen area.

Par 2 After a mate is selected a nest is built by the pair of birds and eggs are laid by the female. Usually it is the mother bird that keeps the eggs warm. She sits on the nest for many days while the eggs are incubating.

Par 3 During the first few weeks after the young birds hatch they eat a huge amount of food. The poor parent birds are kept so busy flying long distances to collect suitable food that they get very thin.

Par. 4 The young bird which hatches first gets most of the food, grows faster and may even push other young birds out of the nest.

Par. 5 While on the ground the fledglings are quickly eaten by enemies such as owls and cats.

Par 6 Some parent birds overcome the food problem by collecting and storing food for a few days before the eggs are due to hatch.

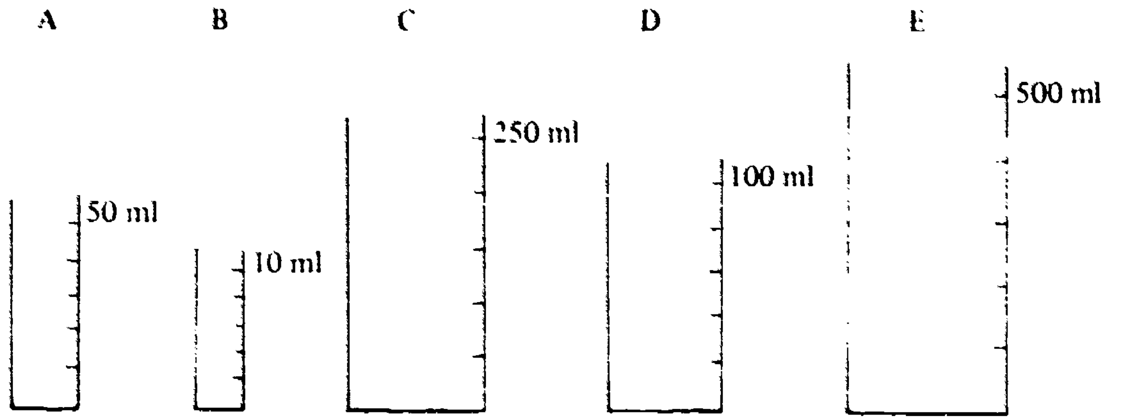
Par. 7 One such bird is the crested bellbird. These birds collect caterpillars, cripple them and place them on the branches of a tree near the nest.

87. The way a bird claims its living area is described in

A Paragraph 1.
B Paragraph 2.
C Paragraph 4.
D Paragraph 5.
E Paragraph 6.

Skill 8 Design of Experimental Procedures

101. Below are five containers which can be used for measuring amounts of liquid.



Which of the above containers would be the *most accurate* for you to use to measure 5 millilitres of water?

A Container A
B Container B
C Container C
D Container D
E Container E

Skill 9 : Conclusions and Generalizations

106. The builder says: "Some Zanians live in round houses. Those people over there are Zanians".

Which one of the statements below follows from what the builder said?

A Those people must live in round houses.
B Those people could not live in round houses.
C We don't know whether those people live in round houses or not.

FROM THE TWO CULTURES TO MONEY: TRENDS IN ENROLMENT IN HSC SUBJECTS IN VICTORIA

R.T. White

The phrase "flight from Science" has been current ever since the publication of the Dainton report (1968), which showed that in Britain the numbers of students in upper secondary school studying science was falling. Shortly after the report appeared there were several investigations of whether there was a similar trend in Australia (Mackay 1968; Thornton 1968, White 1969, Stranks 1969; Fensham 1970). The conclusions of these studies differed to a degree, because of the different methods employed to identify any fall in popularity of Science, but in general it was clear that if there were any trend at all away from Science then it was nowhere near as severe as in Britain. Because there may have been a change in the Australian situation since the first group of studies, this paper presents an updated report of enrolments for the Higher School Certificate (formerly Matriculation) examination in Victoria. No figures are presented for other States. Sheedy (1972) has a comparable, but not exactly parallel, set of data for New South Wales. Since trends in enrolment have important implications for administrators, curriculum constructors, teachers, and the public in general, it is hoped that this paper will stimulate the extraction of corresponding figures for the rest of the Commonwealth.

Conclusions about trends in enrolment are, to some extent, dependent on the index used to measure the trend. Several indices have been used in earlie studies. Their implications are discussed below.

School or university enrolments?

When one is trying to check whether there has been a decline in popularity of Science as a field of study, all of junior secondary, senior secondary, and undergraduate enrolments may seem appropriate for investigation. All suffer from disadvantages. At junior secondary level few students have a wide range of choice of their curriculum: the decisions are made for them by the school, overtly or covertly. Enrolments in subjects at junior secondary level are determined by administrators, not by students. Undergraduate enrolments are similarly directly determined by bureaucratic decision in Victoria, where for many years quotas have been applied to all university faculties. Thornton (1968) and Stranks (1969) included university enrolments in their discussions, but, as measures of popularity of subjects, enrolments and the concomitant figure, number of degrees awarded, are invalid.

Greatest freedom of choice exists in upper secondary school, though quotas have an indirect effect here. If students perceive that it is very difficult to enter a Science or Science-based faculty such as Medicine or Engineering, then many will prefer to aim at an easier target by studying non-Science subjects in their final school years.

None of junior secondary, senior secondary, or university enrolments is entirely satisfactory as a measure of trends in popularity of subject areas, where factors outside the intrinsic appeal of the subjects are to be excluded, but because the greatest element of free choice is in the senior secondary enrolments they are preferable to the others.

The first choice is to operate on senior secondary school enrolments, in this case the numbers enrolled for the Higher School Certificate examination in Victoria. The next choice is between absolute numbers, relative proportions, and proportions of the relevant population, choosing each subject.

Absolute numbers

Because the population is growing, absolute numbers studying each subject can be expected to increase. There is also a general trend for people to stay on longer at school, which also should increase the subject enrolments. Therefore, it is only when the popularity of a subject is in marked decline that the numbers enrolled for it will fall, and only then will absolute numbers provide a clear and useful index. This was the case for Science in Britain from 1964 on, and was the cause for the widespread concern generated by the Dainton report. Up to 1968, when the earlier Australian investigations were carried out, there was no such fall in absolute numbers in this country. The Victorian figures since then show that there are still increasing numbers enrolled for Science subjects. Table 1 (which is shown as a graph in figure 1) contains the comparative figures for England and Wales and Victoria. The British figures are for first year 6th form, and derived from information given on page 9 of the Dainton report. The sum of enrolments for Higher School Certificate Physics and Biology (V.U.S.E.B. 1964-1974) is used in the Victorian figures. This would seem a reasonable index, for few students take both subjects; their total should give a fairly consistent relative measure of the popularity of science subjects so long as other factors remain constant. It would not, for instance, be a sensible measure if a new science subject were introduced. No changes likely to affect the use of this index appear to have occurred in the decade covered by Table 1.

TABLE 1

Numbers of students in science courses in 6th form, relative to 1962 enrolment

	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972
England & Wales	1.000	1.100	1.225	1.174	1.148	1.117	1.059	-	-	-	-
Victoria	1.000	1.112	1.373	1.520	1.620	1.687	1.916	2.300	2.613	2.721	2.783

Since the absolute numbers enrolled for Science subjects in Victoria are increasing at the same time as the population is increasing, little more can be deduced from Table 1 than that there is not so severe a flight from Science in Victoria as there is in Britain, if there is any flight at all. An index independent of the growth in population is needed before more information can be obtained from the enrolment figures.

Relative proportions

The relative popularities of subjects are independent of growth in population.

BEST COPY AVAILABLE

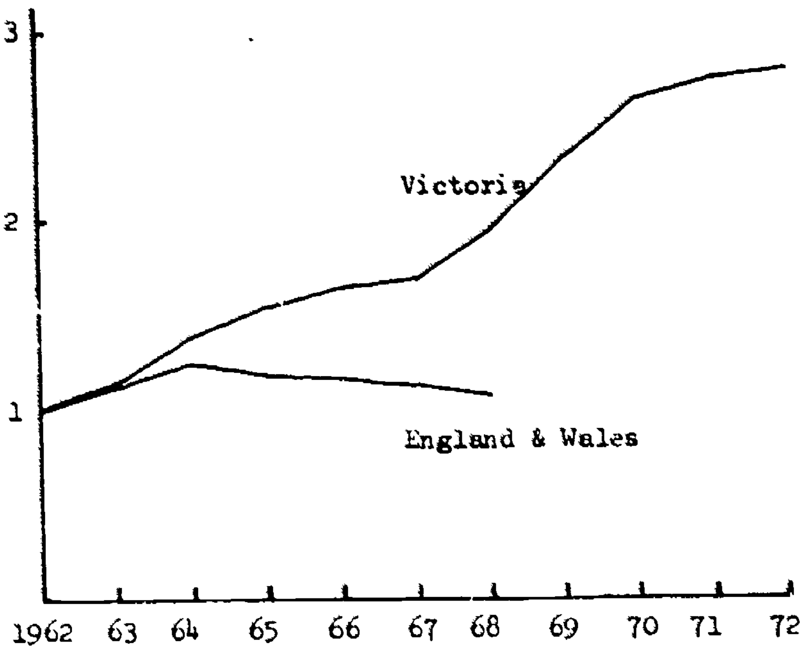


Fig. 1. Numbers of students in science courses in sixth form, relative to 1962 numbers.

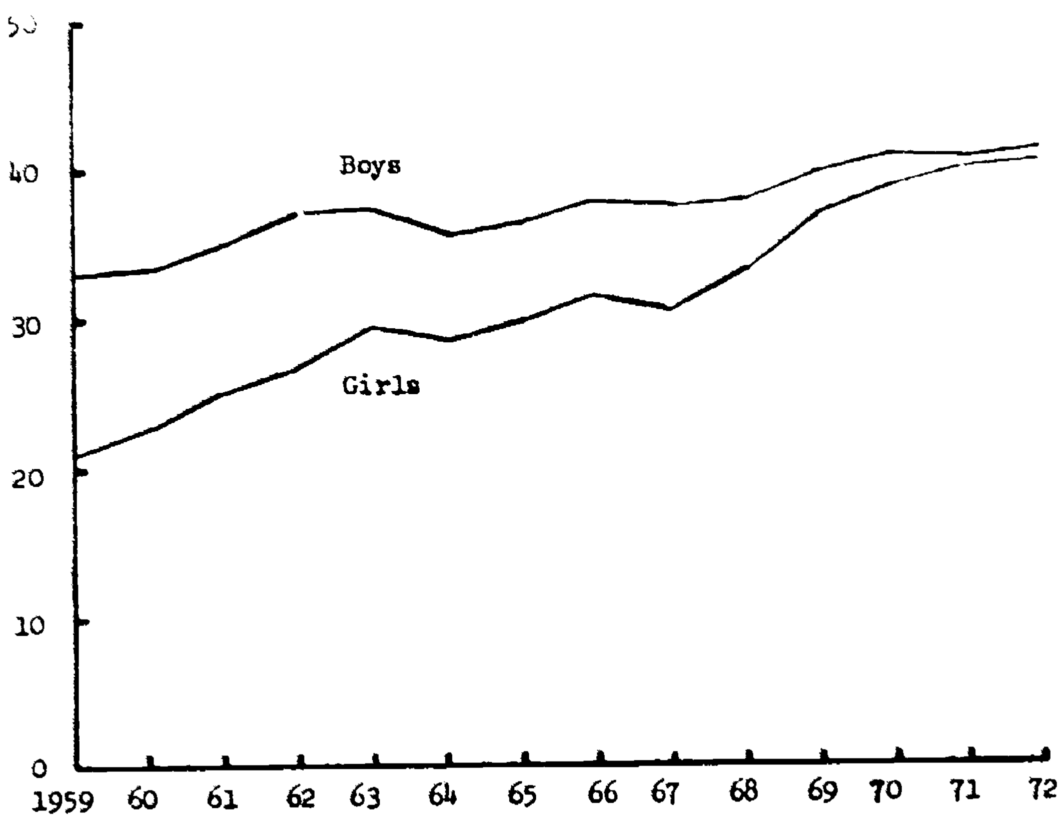


Fig. 2. Victorian sixth form enrolments as a percentage of fourth form enrolments two years earlier.

Mackay (1968) produced graphs of the fraction, for each subject, of the number enrolled for the subject divided by the number of students enrolled for the whole Matriculation examination. The weakness of relative proportions is that they do not take into account shifts in the type of population entering sixth form. Where the shift can be identified and measured, allowance can be made. For example, there has been a shift in proportion of males and females. Table 2 (which is shown as a graph in figure 2) shows the comparatively great increase in proportion of girls staying at school. The figures were derived from tables supplied by the Commonwealth Bureau of Census and Statistics (1960-1973). Different preferences of the sexes for different subjects may cause this shift to produce a fall in relative popularity of some subjects, such as Physics, that have never been popular with girls, and a rise for others, such as Biology. Because of this, Mackay presented separate results for the two sexes.

TABLE 2

Victorian 6th form enrolments as a percentage of 4th form enrolments two years earlier

	1959	1960	19[illegible]	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972
Boys	33.0	33.3	35.1	37.2	37.3	35.5	36.4	37.6	37.3	37.8	39.6	40.6	40.3	41.2
Girls	21.1	22.6	24.7	26.2	29.3	28.5	29.9	31.2	30.2	33.2	36.9	38.5	39.7	40.4

Mackay's results showed that between 1959 and 1967, for both males and females, there was a decline in the relative proportions studying physical science and French, and a rise for Biology and Geography. These trends could, however, be due to shifts in the type of population other than in proportions of the sexes. For instance, as the proportion in the population of sixteen and seventeen year olds remaining to sixth form increases, it can be expected that the fraction of less able students will icnrease (apart from other changes such as the proportions of people from different economic backgrounds). It has been shown that in the past the Sciences attracted the more able students (A.C.E.R. 1964), so it could be expected that as more people seek education a lower proportion of those remaining at school will choose the present science subjects.

Changes such as general ability in the sixth-form population are not easily allowed for when using relative proportions, but become irrelevant when the proportions of the maximum possible sixth form population are calculated and compared for each subject.

Proportions of the relevant population

Calculation of this index is made difficult by the lack of accurate figures for the number of people who could have been in sixth form in any one year had they all stayed on at school. Total numbers in, say, the seventeen year old age group are not satisfactory because the relation between this figure and the possible sixth form population, which would include a spread of ages, is unknown, and may vary with time because of changed admission policies in primary schools long before. Perhaps the most reasonable estimate based on readily-available figures is the total enrolment two years earlier in form four. Use of this figure to estimate the maximum possible sixth-form population assumes that the proportion staying on to fourth form remains reasonably constant from year to year. It would be safer to use third form figures three years earlier, or first form figures five years earlier, but these do not seem to be available. Use of the fourth form figure also assumes that the numbers of mature-age, recent-immigrant, and repeating candidates are either negligible or are in proportion to the total enrolment, which can probably be accepted as reasonably likely. One further assumption is involved in the figures presented below in Table 3. In considering the proportion who go on to Higher School Certificate in Victoria, it would be preferable if enrolments in forms four in high schools and private schools only were used, since very few technical school students enter for this examination. Possibly separate totals for high, private, and technical schools exist, but only the combined total was available at the time of writing. Its use in the construction of Table 3 (which is shown as a graph in figure 3) assumes a relatively constant ratio between technical school and total enrolments in form four from one year to the next.

In constructing Table 3, the numbers enrolled in the subjects were taken from the Victorian Universities and Schools Examinations Board Handbooks (1958-1974), and the numbers in fourth form were taken from the Schools publications of the Commonwealth Bureau of Census and Statistics (1957-1973).

On the 1956 to 1968 figures of Table 3, White (1969) argued that for boys there was no evidence of a flight from science. Until then the Physics and Biology figures showed an effectively constant attraction by these subjects. Since then a change has occurred. In 1972 Physics suffered a serious decline; so great was this decline that the absolute numbers choosing Physics fell while the population rose. Biology, in contrast, showed a marked rise. The same features can be seen in the graphs for girls, though the downturn for Physics is not so severe, and the rise in Biology, which has been going on since 1960, is even more spectacular than for boys. White also argued that for boys Literature and French had suffered a serious decline. The decline has been arrested for Literature, while French has faded away to a very minor subject with prospects of following Latin into oblivion. For girls the same serious decline can be seen for French; Literature, except in the last year, has continued to grow and remains the most popular subject other than English. The rise in Geography, which was apparent up to 1968, now seems to be reversing. What White (1969) called the most striking feature of enrolments up to 1968, the surge in Economics and the subjects likely to be associated with it, Social Studies and Accounting, is even more marked. Physics, which was for a long time the most popular subject for boys, has been passed by Economics. The total enrolment for boys and girls in Social Studies leapt by 40% in 1972. Accounting has risen from an insignificant subject to pass Literature, and, if present trends are maintained, will soon be more popular than Geography for boys.

These figures show that there has not been a flight from science in Victoria, but there has been a marked change in enrolment within the sciences from physical science (the figures for Chemistry are closely parallel with Physics) to biological science; that there has not been a swing to the humanities; but there has been a marked flight to commence.

It is not the purpose of this paper to identify causes of these trends. However, it may be useful to suggest some possible causes, which are open for future researchers to investigate.

TABLE 3

Entries for Higher School Certificate Subjects as a percentage of enrolments in fourth form two years earlier.

Upper figure boys, lower figure girls

	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972
Accounting	1.7	1.7	2.1	1.8	1.9	1.9	2.6	2.8	2.8	3.7	4.4	4.7	5.0	5.7	6.3	7.1	8.4
	0.2	0.2	0.3	0.3	0.4	0.4	0.5	0.6	0.6	0.9	1.0	1.2	1.3	1.6	1.9	2.2	2.8
Biology	3.4	4.3	3.7	3.0	3.2	3.4	4.3	3.8	4.0	4.1	4.3	3.9	4.6	6.4	7.5	8.3	9.0
	5.8	6.3	6.4	6.1	6.1	7.4	7.5	8.5	8.9	10.3	12.6	12.2	13.2	14.9	18.3	20.0	20.8
Economics	5.6	6.7	6.9	7.1	7.5	9.2	11.5	12.0	11.9	13.1	14.4	15.4	16.1	17.7	19.1	20.0	19.9
	1.1	1.1	0.9	1.4	1.5	2.2	2.3	3.4	3.6	4.4	4.9	5.2	6.4	7.5	9.5	9.9	10.5
French	5.7	5.7	5.4	5.1	5.0	4.7	5.4	4.7	4.4	4.3	4.1	3.5	3.6	3.1	2.4	1.9	1.4
	8.9	8.8	8.5	8.8	9.4	10.1	9.9	11.4	11.1	11.6	11.7	11.3	11.0	10.3	9.7	8.4	7.5
Geography	9.5	10.6	10.3	10.6	11.0	11.9	14.3	14.8	12.9	13.7	13.8	13.6	13.6	14.2	14.5	14.7	13.3
	6.1	7.7	6.9	8.7	9.1	10.7	11.2	13.8	13.5	17.0	15.1	14.6	15.5	15.9	17.8	17.5	16.6
Literature	10.8	10.5	9.5	7.8	8.3	9.0	10.2	8.9	7.3	7.2	8.1	7.5	7.8	8.0	8.1	8.1	8.0
	13.4	13.8	12.2	12.8	13.3	15.5	14.5	17.2	16.3	18.1	18.5	17.9	18.8	19.4	21.8	22.0	21.4
Physics	19.8	22.4	21.2	21.7	20.5	20.3	21.6	20.6	19.4	20.3	19.4	18.8	19.2	20.0	19.7	19.3	17.3
	3.6	3.5	3.9	4.0	4.7	4.9	4.3	5.1	5.2	5.5	5.2	5.0	5.6	5.4	6.1	5.7	5.3
Social Studies	4.2	4.7	4.8	4.5	5.0	5.0	6.3	6.3	6.1	6.9	6.6	6.7	7.2	7.5	8.6	10.7	12.4
	0.5	0.9	0.8	1.4	1.1	1.6	1.6	2.0	2.1	2.4	2.9	2.8	3.6	4.3	5.0	7.3	8.8

BEST COPY AVAILABLE

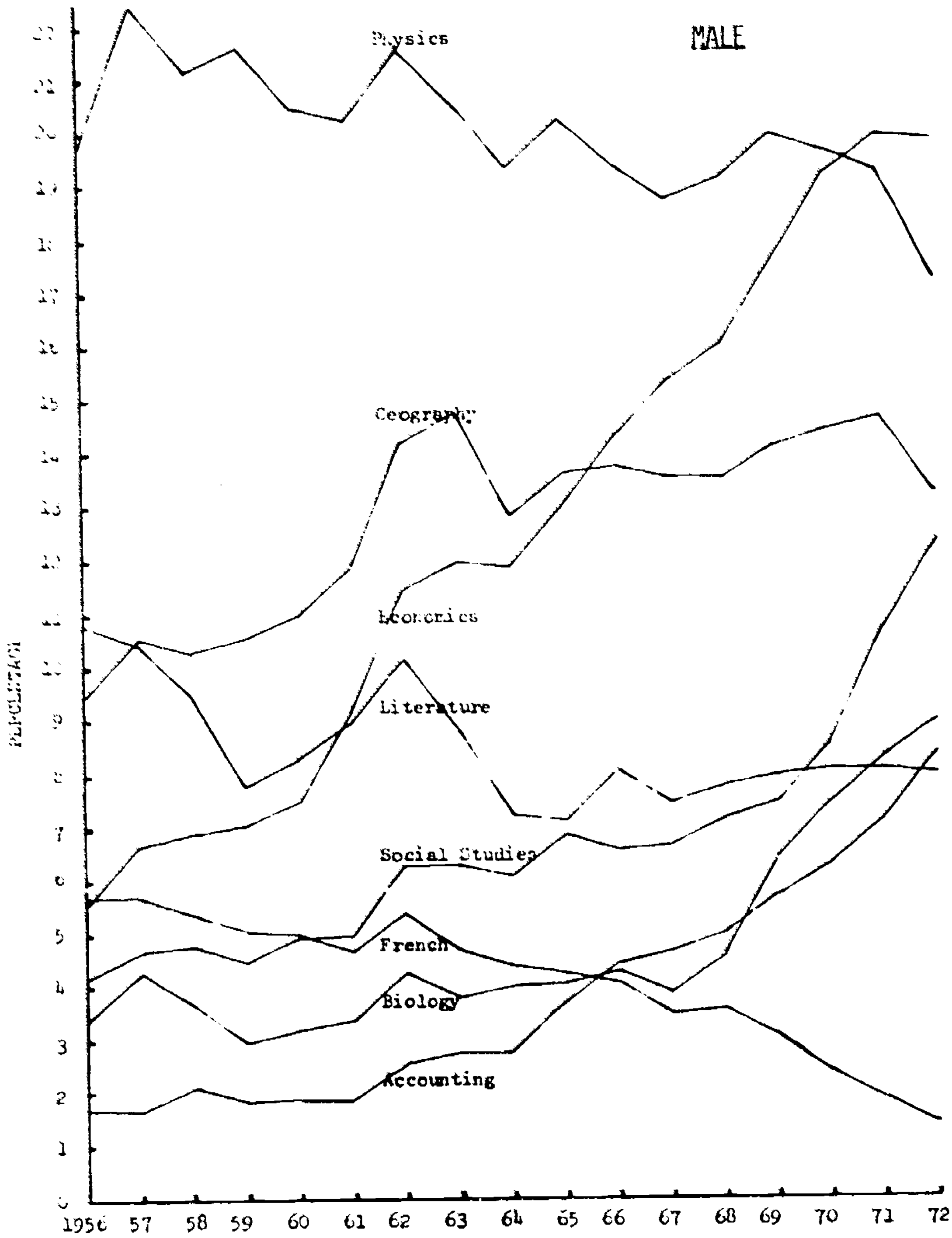


Fig. 3. Entries for HSC subjects as a percentage of enrolments in fourth form two years earlier.

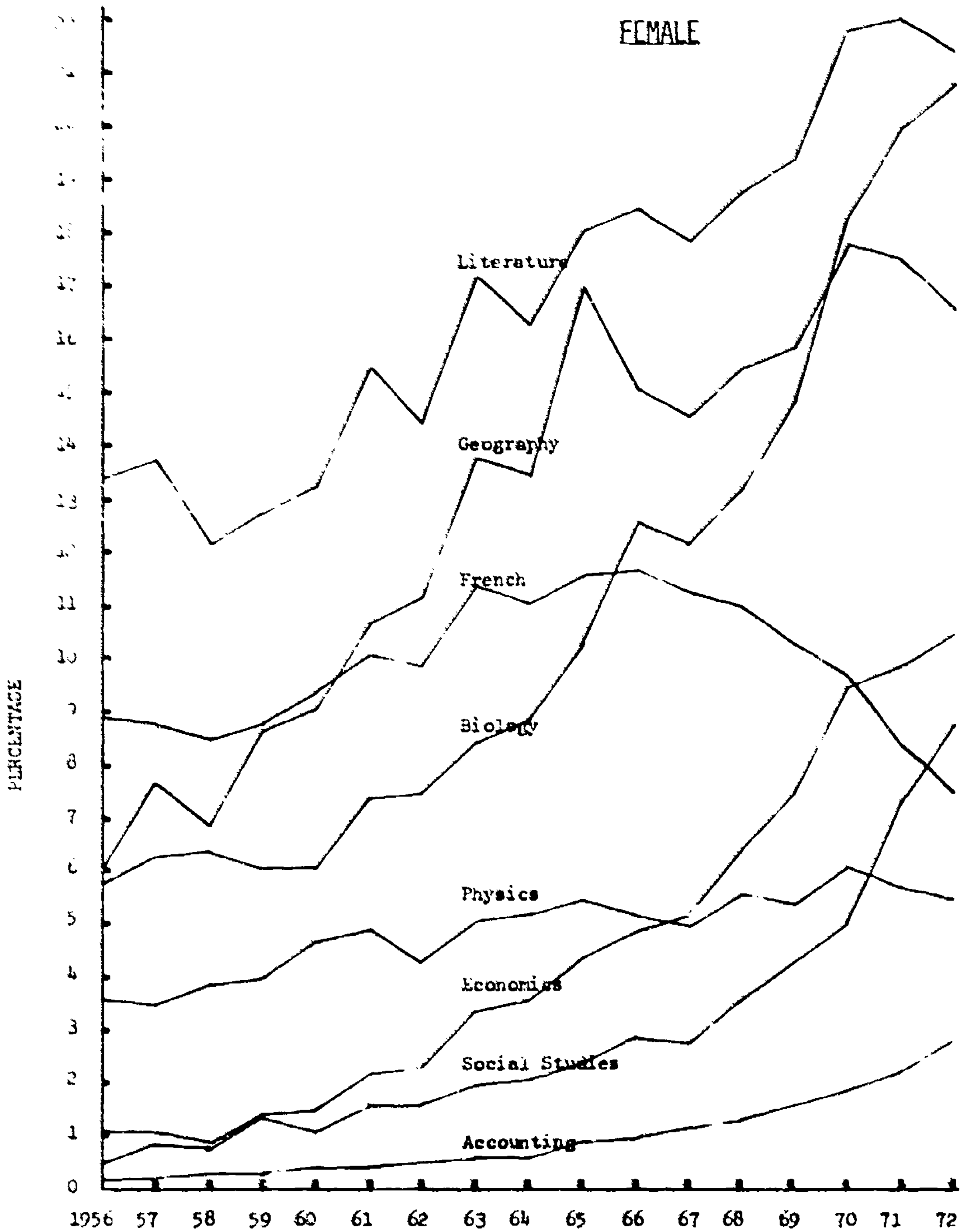


Figure 3 continued

1. With more children staying on at schools, and thus larger sixth forms, schools are more able to establish classes in subjects such as Biology, which were previously uneconomic to run

2. More teachers of Biology, Economics, and Accounting have appeared in schools.

3. The general trend in the community away from authoritarianism has caused a lack of sympathy with subjects such as Physics and French that are perceived as rigorous

4. Choice of subjects is largely determined by perceptions of how difficult the subjects are, and there is a growing tendency to seek the easiest subjects.

5. Subject choice is determined by employment plans, and these have changed because of publicity given to failure of physical science graduates to find suitable positions.

6. Subject choice is determined by experiences in junior forms, and the loss of popularity of Physics (and Chemistry and the mathematics subjects) is attributable to the persistent shortage of qualified and experienced teachers of these subjects.

It must be stressed again that these are hypotheses only, and do not necessarily reflect the opinion of the writer.

The trends that have been identified, and their possible causes (whatever they may be), pose a number of questions for the community. Answers to the questions will be a matter of opinion, but it is preferable that opinions are formed after consideration, and that a consensus should be reached if possible. It should at least be discussed whether the trends are in the community's best interests, rather than to allow changes to occur unconsidered and unremarked. Although it is not intended to give or imply answers to any of the questions that may be of interest, it may be useful to list some possible questions below.

1. Is it desirable for growth to occur in neither of the "two cultures" while rapid growth occurs in subjects outside their scope?

2. Is it satisfactory for about one boy in six and one girl in twenty to be studying Physics at 6th form level?

3. Should generalist subjects like Social Studies be encouraged? Should there be a scientific counterpart?

4. Are enough students studying Literature?

5. Is it desirable that hardly any boys and very few girls are studying a foreign language at sixth form?

6. What changes should tertiary institutions make in their practices in consequence of these trends?

REFERENCES

A.C.E.R. "Survey of general ability of pupils in forms 4, 5 and 6 of Victorian secondary schools 1962", Report No. 1, August 1964.

Commonwealth Bureau of Census and Statistics. *Schools*, 1957-1973.

Dainton Committee. "Enquiry into the flow of candidates in science and technology into higher education — final report". H.M.S.O. Cmnd. 3541, 1968.

Fensham, P.J. "The place of science in Australian education". *Search*, 70, 28-33.

Mackay, L.D. "Enrolments for Physics in Victoria 1959-1967". Unpublished, Faculty of Education, Monash University, 1968.

Sneedy, I.J. "Aspects of the 'swing from science' in New South Wales". Unpublished M.Ed. thesis, University of Newcastle, 1972.

Stranks, D.R. "Drifts from the Sciences". Stanhope Oration, Conasta XVIII, Adelaide 1969, insert in South Australian Science Teachers' Association Journal, December 1969.

Thornton, J.B. "The drift from Science to the Arts". *Australian Journal of Science, 31*, 1968, 206-209.

V.U.S.E.B. *Handbook of Directions and Prescriptions*, 1958-1974, Melbourne.

White, R.T. "The flight to commerce". Unpublished, Faculty of Education, Monash University, 1969.

PART III

ANALYSES OF CURRICULA

CONTENT ANALYSIS AS A TECHNIQUE FOR DETERMINING THE STRUCTURE OF TEXT MATERIALS

Brian W. Carss

Introduction

The origin of this development stems from a project which was funded by The U.S. Office of Education in 1968 called Cognitive Memory. This was an interdisciplinary project based in the Coordinated Science Laboratory of the University of Illinois. The main aims of this project were to develop the software for the retrieval of information from a large, generalised data bank. This was not to be retrieval in the normal sense of the word however, where the entire data bank was to be searched in an exhaustive manner in order to find the pieces of information that were required. Rather, an attempt was made to develop a search strategy which would simulate in some way, the way in which human beings retrieve information from their own memories. Obviously the speed with which a human can 'remember' a given fact (in about 0.1 second) precludes any possibility of carrying out a complete search of all of his 10^9 neurons. Individual facts after having been assimilated by the brain are linked into associative structures and signposted in such a way that retrieval of a fact, also allows retrieval of a mass of associated information. The Cognitive Memory project made an attempt to understand how to store information in a data bank so that it was automatically signposted.

Although a considerable amount of research effort is being devoted to this very problem by a number of research teams, little real headway has been made because of the difficulty in relating written symbols to meaning without ambiguity. Results are encouraging, but an operating system is still a long way off.

My particular area of interest in the Cognitive Memory project had been in the use of word frequencies as a means of Content Analysis, and it was this background that was used as a starting point for this project.

The basic assumption upon which this analysis is based is that there exists sufficient meaning within selected words of a piece of text, and their relative distribution and location to determine major themes. In other words, it is necessary to generate the informational categories with which many content analysis systems begin.

Each document is divided into a number of paragraphs and a word frequency count of a selected group of words is recorded. A paragraph was selected as the informational quantum for analysis purposes because an idea or concept is usually contained within one or several paragraphs but not in anything less than a whole paragraph. Using these frequency data, intercorrelations among all words are obtained. These intercorrelations represent the degree of co-occurrence or association between words as they occur in the various paragraphs of the text materials.

The matrix of intercorrelations is subjected to a principal components analysis to determine, in a systematic fashion, those components which can account for the intercorrelation matrix in some meaningful way. By definition, the number of principal components will be very many fewer than the number of words in the data base. However, the penalties that must be paid for this reduction in the amount of material to be manipulated, are related to machine-size, program running time and factor interpretability. It is the purpose of this paper to describe how this can be achieved.

Description of the System

As in most natural language studies, using a computer demands that the text materials must be in a machine-readable form. This translation of written text into a machine readable form is a very expensive operation, so that any savings that can be achieved must be exploited to the full. In order to overcome unnecessary expense and to increase accuracy, it was found desirable to eliminate all manual coding where possible. In this study, the complete text was transcribed directly onto punched cards, which after proofreading and correcting, served as the raw data for input to the first stage of the analysis.

The first analytical step in the analysis involves determining word frequencies for the whole text, and then on a paragraph by paragraph basis. Certain classes of words are deleted automatically because they occur so frequently or so ubiquitously, that their information content is small. These words all belong to one of the following groups, conjunctions, prepositions, pronouns and the definite and indefinite articles.

Due to the limitations of the size of available computer memory and machine speed, certain other words have to be deleted from the other end of the frequency spectrum. This usually means that all of the words that just occur once are deleted. Deletion of words that occur once appears to be reasonable, since we are particularly interested in the relationships that exist between paragraphs, and these singleton words can not contribute anything to the knowledge of inter-paragraph relationships. Successive deletion of the least frequently occurring words is continued until a corpus of words that will fit into the available computer memory is obtained.

The second pass through the preprocessing program determines the frequency of the selected words in each paragraph and this word frequency/paragraph matrix forms the basis of the input to the second stage of the analysis which is a principal components analysis.

Each principal component that is extracted consists of a number of words that are most highly correlated and taken together define a *Thematic Element.* (Table 1) The principal components are rotated to simple structure using a Varimax rotation in order to make the thematic elements more easily interpreted. The final step in this stage of the analysis is to calculate the factor scores for each of the paragraphs. The factor scores represent the 'amount' of each thematic element in each paragraph.

The matrix of factor scores is printed and stored on a disk file at the same time. The factor score matrix is the input data for the final step in the analysis, a correlation analysis. A diagrammatic representation of the steps in the analysis are shown in Figure 1.

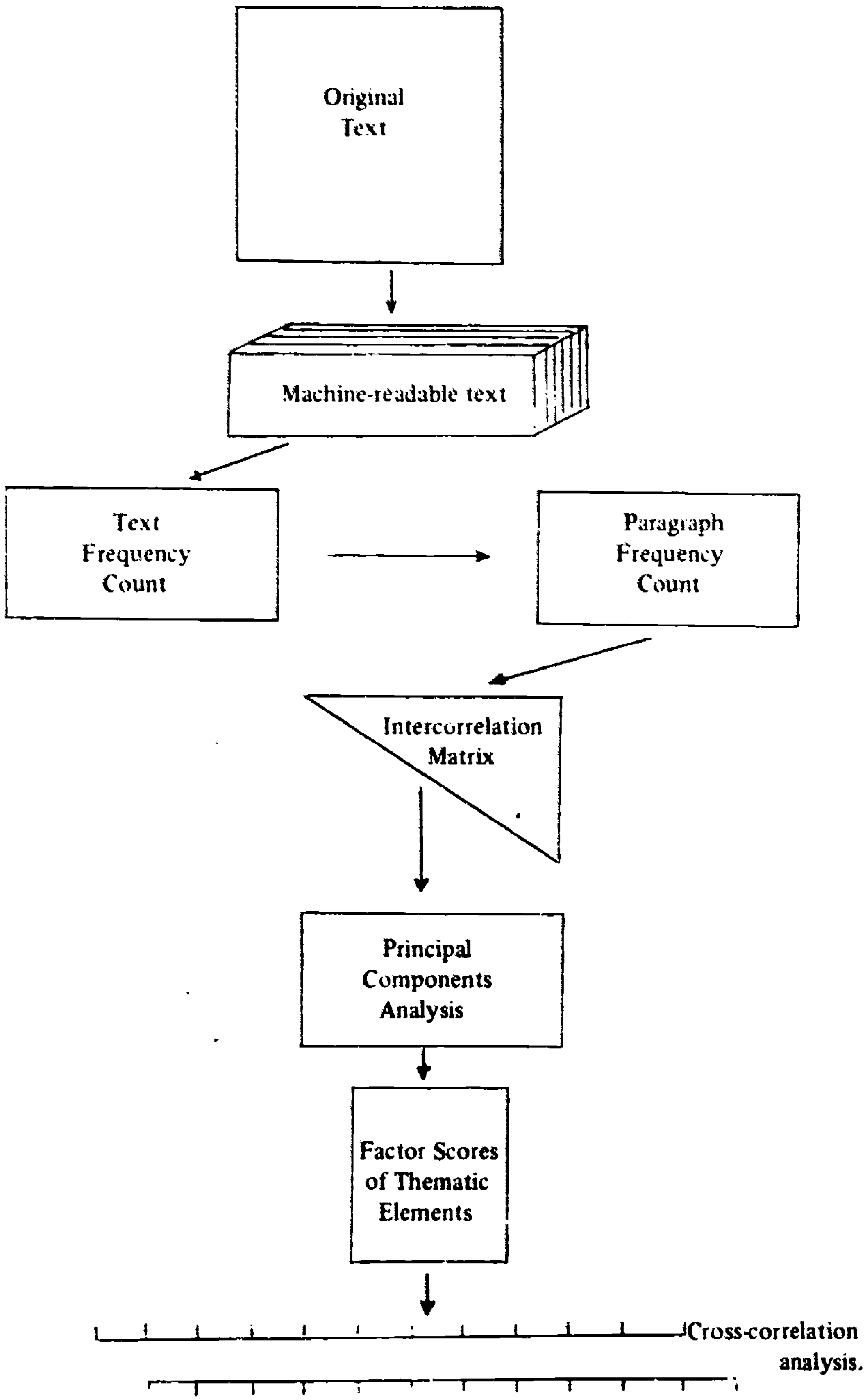
TABLE 1

Educational Psychology : A Cognitive View

Thematic Elements

1.	Ability	80
	Personality	78
	Factory	78
	Intellect	62
	Motivation	51
	Variable	50
	Situation	49
	Social class	40
13.	Instruction	83
	Condition	47
	Variable	41
8.	Learning theory	80
	Teaching theory	79
	Educ. Psychol.	50
2.	Meaningful learning	84
	Reception learning	81
	Task	79
	Rote	77
4.	Problems	78
	Applied	76
	Research	71
	Solution	57
	Relevance	50
	Science	48
3.	Concern	78
	School	77
	Academic	71
	Student	66
	Intellect	47
	Curriculum	40
5.	Commonsense	80
	Psychology	77
	Theory	61
	Principles	49
10.	Verbal	82
	Concept	82
	Process	53
	Empirical	51
6.	Function	75
	Subject matter	39
	Discovery learning	39
9.	Pupils	78
	Teaching	69
	Motivation	41

FIGURE 1


Original
Text
Machine-readable text
Text
Frequency
Count
Paragraph
Frequency
Count
Intercorrelation
Matrix
Principal
Components
Analysis
Factor Scores
of Thematic
Elements
Cross-correlation
analysis.

In the cross-correlation analysis, the factor scores for each thematic element, when plotted on a paragraph basis, can be regarded as a time series. Each factor score trace is cross-correlated with every other trace with lags of 0 and 1.

A lag of 0 means that the traces are set side by side without any lateral shift in the paragraphs. In this case if the correlation was high, then it would be interpreted that for most of the text under consideration these two thematic elements are present together. With a lag of 1, there is a lateral shift of one paragraph for the two thematic element traces. A high correlation in this case, would be interpreted that one theme is followed by another throughout most of the text.

During the cross-correlation phase of the analysis all pairs of thematic element traces are cross-correlated with each other and the results printed. Only correlations which are significant at the 0.05 significance level are considered for further interpretation.

The interpretation phase has two parts. The first, the actual structure of the written text is plotted using the factor scores data and the information gained from the cross-correlation. The decision as to whether a particular thematic element is present in a paragraph depends on whether its factor score exceeds a critical threshold value or not. At this stage in the development of the technique we do not have any definite guidelines as to what is an appropriate threshold value. However, we have applied a cutoff value of + or −1, and this seems to work reasonably well for the text that we have analyzed so far. The cross-correlation indicates which thematic elements can be joined because of the significant correlation between them. The resulting diagram is a pictorial representation of the *text structure*.

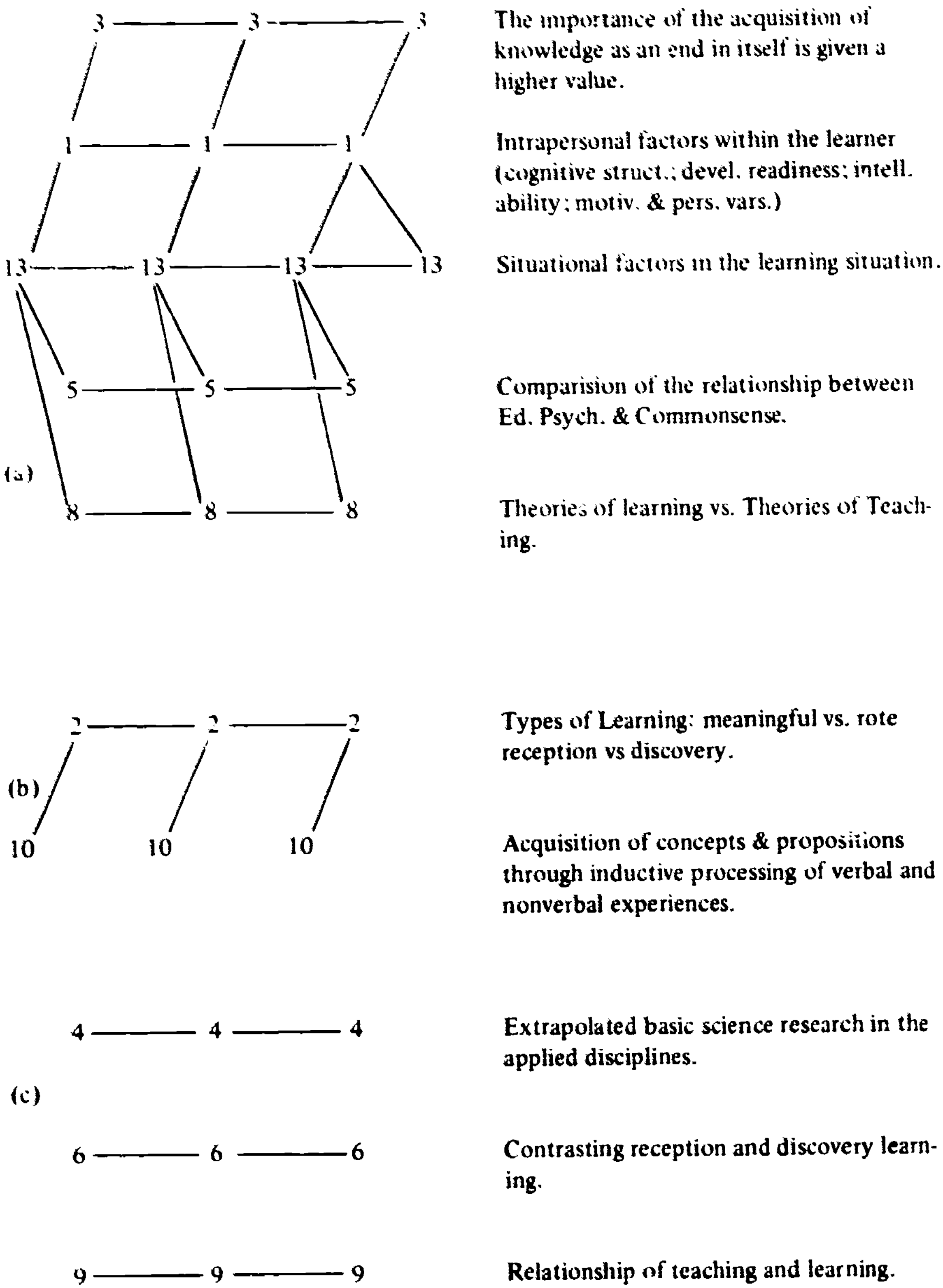
The second part, can be regarded as the *idealized structure* in that a plot is constructed from those highly correlated thematic traces which could theoretically follow each other. In this plot no temporal sequence of paragraphs is implied. By tracing the various pathways through this plot, one can trace the development of a series of concepts. (Figures 2, 3).

Illustrative Examples

The first example is the first chapter of D.P. Ausubel's book Educational Psychology : A Cognitive View, Holt, Rinehart & Winston 1968. The chapter itself is called 'The role and scope of educational psychology'. It was divided into 96 paragraphs, and the analysis carried out as described earlier. Twenty four thematic elements were extracted accounting for 78.5% of the variance and of these 24, only 15 were interpreted. The cross-correlation analysis was carried only on these 15 thematic elements.

The partial idealized structural pattern is shown in Figure 2. Only a partial idealized structure is presented here in order to avoid confusion. What is presented, is enough to illustrate the technique. Figure 2(c) shows three *sequential* thematic elements, where sequential is defined as the continued occurrence of the same thematic element. It would appear that in this type of writing, the same concept is being developed within itself and is not being modified by reference to any external theme.

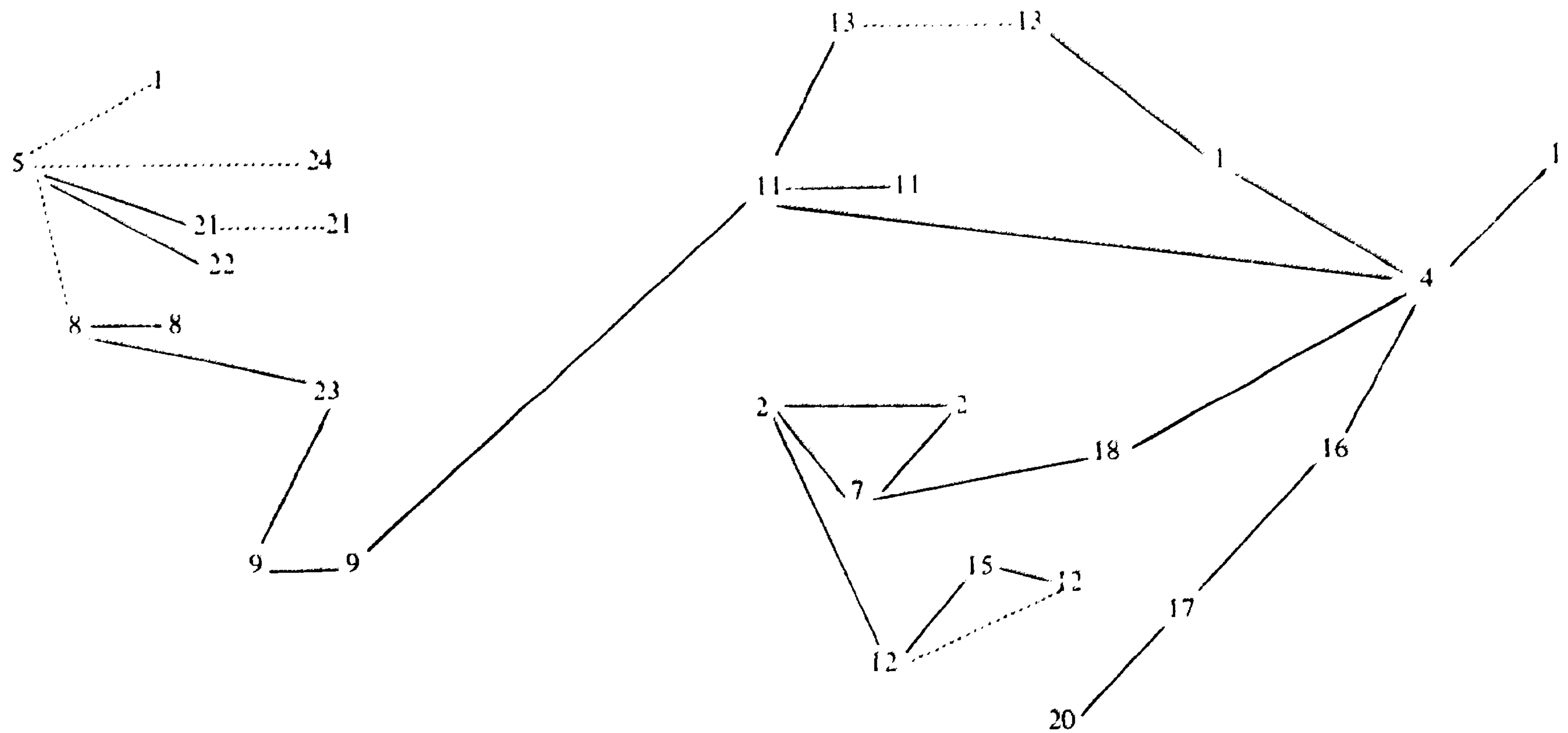
FIGURE 2



3 3 3
1 1 1
13 13 13 13
5 5 5
(a)
8 8 8
The importance of the acquisition of knowledge as an end in itself is given a higher value.
Intrapersonal factors within the learner (cognitive struct.; devel. readiness; intell. ability; motiv. & pers. vars.)
Situational factors in the learning situation.
Comparision of the relationship between Ed. Psych. & Commonsense.
Theories of learning vs. Theories of Teaching.
2 2 2
(b)
10 10 10
Types of Learning: meaningful vs. rote reception vs discovery.
Acquisition of concepts & propositions through inductive processing of verbal and nonverbal experiences.
4 4 4
(c)
6 6 6
9 9 9
Extrapolated basic science research in the applied disciplines.
Contrasting reception and discovery learning.
Relationship of teaching and learning.

FIGURE 3

'Laws & Theories in Geography' D. Harvey (Chapter 9 structure)



In Figure 2(b) by comparison, thematic element 2, develops a comparison between meaningful learning and rote learning and reception versus discovery learning. This theme is progressively modified by another thematic element, theme 10, which discusses the acquisition of concepts and propositions through verbal and nonverbal experiences. In other words theme 10 is being used to supply examples for the development of theme 2.

Figure 2(a) is by far the most complex set of thematic inter-relationships. Four thematic elements 3, 1, 5, and 8 are all sequential in character but with very different kinds of inter-relationship patterns. For example, theme 1 is being progressively modified by theme 13 in exactly the same way that theme 10 modified theme 2 in Figure 2(b). However, theme 1 in turn modifies theme 3 which from its interpretation would appear to be at a higher level of abstraction. Theme 13 also modifies themes 5 and 8 and also serves as a linking theme between 1, 3 and 5 and 8.

The second example is Chapter 9 of a geography book which discusses the structure and underlying concepts of geography as a discipline. The book by D. Harvey is called 'Laws and Theories in Geography'. It can be seen from the idealised structural diagram (Figure 3) that the structure is a much more complex one, than we have just seen for Ausubel. Twenty four thematic elements were extracted, not all of which could be interpreted meaningfully. The whole of the idealised structure has not been presented here because of its complexity. The partial diagram (Figure 3) that is shown here, however, shows several important features. For example, one can see on the far left side of the diagram how thematic element 5 acts as an initiator, and is linked directly to five other thematic elements. Thematic component 5 is actually a discussion of the basic postulates for theory in geography and so in that sense, it serves as an advance organizer. In contrast to thematic element 5, theme 4, on the right hand side of the diagram serves as a summary. The interpretation of the thematic component shows it to develop a general theory in geography which explores the links between spatial form and temporal processes.

In this diagram too, one can trace the development of a particular set of ideas as shown by the pathway through themes 2, 7, 18, 4 and 1. (Table 2). Theme 2 is concerned with the derivation of laws in geography; theme 7 discusses how geographers use very specific examples to illustrate a point; 18 would appear to be some padding and concerned with interaction; 4 explores the links between spatial form and temporal processes and finally theme 1 deals with how the ergodic hypothesis may be used to link temporal behaviour with spatial form.

An even better example of the progressive development of a set of concepts would be the pathway defined by the themes 20, 17, 16, 4, 1 (Table 2). Theme 20 describes how human geographers make use of the methods derived from social scientists in attempting to explain geographic phenomena.

Theme 17 gives definitions of geography which give rise to a set of questions pursued in research in geography; 16 states that geographers must be aware of the empirical status of the laws derived from other disciplines; and we are back to them 4 again.

TABLE 2

HARVEY THEMATIC COMPONENTS

1. The ergodic hypothesis may be used to link temporal behaviour with spatial form.

2. Methodological universality and the derivation of laws in geography.

4. General theory in geography explores the links between spatial form and temporal process.

5. The establishment of basic postulates for theory in geography.

7. Geographers frequently use statements of low order generality.

8. Physical geography applies laws derived from physics.

9. The concept of region in geography may perform the same function as that of class in science.

11. Geometry provides the language used in discussing the sets of relationships inherent in problems of geographical interest.

12. It has been argued that universal laws have no relevance to cultural geography because of the differences in the value systems of different cultures.

13. The literature of the subject provides a wealth of source material from which to derive an understanding of the nature of geography.

15. There are problems of formulating laws in geography as is apparent in the studies of several research workers.

16. Geographers must be aware of the empirical status of the laws derived from other disciplines.

17. Definitions of geography give rise to a set of questions pursued in research in geography.

18. Probably discard. Very generally concerned with interaction.

21. Some concepts used by geographers are derived from other disciplines.

22. Some concepts are indigenous to geography (and may be used in the development of theory in geography).

Areas of possible application

Five main areas of application of this technique spring readily to mind and they are in.

(1) Information storage and retrieval
(2) Development of reading tests of the Cloze type.
(3) Analysis of dialogue
(4) Textbook evaluation
(5) Curriculum development.

If the idealized structural network was developed using words from abstracts of a selection of research papers and books then the network diagram itself would represent a developmental history of that particular subject area. The various pathways through the network would show how a particular idea was developed and how it was related to other ideas. For example, if one was working in the field of Early Childhood Education, there is a considerable body of research literature concerned with human intellectual development and it is possible to trace the development of current practice through a sequence of related research documents. There is also, a body of related knowledge which deals with the relationship between nutrition and intellectual development. Where these two pathways intersect in the network diagram one would expect to find a selection of documents dealing with aspects of nutrition and its effect on intellectual development in ECE. There is really nothing new in this idea because in fact, this is precisely how the present library card catalogue system works. However, one can imagine with the idea or concept linked catalogue, a user of the library would be supplied with a map of his particular area of interest instead of having to rely on the card catalogue.

Secondly, in the development of Cloze type reading tests where certain words are systematically deleted and must then be replaced by the person taking the test, deletion is carried out systematically, say every eighth word. The use of text structural analysis would allow systematic deletion of words occurring in only one thematic component.

Thirdly, in the analysis of dialogue such as may take place between a teacher and a student. It would seem that the skill of a good teacher lies in his ability to match his level discourse to the conceptual map of the student. By continuously monitoring and analysing the feedback that he receives from his students, the teacher will be able to judge whether the material he is trying to teach is beyond their grasp at that point in time or not. Analyses of this kind could be used in teacher training programs to analyse discourse in microteaching sessions.

Fourthly, the analysis of textbooks. So often the adoption of a textbook or the recommendation of a particular textbook is an idiosyncratic decision. Maybe it is the layout of the pages or the particular type font that appeals to the teacher or lecturer that leads to his recommending the book for use in a particular course. How much better it would be to have some objective means of evaluation, so that the teacher can be sure that all of the stated objectives of the book have been met. A thematic and structural analysis of such a textbook would allow such an objective evaluation. In fact a doctoral dissertation is being completed here at the University of Queensland, at the moment in this very area, where two geography textbooks and several research papers are being analysed in this way in order to see if they have been written according to the accepted structure of the discipline and also to see if they are consistent with any learning theory in educational psychology.

Finally, there is the area of curriculum development which is really an extension of the textbook analysis except that it would be an ongoing adaptive procedure in that the analysis would be carried out continuously during the development of the curriculum materials. It would not mean that there would be no more field trials or a need for evaluation, but rather it would speed up the whole process of producing a final version of the materials. So much time, money and energy is presently wasted in trying to evaluate new curriculum materials, when it is an almost impossible task due to the interference effects of the materials, the classroom environment, the teacher's characteristics and the student's own motivational and personality variables.

The second half of this presentation is a description of just such an experiment in curriculum evaluation where one small part of one of the ASEP units has been rewritten and then taught in one of the Brisbane metropolitan High Schools. (See Clarke's article which follows.)

THE ROLE OF THE CONTENT AND STRUCTURE OF CURRICULUM MATERIAL IN COGNITION

John A. Clarke

Introduction

In recent years, numerous writers have highlighted the need for various types of studies of ASEP materials; in particular evaluative studies designed to determine whether the picture of science presented in the materials is valid. (Moritz, 1972; Power, 1970). What is reported here is an evaluation of the core of an ASEP unit "Pushes and Pulls" (ASEP; 1972a) without the possible distorting influence of teachers, pupils and school environment, a revision of the core on the basis of the evaluation, and a field trial of the original and revised versions.

Curriculum evaluation

A conventional approach to curriculum development is shown diagrammatically in Figure 1, where "formative" evaluation (Scriven; 1967) takes place during development and by the use of field trials.

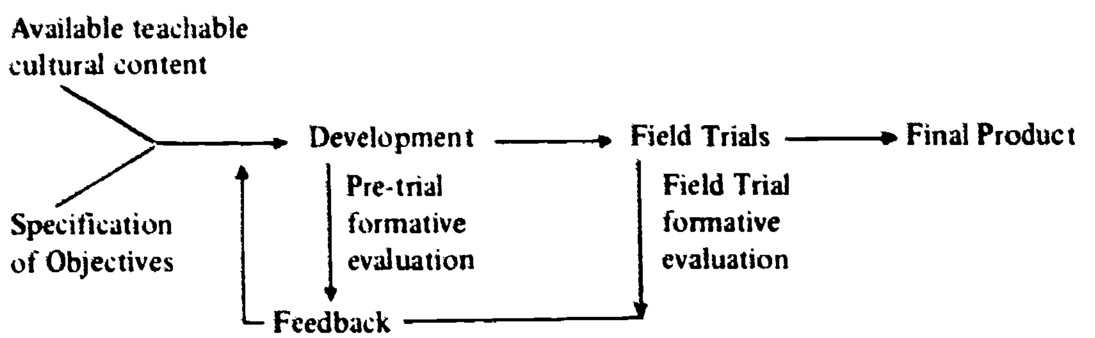


FIGURE 1

Pre-trial formative evaluation is built into most curriculum development schemes in the form of evaluation committees, (ASEP, 1971; Ramsay, 1971), referral to "experts" (Grobman, 1969; Lockhard, 1968; Welch and Walberg, 1968), etc. Questions have been asked about the objectivity of such procedures (e.g. Herron, 1966). However, most formative evaluation seems to come from within the classroom, (Lockhard; 1968) and perhaps more important are questions such as, how useful is the field-trial formative evaluation feedback, and how much is this information a true indication that the unit's objectives have been achieved? It is almost impossible to determine the amount of influence that teacher, pupil and environmental factors may have in enhancing or inhibiting the objectives of the curriculum developers. It may be suggested that what really matters is how the curriculum "works" in the field. The problem still remains however of differentiating between outcomes due to the material and outcomes due to factors other than the curriculum and knowing how much possible output has been lost due to, or distorted by, 'interference'.

Curriculum theorists Johnson (1967) and Macdonald (1965) recognize this problem when they note that curriculum evaluation is all too often conducted at the output point of instruction rather than at the input position, with the result that

> ". . . . . curriculum evaluation is confounded with instructional evaluation. Curriculum serves as a criterion for instructional evaluation, variations in instruction cannot be permitted to enter into the evaluation of curriculum" (Johnson, 1967:135).

Some suggestions have been made as to how to overcome the 'interference' problem. Johnson, for example, suggests that ". . . . differences in instructional effectiveness must be controlled, randomized, or partialled out". (Ibid) These methods may be difficult and perhaps not completely successful. The only sure way would seem to be to evaluate *before* any field trials.

This pre-trial formative evaluation, should, in part, allow an evaluation of the *potential* of the material to allow pupils to achieve stated objectives, allowing a restructuring of the content, if necessary, to more closely approximate stated objectives *before* interfering field-trial influences distort the outcomes, and hence allowing more reliability to be placed on the outcomes of field trials.

In an attempt to do this, what is being suggested is, in *addition* to the perhaps too subjective (?) pre-trial formative evaluation strategies presently being used, and in keeping with the idea that it is necessary to include elements "from the outside" in formative evaluation (Ahmann, 1967:87), that there be a more *objective* evaluation of the final pre-trial product by means of an analysis of its content.

Content Analysis

Content analysis, defined broadly as

> ". . . . . a multipurpose research method developed specifically for investigating a broad spectrum of problems in which the content of communication serves as a basis of inference" (Holsti, 1968:597).

or more specifically as

> ". . . . . a research technique for the objective, systematic and quantitative description of the manifest content of communication" (Berlson, 1952:18).

has been used to investigate a diversity of problems.

A recent comprehensive review of content analysis research is given by Holsti (1969). Most of these studies, however, use either a sophisticated analysis specifically designed for a given purpose (e.g. The General Enquirer; see Stone *et al*, 1966), or very simple frequency counts, both of which have limitations. Further, there seems to be little study in the education area, or more specifically in the science education area, most research seeming to concentrate on history, politics, etc.

The exception to this may be readability research where attributes of content are correlated with ease of comprehension. Characteristics of text which have often been tested include various aspects of vocabulary (diversity, hard words, long words, abstract words), sentence structure (length, type, number of prepositional phrases or indeterminate clauses) and human interest elements (personal pronouns, colourful words). (Holsti, 1969). Gardner (1968), using procedures based on earlier work by Flesch (1948) evaluated the readability of the P.S.S.C. text for Australian students. ASEP also used the Flesch readability formulas to prepare their materials ("The ASEP Handbook; 46-50), and have developed a checklist for analysing the "style" of their materials (Shepherd, 1972). Although measures of style and readability are obviously essential, the formulae etc. can only check on the comprehensibility of the text. They can, in no way, check that the content itself is valid. The approach here is to use a technique reported in an earlier paper (Carss; 1973) on the core of "Pushes and Pulls" (ASEP; 1972a).

(a) to isolate the important themes of the content,

(b) to detect structure, if any, and

(c) to see if the themes and the structure have the potential to achieve the stated objectives.

Once this has been done, *useful* objective pre-trial formative evaluation feedback will be available as a guide for any rewriting that is necessary.

The original curriculum (C_o)

1. *Thematic Elements*

The Principal Component analysis isolated thirteen significant components which accounted for 55.1% of the variance. In order to name these thematic elements effectively, it was necessary, occasionally, to look at the text to see in what context the words occurred. A summary of the themes appears in Table 1.

Theme 1 is obviously a "measuring" theme, combining both how to measure (stretch, mark) and what to use (rubber-band, metre-rule). The high loading of "centimetre" suggests a quantitative as well as a qualitative method of measuring. Theme 2 outlines the structure of the total unit, including the core and options plus some idea of how to record answers to small self-evaluation tests that occur throughout the unit. In Theme 3, apparatus (rubber ball on table) is being used to pose questions about forces.

Themes 4, 5 and 6 appear to be themes that one would expect to be related if any structure exists. They describe a way of "seeing" forces act (6), and what they do when they act (5), logically leading to an operational definition of detecting and measuring (4). Whether these themes are, in fact, structurally related will emerge later.

Themes 7 and 8 are "comparison" clusters. Theme 7 seems to suggest a "sensitivity" theme small and large force-measurers being calibrated against a standard. This theme may be weakened by the lack of some quantitative words e.g. tenth, unit. Theme 8 is an example of a theme where text consultation was required before a suitable name could be applied. It is suggesting, in fact, a preparation for calibr ''on using washers that have the same or different weights.

TABLE 1

	THEMATIC ELEMENT	%VAR	WORDS IN THEMATIC ELEMENT (LOADINGS)	
1	Introductory remarks on the phenomena of measurement	5.9	centimetre (.91) stretch (.90) rubber-band (.85) graph (.63) show (v) (.62)	amount (.56) mark (n) (.53) metre-rule (.52) length (.42) pull (n) (.41)
2	Outline of the unit	5.3	option (.89) test (n) (.84) core (.80) book (.70)	record book (.71) answer (n) (.59) each (.49) before (.42)
3	Posing Questions	5.2	rubber-ball (.93) happen (.87) suppose (.87)	table (.83) question (.73)
4	Definition in operational terms	4.6	definition (.83) detect (.83) how (.72) tell (.65)	try (.56) measure (.55) only (.52)
5	Comparative effects of change of state	4.5	change (v) (-.85) shape (n) (-.85) act (-.76)	much (-.56) know (-.52) force (n) (-.45)
6	Concept of a force	4.3	push (n) (.81) pull (n) (.80) see (.71) way (.53)	length (.51) word (.48) something (.41)
7	Sensitivity of measuring instruments	4.2	large (-.79) newton (-.76) small (-.60)	force-measurer (-.46) weight (-.43) blade (-.40)
8	Preparation for calibration	3.9	different (.77) weight (.58) same (.54)	many (.47) result (.46) each (.46) washer (.41)
9	Assembling apparatus and developing standard scale	3.7	card (.60) mark (v) (.59) up (.57) zero (.53)	force-measurer (.47) now (.42) your (.40) blade (.40)
10	Model of a scientist and his methodology	3.5	scientist (-.88) like (-.79) word (-.58)	find (-.45) give (-.42)
11	Instructions on performing experiments, recording and examining data	3.4	add (-.78) sinker (-.63) do (-.43)	
12	Instructions on calibrating	3.3	stop (v) (.79) mark (n) (.66) washer (.58)	blade (.56) unit (.47)
13	Procedural instructions and summary	3.3	check (-.65) page (-.62) sure (-.56)	answer (n) (-.51) before (-.46)

Theme 9 is a further "measuring" cluster but with an additional "calibration" dimension (mark, zero). Logically, it would link with or run concurrent with, Themes 11 and 12, which are "instructional" themes, one giving instructions on measuring (11), the other instructions on calibrating (12). Again any structural relations will be seen later.

Theme 10 seems to be trying to say "Scientists do things like this" or "We are going to do things the way scientists do". Theme 13 is obviously one of instructions concerning the self-evaluation that occurs throughout the unit. It is much more specific than Theme 2 which relates to the whole unit, but one would expect perhaps some structural relationship with Theme 2 and other quantitative themes (e.g. 9, 12?)

2. *Structure*

The Cross-correlation analysis makes it possible to see at which stage themes emerge and recede, and also which themes and subthemes are linked. The results of this analysis are shown diagrammatically in Figure 2. It can be seen from Figure 2 that there are four distinct parts to the unit. Part 1 (para. 1-25) appears to be the most integrated of all the parts and contains three major themes, 4, 5, and 6. It begins with a two-part introduction of eight paragraphs. Instructions, and the predicted linking of Themes 2 and 13, occur in the first part while some hint of things to come occupys the second part viz. that pupils will be "seeing" forces (Theme 6), observing their effects (Theme 5), measuring (Themes 7, 8), evaluating (Theme 13) and defining (Theme 4) these effects "just like scientists" (Theme 10). This latter section acts as an "advance organizer" for the whole core. Then follows a six paragraph block (8-13) in which forces are actually "seen" i.e. the concept of a force (Theme 6) is developed. Important sub-themes here are Themes 1 and 12 where the idea of measuring forces is introduced. What a force *is*, leads logically to what a force *does* and this theme (Theme 5) dominates the next major paragraph block (14-18). Questions are asked about what a force does (Theme 3), and links are made with what a force is (Theme 6) and a range of measuring and evaluating themes. The section culminates in paragraph 18, where some self-evaluation (Theme 13) concerning forces (Theme 6) occurs. Paragraphs 18-20 form a bridge between what forces do and the next major theme, Theme 4, which is concerned with developing a method of defining force in operational terms viz. detecting and measuring. Experimental activity is again present to some degree (Themes 11, 12, 13). Part 1 finishes in paragraphs 24-25 with a summary and self-evaluation (Theme 13) "like scientists do" (Theme 10).

The suggestion that Themes 4, 5, and 6 should logically have some structural relationship have in fact occurred. At the end of Part 1 there is a definite hiatus suggesting perhaps that a new sequence of events is about to occur.

This new sequence does, in fact, eventuate in Part 2 where, generally speaking, there is change from measuring and detecting to calibrating. One would expect, with this more quantitative approach, a need for greater precision and this is reflected by a strengthening of the evaluation strand (Theme 13), and more instructions on experimenting – assembling apparatus (Theme 9), preparing for and calibrating (Themes 8, 12) and recording and examining data (Theme 11). Some of the earlier dominant themes have all but disappeared – the "detection" (Them 4) and "examples" (Theme 6) job being virtually finished. They have laid the foundations for the more difficult calibrating themes that dominate this section.

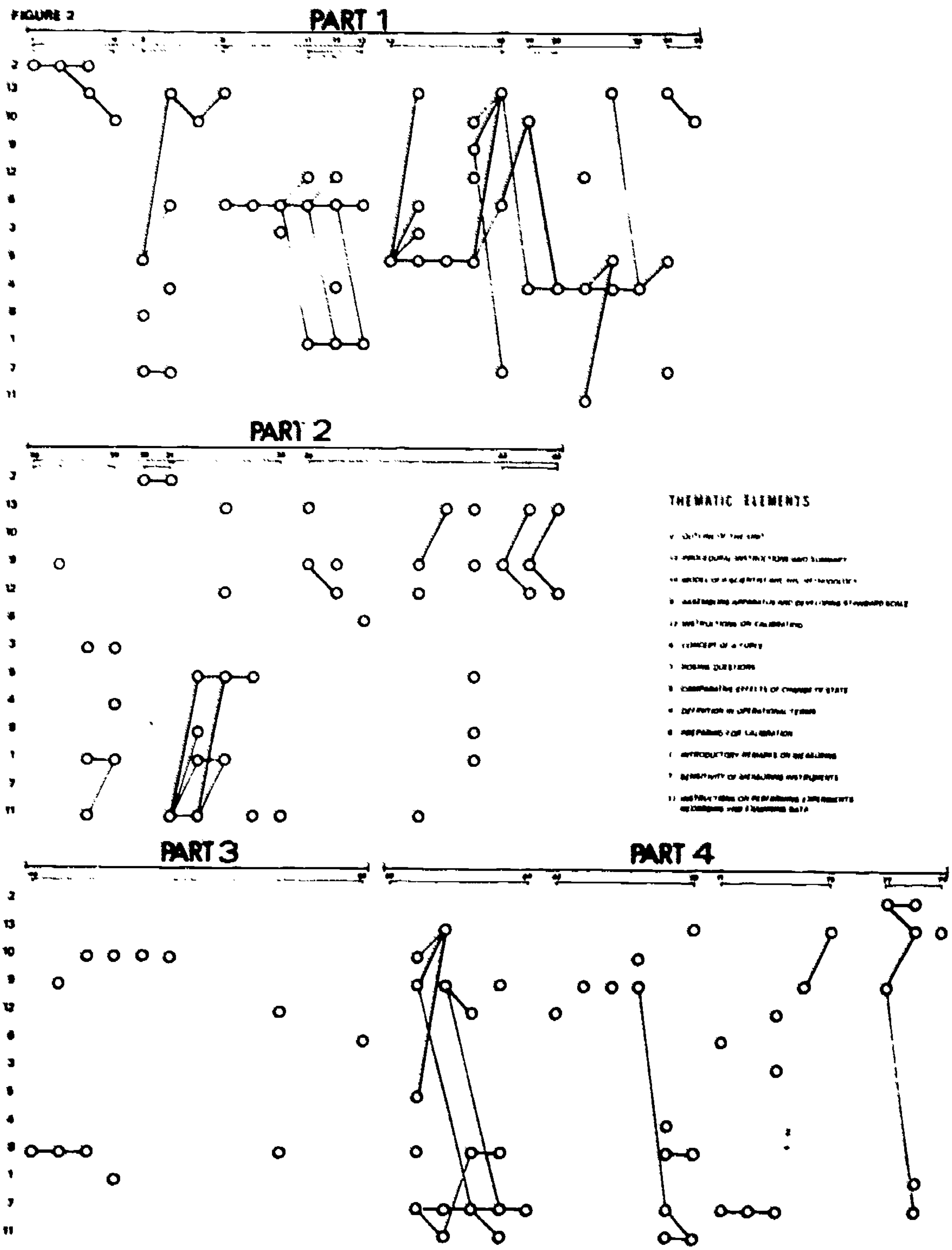


FIGURE 2
PART 1
PART 2
THEMATIC ELEMENTS
PART 3
PART 4

Part 2 also has a two part introduction but of six paragraphs only. (26-31). The first four "set the stage" for the whole of Part 2 when questions are posed (Theme 3) and there is preparation for experimentation (Themes 1, 9, 11). The remaining two paragraphs remind pupils about how to use their record booklets for this new quantitative experimentation (Theme 2). In paragraphs 31-35, the effects of forces (Theme 5) are in fact measured quantitatively (Themes 1, 11) and some hint of calibration occurs (Themes 12, 13). The remainder of Part 2 (36-45) is devoted completely to this calibration. The predicted structural relationship of Themes 9, 12 and 13 is developed as all three strands strengthen, until with the assistance of Theme 8 (para. 42) the linkages occur near the end of Part 2 (43-45). Theme 13 also acts as a means of summarizing and giving further instructions.

Part 3 is almost devoid of structure apart from an early attempt to prepare for further scientific experimentation (Themes 8, 9, 10). If the analysis is valid, one could suggest that there seems to be little of consequence occurring as far as the development of the concept of force is concerned. An examination of the content (ASEP: 1972a: 20-22) shows that it is dealing with "Good Standards for Measurement" and is attempting to develop the concept of a standard, leading eventaully to the newton. The only possibly disturbing feature of the failure of this 'standards' theme (Possibly Theme 7?) to emerge is that it is required for the next section, Part 4. Part 4 (59-79) does not seem to be as highly integrated as Parts 1 and 2 apart from a structurally complex introduction (60-64). The general themes are still those of calibrating and evaluating (Themes 9, 12, 13) but including as well Theme 7, on which "newton" loads highly (.76) suggesting comparison with a standard.

The introduction lays the basis for this standardization but the attempt does not seem to succeed, (probably due to the lack of support from Part 3) as evidenced by the middle section (65-75) which is rather disjointed. The first part (65-70) attempts to carry out the standardization, while the second part (71-76) attempts to extend the new ideas "suppose you wanted to measure something heavier than one newton" (ASEP; 1972a:27). The remainder of Part 4 (77-79) contains a core self-evaluation test and instructions on how to proceed onto the options (Themes 2, 13). There seems to be no evidence of any attempt to summarize the whole core.

The "ideal" or "essence" of the structure can be achieved as indicated in an earlier paper (Carss, 1973). The "essential structure" is shown in Figure 2A which shows that the themes divide themselves into two interlinked chains — qualitative and quantitative. The main aspects of structure that emerge are that

(a) Themes 5, 6, 9 and 11 are the important "pivotal" elements or "growing points" of the structure;

(b) Themes 1, 3 and 12 are terminal themes;

(c) The concepts present in the structure can be obtained by tracing the numerous pathways e.g. 2-13-4-5-6-10; 9-13-4-5-6-1 etc.

"ESSENTIAL STRUCTURE"

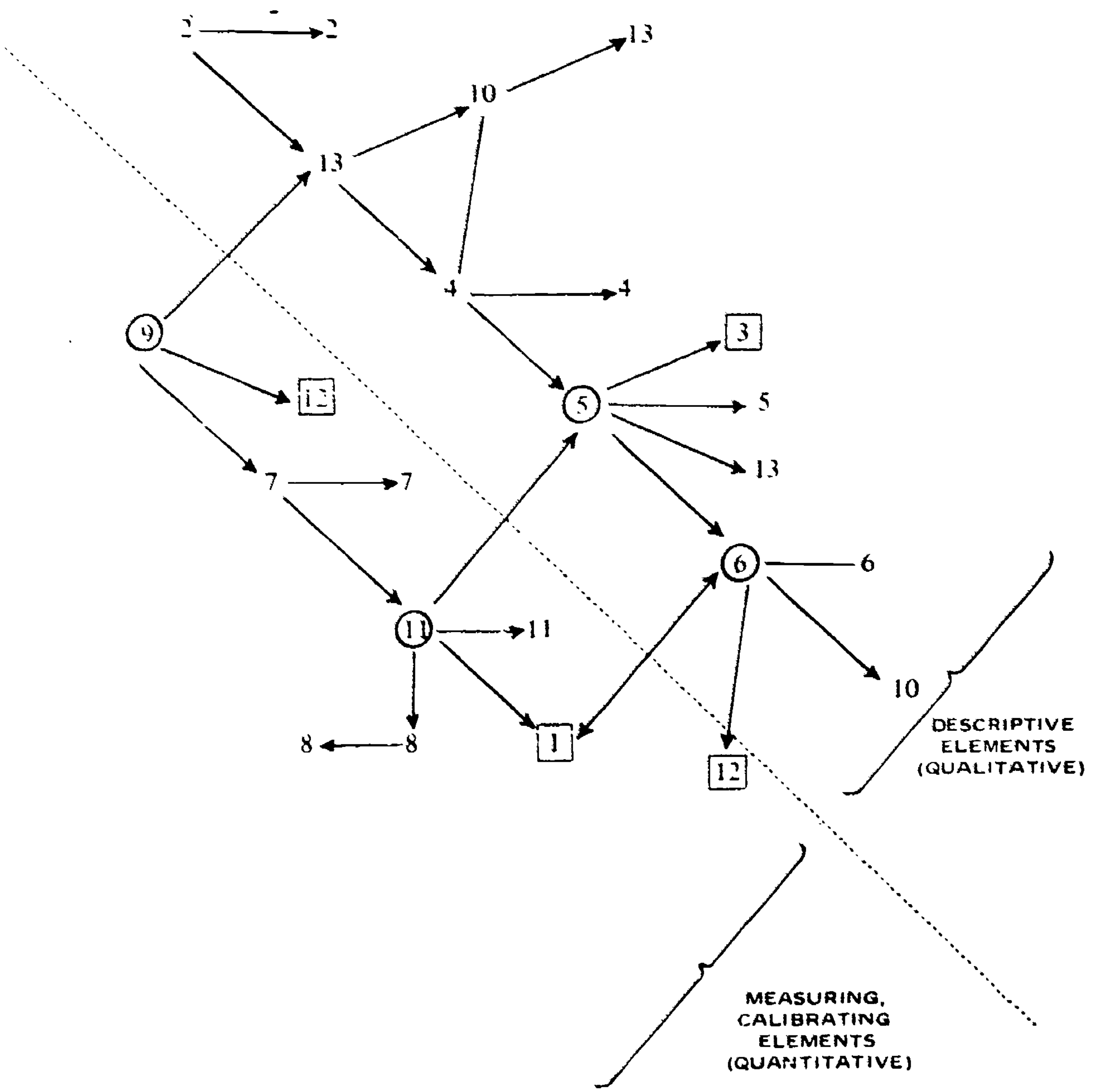


FIGURE 2A

3 *Relationship of Thematic Elements and Structure to Stated Objectives*

To determine how well the core has the *potential* to achieve the expected outcomes, a comparison can be made between the content and logic implied in the thematic elements and their structure, and the statement of expected knowledge outcomes.

Expected Knowledge Outcomes (KO) of the Core (ASEP; 1972b:vi)

K.0.1 The first activities are designed to allow the students to visualize unseen forces by observing their effects. The students are shown that forces cause observable reactions. Forces cause objects to:

1.1 Change shape.
1.2 Start or stop moving.
1.3 Slow down or speed up.
1.4 Change direction.

K.0.2 A scientist gives a description (or definition) of a thing by answering two questions.

2.1 How can I tell if I have some of this?
2.2 How can I tell how much of it I have?

K.0.3 You can tell how much force you have by the amount that it changes the shape of something.

K.0.4 It is necessary to have a calibrated scale to measure something.

K.0.5 It is important to have standards of measurement.

K.0.6 When forces cause certan objects to change shape these objects do not return to their original shape. This characteristic makes them less useful than others as force measurers.

K.0.7 Weight is one kind of force.

K.0.8 An instrument must be zeroed before making a measurement.

K.0.9 The newton is the standard unit of force.

K.0.10 An instrument can be calibrated against a standard instrument.

K.0.11 The range of an instrument can be extended by chaning its sensitivity.

K.0.12 A force measurer can be used to measure forces other than weight.

K.0.13 Knowledge that a standard must be plentiful, of suitable size, easily duplicated, agreed on by everyone.

TABLE 2

THEMATIC ELEMENT	EXPECTED KNOWLEDGE OUTCOMES												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1				X								X	
2													
3	X	X	X										
4		X	X									X	
5	X		X			X							
6	X												
7					X?		X?		X?	X	X		
8					X?		X						
9				X				X					
10		X											
11				X				X					
12				X									
13													

A comparison of thematic elements and knowledge outcomes is shown in Table 2, and suggests that the majority of expected outcomes are represented to varying degrees. However, some either do not occur (K.0.13) or are only weakly represented (K.O's 5, 6, 7, 9, 10, 11). Of the latter, K.O's 6, 7 seem highly specific in that they would probably need only to occur once. The remainder, along with K.O.13 form a "standards" cluster. Examination of the content show that the development of the concept of the standard newton and calibration of instruments of varying ranges against this standard is the responsibility of Parts 3 and 4 of the core. It is in these areas that the material is most poorly structured.

The broad aim of this ASEP unit is

> ". . . to allow the students, through his own investigations, to arrive at an understanding of concept of a force. In the process, the student builds a force-measurer and calibrates a scale in newton units . . . . In addition, the student activities are designed to develop a scientific approach to problem solving, the gathering of data for interpretation and the presentation of data." (ASEP; 1972b:ii)

and the analysis suggests that the content has the potential to do this up to a point. The analysis also shows that the content seems to be aimed at the correct cognitive readiness level of students. Ausubel, working within a Piagetian framework suggests that

> ". . . it would appear that, (in the concrete operational stage), given only the stimulus support provided by concrete images of examplars of criterial images, the child would be able mentally to construct a representative image embodying the meaning of the concept."
> (Ausubel and Robinson; 1969:185).

and the high emphasis on "doing" and "seeing" would help to achieve this.

Questions that the curriculum developers may now ask, however, are:

(i) Do themes emerge and fade at appropriate times? The non-emergence of a "standards" theme at the appropriate time (i.e. paragraphs 48-58) may be a cause for concern.

(ii) Is there any "dead wood" or poorly structured material present that may be interfering with the development of a theme or the concept as a whole e.g. in paragraphs 48-58, 65-75?

(iii) Is there any need to "tighten up" the structure of the latter sections which are not as well integrated as the early sections.

The inadequacies of the core highlighted by this analysis of the content, may be overcome by a revision of Parts 3 and 4.

The revised Curriculum (C_R)

1. *Theoretical Basis of the Revision*

Bruner (1966:18), in discussing curricula construction, regards one of the major problems as being

> ". . . . . how to have basic subjects rewritten and their teaching materials revamped in such a way that the pervading and powerful ideas and attitudes relating to them are given a central role."

Some learning theorists have delved deeply into this idea that the *structure* of the material presented to learners is of the utmost importance, the best known perhaps being Ausubel who says, for example, that

> " . . the major implication for teaching is that control over the accuracy, clarity, longevity in memory and transferability of a given body of knowledge can be most effectively exercised by attempting to influence the crucial variables of cognitive structure . . . . e.g. by employing suitable programmatic principles of ordering the subject matter and constructing its internal logic and organization." (1964:168)

His ideas of "advance organizers" etc. are well documented (Ausubel; 1960, 1963) and have been used successfully in science education research (see for example, reviews by Kuhn, (1972) and Novak *et al* (1971).

Bruner (1964:309) suggests that the structure of any domain of knowledge may be characterized in three ways, each effecting the ability of any learner to master it, the mode of presentation, its economy and its effective power. Figure 2 and the subsequent discussion suggest that some of these conditions particularly the last two may not be present as much as they could be. What seems to be required is some restructuring and rewriting of Parts 3 and 4 (or slightly before), with emphasis being given to economy (no "dead wood") and effectiveness (potential to achieve objectives) without loss of appropriate level of difficulty, readability, visual impact etc. (mode of presentation). The Ausubelian ideas of "advance organizers", "integrative reconciliation" etc. may act as theoretical guides for the rewriting. This new version can then be reanalysed by content analysis to see if the rewriting has been successful in overcoming the apparent deficits of the original content.

2. *The Revised Curriculum*

Paragraphs 46 to 79 were rewritten and reduced by three paragraphs. The source of the original ASEP material proved to be most useful here i.e. "Probing the Natural World. Vol.1" (ISCS:1970). This reduction occurred in Part 3 where most of the "dead wood" seemed to be. An example of the rewritten material is as follows:

ORIGINAL CURRICULUM (C_O)	REVISED CURRICULUM (C_R)
"Pushes and Pulls" (1972a:20)	
Para. 47 (Refers to Figure 2)	47R
By now you know that your "washers" card is different from the cards of some of your neighbours. You may have even guessed why. Although they looked the same the piles of washers in each pile weighed the same but those in different piles had different weights. You can check this by 'feeling' the weight of a pile of each colour. This is why you and many of your classmates got different results when you weighed the sinker.	By now you know that your "washers" card is different from the cards of some of your neighbours. You may even have guessed why. Go back and think about how you made the scales on your "washers" card.[1] You made a mark each time a washer was added to the cup. Anything (such as a washer) that is used to make a scale is called a *standard*.[2]
48	48R
The scales made with different washers will not be the same. They differ because members of your class used different standards[2] to make their scales. If everyone had used washers that weighed the same, the marks on all cards would have been in the same place (the scales would be the same).	Now compare your scale with the scale of your classmates. If the "washer units" on your scale are not the same, it means you must have used different standards to make them. Thus, the washers of different colours must have different weights. You can check this by 'feeling the weight' of a pile of each colour. This is why you and many of your classmates got different results when you weighed the sinker. Q.23. How would all members of your class be able to produce the same scales?[3]

1. Integrative reconciliation (Ausubel: 1963:80-1).

2. In C_O, this is the first mention of the word "standard". There would seem to be a need to introduce this earlier to avoid confusion between "standard" and "scale". This is done in C_R by means of a definition an "expository organizer". (Ausubel: 1960).

3. To promote problem solving and test understanding of the section.

The new core (paragraphs 1-45 of the original core and the revised paragraphs 46-76), was then analysed as before. The resultant thematic elements are shown in Table 3 and the structure of paragraphs 46-76 is shown in Figure 3.

TABLE 3

	THEMATIC ELEMENT	%VAR	WORDS IN THEMATIC ELEMENT (LOADINGS)	
1	Introductory remarks on the phenomena of measurement	5.6	centimetre (.92) stretch (.92) rubber-band (.81) graph (.69) show (v) (.64)	amount (.51) metre-rule (.49) mark (n) (.49) pull (n) (.42) need (.40)
2	Outline of the unit	4.9	option (-.92) core (-.83) test (n) (-.80) book (-.71)	record-book (-.66) answer (-.51) each (-.50)
3	Posing questions	5.0	rubber-ball (.94) suppose (.91) happen (.86)	table (.83) question (.73)
4	Definition in operational terms	3.7	definition (.81) how (.67) detect (.65) tell (.62)	measure (.51) try (.50) only (.42)
5	Comparative effects of change of state	4.2	shape (.76) act (.75) change (.74)	much (.66) know (.59) force (n) (.52)
6	Concept of a force	4.2	see (.73) push (n) (.70) pull (n) (.68) way (.66) find (.53)	measure (v) (.49) something (.49) length (.44) word (.43)
7	Sensitivity of measuring instruments	4.9	tenth (.86) weigh (.76) newton (.71) weight (.68) small (.60)	unit (.57) sinker (.51) do (.43) try (.41)
8	Preparation for calibration	3.2	collect (-.73) large (-.71) sure (-.62)	before (-.57) blade (-.41)
9	Assembling apparatus and developing a standard scale	3.7	zero (-.75) force-measurer (-.56) card (-.54)	mark (-.52) paper-hook (-.45) now (-.41)
10	Model of a scientist and his methodology	3.2	scientist (.82) like (.77) word (.59)	many (.48) give (.47)
11	Instruction on performing experiments, recording and examining data	3.5	time (-.66) add (-.62) each (-.60)	amount (-.53) same (-.48) sinker (-.41)
12	Instructions on calibrating	3.1	stop (-.77) mark (n) (-.65) washer (-.53)	move (-.41) blade (-.41)
13	Procedural instructions and summary	3.7	different (.78) check (v) (.76) page (.71)	answer (n) (.60) answer (v) (.52)
14	Characteristics of a good standard	4.2	standard (.84) measurement (.82) good (.76) object (n) (.48)	make (.47) choose (.46) all (.43)

3. *Thematic Elements*

As might be expected, there was a redistribution of variance and changes in the loadings of words but the order of the original thematic elements has been maintained in Table 3 to allow easy comparison. As might also be expected, there was a large degree of correspondence between the two sets of themes. There were however, four significant changes. Firstly, and perhaps most importantly, a completely new thematic element emerged (Theme 14) viz. characteristics of a good standard. This theme should have emerged in Part 3 of the original curriculum. It should appear in this region in the structure of the revised curriculum. Secondly, Theme 7 was strengthened by the inclusion of words such as "tenth" and "unit" and can now be regarded with greater confidence as a "sensitivity" theme. Thirdly, in similar fashion, Theme 11 is now a much more powerful experimental instructions theme as a result of the inclusion of more relevant words such as "each", "amount", "same", etc. Finally, the role of Theme 8 was clarified as being one that represents preparation for a calibrating activity.

4. *Structure*

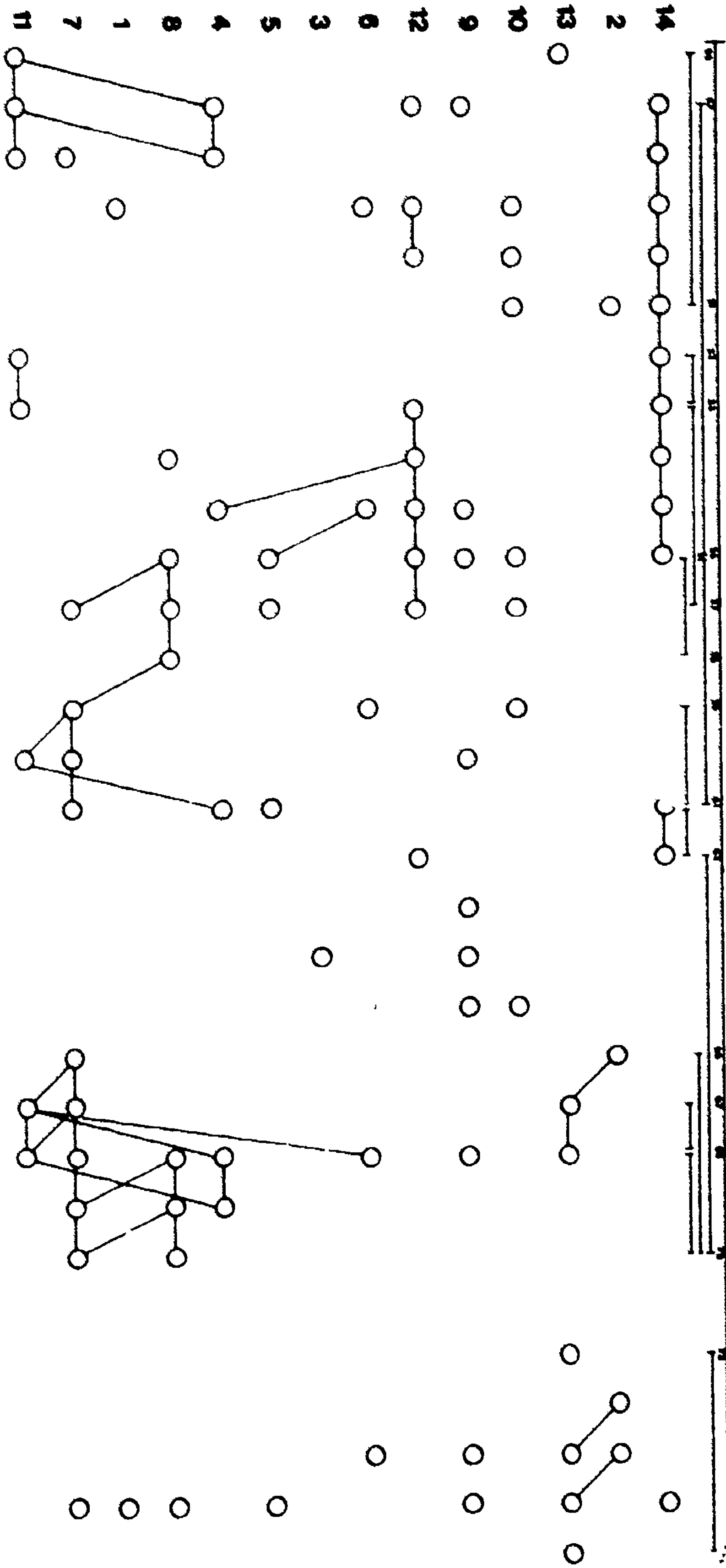
Since Parts 1 and 2 of the original core were not rewritten, their structure can be regarded as remaining the same. Figure 3 immediately suggests two significant differences when compared to the corresponding sections of Figure 2. Firstly, the rewritten section has much more structure and secondly, is more integrated to the point where it can be regarded as one section (Part 3R). It consists of a series of sub-sections that either overlap or are connected by "bridging" themes. The exception to this integration may be paragraphs 73-76 which act as a core summary.

Two structures dominate Part 3R — the new theme (Theme 14) which binds the early sub-sections together and paragraphs 66-70 where the all-important standardization occurs. These aspects were either missing or only weakly represented in the original core. Theme 14 develops over paragraphs 47-56 and during this development, the need for a standard is tied to the measuring and calibrating skills developed in Parts 1 and 2 and reintroduced here as a preliminary to the development of the standard unit of force after Theme 14 recedes. In the first sub-section (46-51) experimental instructions (Theme 11) on detecting and measuring (Theme 4) and calibrating (Theme 12) forces are reintroduced with, however, a new aim — the development of a standard. Experimental instructions (52-53) act as a bridge to the second sub-section (53-57) and an overlapping third subsection (56-62) where calibration (Themes 8, 12), now using a standard scale, leads to the emergence of the standard unit of force — the newton (Theme 7; 59-61).

Theme 14 then acts as a bridge to the last major sub-section of Part 3R (62-70) where the standard newton is then used to measure small and large forces. This can be predicted from the dominance of paragraphs 66-70 by the now strengthened sensitivity theme (Theme 7) and the presence of experimental and evaluation themes (Themes 11, 13: 67-68; Theme 8: 68-70).

The final section (72-76) summarizes what has occurred throughout the core, as evidenced by the presence of most of the thematic elements, as well as giving instructions as to how to proceed with the options in a scientific way (Themes 10, 13). This "tying together" was lacking in the original core.

FIGURE 3



5. *Conclusion*

It now seems that "on paper" the revised version of the core (C_R) has more potential than C_o to achieve the stated objectives of the curriculum writers. The new th Theme 14 and the strengthened Themes 7 and 11, along with the improved structure of the latter sections should overcome the inadequacies of the original core. One method of checking on this would be to compare the effects of both versions on pupils "in the field". Before reporting on the field trial, it should be noted that C_R is by no means the "best" possible version of the core. C_R itself could easily be subjected to further rewriting and content analysis any number of times e.g. para. 62-66 may well need further revision.

Field Trial

1. *Organization and Sample*

During 1973, science teachers under Mr. S. Mackenzie at Wavell State High School in Brisbane changed the normal Grade 8 (first secondary year) science program to one of eight ASEP units, four to be done per semester. "Pushes and Pulls" was one of the first semester units. Due to school organizational "problems" (see Clarke; 1973a) eight class groups (four groups of two) were chosen from an available twelve to act as the sample. The groups were chosen to avoid groups with unique characteristics, to minimize the teacher variable and to provide an opportunity for replications. The sequence of groups is shown in Table 4.

TABLE 4

GROUP	TEACHER	SEQUENCE			
		29/1 to 2/3	5/3 to 30/3	2/4 to 4/5	21/5 to 15/6
1 8A7	Mrs. A	P	-	-	-
8A9	Mrs. A	P	-	-	-
2 8A3	Mrs. B	-	P	-	-
8A5	Mrs. B	-	P	-	-
3 8A2	Mr. C	-	-	P	-
8A4	Mr. C	-	-	P	-
4 8A6	Mr. D	-	-	-	P
8A10	Mr. D	-	-	-	P

"P" : Pushes and Pulls "-" : Other ASEP units.

2. *Instruments*

(a) ASEP booklets.

In preparing the materials, an attempt was made to maintain the "ASEP atmosphere" by having individual core and record booklets and by making the revised version very similar in appearance to the original version e.g. the same illustrations in much the same position. There were, however, some deviations from the normal ASEP format viz. booklets were reproduced in foolscap form and a special separate "Check Your Answers" sheet was prepared to replace the "Turn to p.119 and check your answers" type of instruction. The core booklet, record booklet and "Check Your Answers" sheet for each version had a consistent colour difference.

(b) Pre- and Post-tests.

Short objective pre- and post-tests were produced by the author. Each test contained two distinct sub-tests — one sub-test contained questions related to the section of the core not revised (para. 1 to 45) and the other sub-test contained questions related to the section of the core that was revised (para. 46-79). Details of the construction of these questions, in particular their relation to relevant K.O's, are available elsewhere (Clarke, 1973a). Reliability measures, however, are not yet available.

3. *Methodology and Design*

Teachers were given verbal and written instructions (see Clarke, 1973a) and during the first "Pushes and Pulls" meeting, administered the pre-test, randomly assigned pupils to control (original ASEP core) and experimental (revised ASEP core) groups, distributed the booklets and introduced the pupils to the "hardware" they would be handling during the unit e.g. force-measurers, sinkers etc. The distribution of the sample to this point of time, is shown in Table 5. The original sample has been reduced by 29 so far, due to pupils not completing both pre- and post-tests, not having a school record card, transferring to another school etc. Table 5 shows the reduced figures.

TABLE 5

Group	1	2	3	4
Teacher	A	B	C	D
Control	27	27	34	*
Experimental	35	28	23	*
Total	62	55	57	*

* Unavailable at this time.

As each individual finished the core (4-5 forty minute periods) they completed the post-test. The pupils were then issued with the "real" ASEP books and allowed to proceed onto the options. This pre test, post-test, control group design, with random assignment of individuals to experimental and control groups, is regarded by Campbell and Stanley (1967: 183) as a true experimental design.

Teachers and pupils were asked to indicate problem areas throughout the core by teachers keeping a list of pupil questions that related specifically to understanding of the content of the core, and by pupils marking sentences or paragraphs they had difficulty in understanding. This technique teacher and pupil reports is the most common formative evaluation technique (Lockhard, 1968) but is regarded by reviewers of curriculum evaluation research (Clarke, 1973b; Lockhard, 1968; Welch, 1969) as being of questionable use for a number of reasons e.g. the feedback is not suitable to allow changes to be made in the materials. A preliminary analysis of the feedback has already shown the advantage of content analysis. A control group pupil's question "Why do we have to change the hacksaw blades?" reflects confusion concerning sensitivity. Content analysis of the original core could have predicted such confusion. However, in many cases it seems that teachers and pupils are just too busy in the ASEP classroom to provide feedback. There may be a need to provide teachers and pupils with some structure and system within which they could provide usable feedback, or to get the feedback independently of them e.g. video-tape.

4. *Results and Discussion*

Scores obtained in the four sub-tests were analysed using a stepwise discriminant analysis. Results are shown in Table 6.

TABLE 6

		GROUP 1		GROUP 2		GROUP 3		TOTAL	
		F	df	F	df	F	df	F	df
Pre-test	sub-test 1	2.28	61	-	-	1.63	52	2.01	172
	sub-test 2	1.23	59	0.21	51	2.18	53	-	-
Post-test	sub-test 1	1.64	60	0.28	52	3.10	54	2.95	173
	sub-test 2	**3.77	62	0.40	53	*4.13	55	*5.50	174

- insufficient for computation
* significant at 0.05 level.
** F = 3.99 is significant at 0.05 level.

A significant difference between control and experimental groups was found with group 3 and the total sample, while group 1 approached significance, the one discriminator in each case being Post-Test sub-te 2 the set of questions related to the rewritten paragraphs. Group 2 differences were not significant but again, the same sub-test discriminated most. It seems that some of the "instructional evaluation interference" metnioned at the beginning of this paper may account for group 2 results. The pupils in this group took nearly three times as long to complete the core as did those in groups 1 and 3. (10-12 periods c.f. 4-6 periods). Over this extended period one could see perhaps an increase of teacher influence at the "expense" of the text influence.

At this early stage of data analysis, the results are very encouraging but can only be accepted with optimistic caution. What has yet to be done is to establish beyond reasonable doubt, the reason for the significant difference. Hopefully, it will be just the text material but other variables may explain the difference e.g. sex, intelligence, verbal abilities etc. Measures of general scholastic aptitude (intelligence), word knowledge and reading comprehension are available from the results of a standardized test battery administered to all Queensland state primary schools late in their Grade 7 year and multiregression analysis should isolate the discriminating variable.

Other contemplated analysis will be looking for individual *questions* (as opposed to sub-tests) that discriminate between the control and experimental groups, a more detailed examination of the teacher and pupil "feedback", a look at the effect of the time taken to complete the core, and a look for any "ASEP familiarity" effect that may emerge in the later groups as a result of their exposure to other units before "Pushes and Pulls".

Summary and Conclusion

What has been reported here is an objective, pre-trial formative evaluation of the core of an ASEP unit designed to show its potential in terms of themes and structure, a revision on the basis cognitive-field learning theory of those sections of the core shown by the analysis to be possibly deficient, and an ongoing field-trial of the original and revised cores.

The field-trial results, though only tentative, have demonstrated the use of the Carss (1973) content analysis technique as it may be applied to formative curriculum evaluation since it supplies ". . . . . the diagnostic analysis which is at the heart of formative evaluation" (Baumgart, 1972:9). The technique produces, prior to field-trials, usable information with respect to the potential inherant in the text, showing for example where structure is poor or where there is irrelevant or too much information. The technique also allows feedback from field-trials to be used more constructively since general statements or questions, hitherto of little use in giving writers directions for revision, can be linked to themes, the content and distribution of which are available.

The use of this technique for the analysis of the manifest content of communication-written or verbal seems unlimited. The analysis of the teacher and pupils verbal communications in parallel with the textual analysis, and the analysis of "expert" ideas on "what is science?" to obtain the "essence" of science are but two. The technique itself needs further refinement and this refinement and possible uses offer an interesting challenge for future research.

REFERENCES

Ahmann, J.S. 1967. 'Aspects of Curriculum Evaluation: A Synopsis' in *AERA Monograph Series on Curriculum Evaluation 1. Perspectives of Curriculum Evaluation.* R.W. Tyler *et al* Rand McNally and Co. Chicago, 84-89.

ASEP. 1971. *Newsletter No. 5.* Ramsay, G.A. (ed).

ASEP. 1972a. *Pushes and Pulls. National Trial Print.* Toorak.

ASEP. 1972b. *Pushes and Pulls. Teachers Guide. National Trial Print.* Toorak.

ASEP. 401. *A Guide to ASEP. The ASEP Handbook.*

Ausubel, D.P., 1960. "The Use of Advance Organizers in the Learning and Retention of Meaningful Verbal Learning". *Journal of Educational Psyschology*, 51, 267-272.

Ausubel, D.P., 1963. *The Psychology of Meaningful Verbal Learning.* Grune and Stratten. N.Y.

Ausubel, D.P., 1964. "Some Psychological Aspects of the Structure of Knowledge" in S. Elam (ed) *Education and the Structure of Knowledge.*

Ausubel, D.P. and Robinson, F.G., 1969. *School Learning: An Introduction to Educational Psychology.* Holt, Rinehart and Winston. N.Y.

Baumgart, N.L., 1972. "A Model for Curriculum Evaluation". *Research 1972.* R.P. Tisher (ed). University of Queensland Press. Brisbane.

Berelson, B., 1952. *Content Analysis in Communication Research.* Glencoe, Ill. Free Press.

Bruner, J.S., 1964. "Some Theorems and Instruction Illustrated with reference to Mathematics" in *Theories of Learning and Instruction. Part I.* E.R. Hilgard (ed). N.S.S.E. Chicago.

Bruner, J.S., 1966. *The Process of Education.* Harvard Uni. Press. Cambridge.

Campbell, D.T., and Stanley, J.C., 1967. "Experimental and Quasi-Experimental Designs for Research on Teaching". *Handbook of Research on Teaching.* N.L. Gage (ed). Chicago. Rand McNally 171-246.

Carss, B.W., 1973 — preceding paper, this publication.

Clarke, J.A., 1973a. "The Role of the Structure of Curriculum Content in Cognition". M.Ed.St. Thesis. University of Queensland. In process.

Clarke, J.A., 1973b. "Curriculum Evaluation: Strategies, Problems and a Possible Solution". Unpublished manuscript. Psychology Department. Kelvin Grove College of Teacher Education, Brisbane, Queensland.

Flesch, R., 1948. "A New Readability Yardstick". *Journal of Applied Psychology*, 32, 230.

Gardner, P.L., 1968. "The P.S.S.C. Text: a measure of its readability". *Australian Science Teachers' Journal*, 14(3), 71-74.

Grobman, A.B., 1969. *The Changling Classroom: The Role of the Biological Sciences Curriculum Study*. Garden City, N.Y. Doubleday.

Herron, M.D., 1969. "The Nature of Scientific Enquiry as Seen by Selected Philosophers, Science Teachers, and Recent Curricular Materials". Doctor's Thesis, Univ. of Chicago.

Holsti, O.R., 1968. "Content Analysis" in *The Handbook of Social Psychology, Second Edition, Volume 2*. G. Lindzey and E. Aronson (ed), p.596-692.

Holsti, O.R., 1969. *Content Analysis for the Social Sciences and Humanities*. Addison Wesley, Mass.

ISCS, 1970. *Probing the Natural World. Volume 1*. ISCS Program, Level One. Florida State University.

Johnson, M., 1967. "Definitions and Models in Curriculum Theory". *Educational Theory*, 17 (3), 127-140.

Kuhn, D.J., 1972. "Science Teaching, Concept Formation and Learning Theory". *Science Education*, 56 (2), 189-196.

Lockhard, J.D. (ed), 1968. *Sixth Report of the International Clearinghouse on Science and Mathematics Curricular Developments*. College Park, Univ. of Md.

Macdonald, J.B., 1965. "Researching Curriculum Output: The Use of a General Systems Theory to Identify Appropriate Curriculum Outputs and Research Hypotheses". Paper presented at AERA Meeting.

Moritz, K.W., 1972. "Some Suggestions for Research related to ASEP". *Research 1972*. R.P. Tisher (ed). University of Queensland Press. Brisbane.

Novak, J.D., Ring, D.G. and Tamir, P., 1971. "Interpretation of Research Findings in Terms of Ausubel's Theory and Implications for Science Education". *Science Education*, 55 (4): 483-526.

Power, C.N., 1970. "A Review of Current Research in Science Education in Australia". *Australian Science Teachers' Journal* 16 (2), 51-56.

Ramsay, G.A., 1971. "Formative Evaluation and the Australian Science Education Project". *Research 1971*. R.P. Tisher (ed), Toorak.

Scriven, M., 1967. "Methodology of Evaluation". *Perspectives of Curriculum Evaluation*. R.E. Stake (ed) AERA Monograph Series on Curriculum Evaluation No. 1. Chicago. Rand McNally. 39-93.

Shepherd, S.T., 1972. "A Checklist for Analysing the 'Style' of Instructional Materials", in *Research 1972*. R.P. Tisher (ed.) University of Queensland Press. Brisbane.

Stone, P.J., Dunphy, D.C., Smith, M.S. and Ogilvie, D.M., 1966. *The General Enquirer: a computer approach to content analysis in the behavioural sciences.* Cambridge. M.I.T. Press.

Welch, W.W., 1969. "Curriculum Evaluation". *Review of Educational Research*, 39 (2): 429-443.

Welch, W.W., and Walberg, H.J., 1968. "A Design for Curriculum Evaluation". *Science Education*, 52, 10-16.

REFLECTIONS ON THE SOCIAL CONTENT OF SCIENCE EDUCATION MATERIALS

Peter J. Fensham

This paper is a reflection on the findings of a research investigation rather than the investigation itself. It could be rewritten in a more traditional form which would suggest that a deliberate plan and design had laid out the procedures. In the complex fields of a social science like education, it is, of course, important to keep such planning and design in mind but its dangers need also to be remembered. In the physical sciences, good design is usually a much simpler matter and carefully planned studies can be carried out. However, many of the mainsprings for the more important findings and lines of advance come from the unplanned by-products of earlier investigations, or really critical appraisals of the restricted thinking that is inevitably associated with a tight design.

The planned study in this case was the development of methods of content analysis for use with the materials of various science curricula. The content was to be analysed for various social dimensions. There has been a growing interest since 1968, the year of the Dainton Report, in the social relevance of science education.

Two social aspects of science led to the first two dimensions.

The first is the social nature of science itself. Pragmatically, this is mentioned by the Dainton Committee when it says that students should study science because it is one of the great achievements of humanity and so often provides examples of the best of man's creativity and inventiveness. The theoretical base for this social dimension of science comes from the work of Storer[2], Merton[3] and other sociologists of science. They argue that science and scientists form a social system with its own value system and special means of communication that transcend national and cultural boundaries. In science education it may be that some of these aspects of scientific work could be appreciated and learnt. Several primary level courses have emphasised this point[4]. The presence of this dimension in the materials of a course could take many forms. Among these would be case studies of scientists and of the exchanges between them that led to the development, and reputation of experimental results, concepts and theories. Exercises that require the pooling of data would be emphasised and teachers, texts and data tables as sources would be repeatedly related to the energies and work of past and contemporary scientists. The contributions of the technicians who lie behind apparatus and instruments would be made explicit.

The second social dimension is also to be found in the Dainton Committee's critique of the teaching of science in English secondary schools. It is the application of scientific knowledge in society technology and all its implications. Examples of past and contemporary application of science, and of how scientists were set on the track of many of their discoveries from real life situations, would be evidence of this dimension in the materials of a course.

The third dimension is derived from the second one the interaction of scientific knowledge with society. It is essentially a set of ideological positions that has become meaningful in relation to the curriculum of formal education since the year of the Dainton Report. It is hard, only five years later, to believe that such a major report on science education could be written without the phrase "social responsibility of science" being used. The environmental issue, the rise of SRS groups in many countries and the counter-culture movements in these intervening years, have made us aware that there can be differing ideological positions about the interaction of science with society. The Dainton Report came at the end of an era when the established view of science and of science education was that 'it was good, meet and right so to do'. It is now recognised that, as well as the position that sees applied science as problem solving for society, there is a position that sees it as creating problems. In other words, science is viewed as a source for the improvement of the quality of social life or for its deterioration. Both positions can also be held at once and such an ambivalent view of applied science is now increasingly the norm. As soon as a medicinal breakthrough is announced, the question of its side effects are raised, etc.

Procedures and results

The procedures for making the content analysis on these three dimensions have been described in detail elsewhere[5]. The first two dimensions are finally scored on the five point scale 0, very weak, weak, strong, very strong. The last is scored with + and – depending on what positions are implicitly or explicitly taken in the inclusion and discussion of examples of social application of science. The number of + or – signs signifies in a crude manner the strength of the presentation of the ideological positions.

Table 1 shows the results of the analysis for a number of chemistry text books that are used in conjunction with courses in Australia or elsewhere.

TABLE 1

Social content of textbooks for secondary chemistry courses

TEXT	SOURCE	SOCIAL CONTENT		
		Nature	Application	Ideology
Chemistry Takes Shape	Johnson and Morrison Scotland	weak	strong	+ +, –
A Modern Approach to Chemistry	Stove and Phillips Victoria, Australia	weak	very strong	+ + +, – –
Chemistry – A Structural View	Stranks *et al* Victoria, Australia	very weak	very weak	+
CHEM STUDY	Parry U.S.A.	very weak	very weak	+
C.B.A.	U.S.A.	weak	0	0
Chemistry The Sample Scheme	Nuffield Foundation England	weak	very strong	+ + +
Chemistry	Nuffield Advanced Science England	weak	strong	+ +, –

All the books except the first edition of Stove and Phillips are associated with curricular reforms of the substantial variety that began with the PSSC project in the late 1950's. This exception was the spare time work of two practising school teachers and was a text for a course that had known only the minor revisions that science curricula had enjoyed for most of this century until the 1960's. Although published in 1963, after the major overseas projects had begun it was uninfluenced by them. It was an excellent example of the ype of text in common use before the project era.

The data in Table 1 indicate a polarisation of the texts on the second dimension. They are either "very weak" in social applications of chemistry or "strong" to "very strong". The score on the third dimension is, of course, affected by the result on the second so the "weak" texts have a positive position but largely implicitly. The others present their positions strongly, but no examples of strong negative positions appear until the texts of the late 1960's and 70's.

Table 2 presents the results of a similar analysis for the texts of some other science course of this same period. The results are similar.

TABLE 2

"Social" content analysis of secondary science materials

	Published	Level	Science as "Social"	Social Application	Ideological Stance
J.S.S.P.	1967	7-8	0	weak	+
Discovery in Science	1970	7-10	0	weak	+
Discovery (Revised)	1972	7-10	weak	strong	+,-
Nuffield Combined	1970	7-8	0	weak	+
Nuffield Secondary	1971	9-11	strong	very strong	+
A.S.E.P. (Aust)	1974	7-10	weak	very strong	+,-
S.C.I.S.P. (U.K.)	1974	9-11	very strong	very strong	+,-
P.S.S.C. Physics (U.S.A. & Victoria)	1959	11-12	weak	very weak	+
Nuffield "0" Physics	1966	7-11	0	weak	+

JSSP, early Discovery, PSSC, and Nuffield Combined Science are examples of "weak" social content. Nuffield Physics and revised Discovery are more like the "strong" group of chemistry texts and the post – 1970 ones, Nuffield Secondary Science, ASEP and SCISP are all "very strong".

The procedures have now been applied by several analysts to a number of texts and consistent scoring is obtained in relation to the defined criterion of the dimensions.

We may therefore conclude that Australian secondary science is still largely gripped by peculiarly a-social courses, since PSSC, CHEM STUDY and its derivates, and JSSP are very prominent source materials. Indeed, none of these courses can exist apart from their textual materials and the texts have become the course for most teachers and students.

Discussion of the results of the analysis

The results in Tables 1 and 2 can be interpreted in a number of ways. One simple correlation is to look at the sorts of people who wrote the various texts and determined the details of the projects of which many of them form the basic part. If the projects are separated into those in which research-oriented scientists were prominent and those essentially in the hands of experienced secondary teachers there is fairly good correspondence with the "weak" and "strong" groups respectively. Nuffield Combined Science, Nuffield Physics, JSSP and early Discovery seem to be exceptions. Combined Science was a pruned amalgam of courses with more social content and in the squeeze the straight scientific content won precedence. Nuffield Physics really had a research-oriented person since Rogers moved into academic physics after the project.

JSSP though written by teachers did draw on academic scientists and had PSSC and Chemistry a Structural View as reference points. Discovery was written for course outlines which were still heavily oriented towards university requirements.

However, this relationship just raises another question, Why do these two types of persons interpret the content of school science so differently?

A closer look at some of these curriculum projects does provide part of the answer to this more basic question. Any persons writing material for a science course will ask the question, What is the content of the subject I am to write for?

In the 1950's, Joel Hildebrandt, the doyen dean of Berkeley's chemistry department answered for his subject, 'Chemistry is what chemists do'. Cryptic, but profound remarks often tend to be misinterpreted and oversimplified.

A number of the curriculum projects whose material we have considered were influenced by the particular philosophical approach known as the Structure of Knowledge. Allied to the slogan of Bruner[6] about the learner of physics being like the research physicist in his laboratory, this appeared to be entirely consistent with the Hildebrandt definition. Furthermore it provided the way to put the flesh and bones on to his chemist. Hirst[7] and Phenix[8], on either side of the Atlantic developed grand schema within which the science subjects in the essentials of research type science could be credited with a special place. Schwab[9] and Zacharias[10] translated the schema for biology and physics and BSCS and PSSC took shape. Less directly, but still using this view that the essential content of a science is the substance and syntax of its pure researchers, Pimentil[11] and Strong[12] fashioned the two chemistry projects in the U.S.A. Nuffield Physics, Victorian Chemistry and JSSP of all the other projects were those most closely in touch with these sources of the structure of knowledge.

It matters little to education that other philosophers have attacked the validity of this whole philosophical approach. Its influence has been strong in education, particularly in science and will go on being so.

In his recent challenging book, Whitfield has expounded the theory in a way that will appeal to many educators and certainly extend its influence. He gives an answer to the ... st basic question secondary educators are asking about the curriculum of the school. 'The curriculum must therefore draw upon analyses of the nature of knowledge and the inherent human abilities it develops in order to determine its nature, prior to analyses of society, the learner and the learning process'.

This consequential statement from the structure of knowledge theory immediately explains why the texts in which it is an influence are so weak in social content. It is the knowledge for the knowledge sake that is the priority determinant of the content in the school. Only secondary are aspects of its relation to society, the characteristics of the learner or even how learning occurs. No wonder with vast knowledge structures like the sciences, these secondary features are not reached in texts of this persuasion.

Some chemists do do research but the great majority do not. What they do is to apply existing knowledge in situations of immediate concern to society? The structure of knowledge is not contrary to Hildebrandt's dictum, but it relates to only the minor part of it. Reflections on the discussion of the results.

If, as I have suggested, there is a clear link between the "structure of knowledge" philosophy and the spate of a-social materials for science courses, what lies behind those with a good measure of social content?

One explanation would be that their authors were not influenced by the S.O.K. either through ignorance of it or because they rejected it in favour of some other framework for their writings. As has been pointed out above, the main authors of the "strong" social texts were successful secondary school teachers. Perhaps we might argue their resulting emphases are the natural outcome of recognising that motivation is important in learning. Relevant, textual material may enhance the interest of the learner and experienced teachers are likely to be well aware of this.

However, this explanation is not the only one that can be advanced and the recent debate on the social significance of chemistry suggests that what is occurring here is a conflict of philosophies. Just because the authors of the "social" materials for science education do not draw on *the* structure of knowledge approach does not mean that they do not have a philosophical approach to the content of their subject.

A person whose philosophy is essentially pragmatic or utilitarian is very likely to perceive the content of science and its origins in a way that relates it to himself or to society more broadly. Could a Marxist write a chemistry book like CBA or a physics text like PSSC? The fact that East European texts for science are also, on occasions, sterile socially does not disprove these philosophical emphases. We would need to be sure that the authors in these socialist countries were in fact wholly integrated persons in their political philosophy and practice. Many Christian scientists are able to compartmentalise their faith and daily work almost completely. They can and do write depersonalised science while holding intensely personal, anthropomorphic religious beliefs.

However, a Christain existentialist like John Wren-Lewis[13] sees a continuous link between technology and science with the former as the real essence of the whole area. Similarly, explicitly Marxist-oriented scientists like Joseph Needham[14], Haldane[15] and Bernal[16] in their writings repeatedly inter-related science with its application in society.

If the content of science education can be determined so differently due to the various "philosophies of knowledge" held by its writers, then research in science education may be likewise affected. In other words, what sort of questions are we likely to ask in a research sense if we are essentially using the structure of knowledge as our philosophical base? How would these be different, asked by persons with a utilitarian position?

Concept attainment, learning hierarchies and TOUS type investigations are congruent with the former position and have been major interest areas in the U.S.A. and Australia where "structure of knowledge" courses have been popular. Diffusion of curricula and "supply and demand" type studies have been best developed in the U.K. where the science courses have retained a higher social (and more utilitarian?) emphasis.

At this stage in the development of research in science education, there are a number of reasons for us all to step back and critically analyse our efforts from time to time. An obvious one is the fact that much of what we do is because it can be done, and not because it is related to the most important or most basic questions. As in all fields of research, but education and other social sciences seem particularly susceptible, we must not use possibility as an automatic criterion of importance. Myron Atkin[17] at the Kiel Conference of 1970 pointed out a second reason that relates to the likelihood that curriculum evaluation becomes bounded by the terms of reference laid down by the particular curriculum being studied. This paper, springing as it does from some findings in content analysis, may provide yet another reason related to the philosophical frameworks from which we work.

REFERENCES

1. Dainton Committee's Report. Enquiry into the Flow of Candidates in Science and Technology into Higher Education HMSO (1968).

2. Storer, N.W. *The Social System of Science*, Holt, Rinehart and Winston, New York, (1966).

3. Merton, R.K. *Social Theory and Social Structure*, Chap. XV, The Free Press of Glencoe, New York, (1957).

4. Elementary Science Study in the U.S.A.
Nuffield Junior Science in the U.K.

5. Fensham, P.J. *Social Content of Chemistry Courses* Proceedings of the 1973 Conference of the Division of Chemical Education, R.A.C.I.

6. Bruner, J.S. *The Process of Education*, Harvard U.P., Cambridge, Mass., (1960).

7. Hirst, P.H. in *Philosophical Analysis and Education*, ed. R.D. Archamboult, Routledge, London, (1965).

8. Phenix, P.H. *Realms of Meaning*, McGraw Hill, New York, (1964).

9. Schwab, J.J. *Theory of Knowledge and Problems of Education*, ed. D. Vandenberg, U. of Ill. (1969).

10. Zacharias, J.R. "PSSC Into the Laboratory", *The Science Teacher, 24*, 324 (1957).

11. Pimentel, G.C., and Campbell, J.A. Foreword, *CHEM STUDY Text*, Freeman and Coy., San Francisco, (1963).

12. Strong, L.E. *J. Chem. Ed.*, *39*, 126 (1962).

13. Wren-Lewis, J. "Science, Technology and the Dilemma in Ethics, 57, 260-272, *The Student World*, W.S.C.F. Geneva, (1964).

14. Needham, J. *Science and Civilisation in Ancient China*, Cambridge U.P. (1954-65).

15. Haldane, J.B.S. *Science and Every Day Life*, Penguin Books, Middlesex (1941).

16. Bernal, J.D. *Science in History*, Penguin Books, Middlesex, (1969).

17. Atkin, J.M. "Curriculum Design: The Central Development and the Local Teacher", in I.P.N. Symposium 1970, Kiel 1971.

CONTENT ANALYSIS CRITERIA FOR ENVIRONMENTAL EDUCATION

R. D. Linke

1. Environmental Emphasis

This index represents a measure of explicit emphasis associated with the theme of anthropocentric environmental interaction, or the relationship between man and his natural (biological and physical) environment. The assessment is based on a four-point rating scale, as shown in the table below, and is determined by the degree of recurrence, or frequency distribution, of statements which incorporate some form of explicit interaction between man and his biophysical environment. This condition excludes all moral, social, economic and political issues, unless related to some aspect of the natural environment. Thus a reference to population control based only on moral or social issues (example 1) would not be classified as a statement of environmental interaction, although relation to more practical problems (example 2) may justify positive classification. In a similar way the discussion of economic and political issues (example 3) need not involve any form of environmental interaction, unless related to the production or consumption of specific resources (example 4). It should also be emphasised that ecological issues, although concerned by definition with the natural environment, must also be related in some way to man, according to the criterion of explicit anthropocentric interaction.

EXAMPLES

1. "Infanticide allows for selection of personal characteristics of the offspring such as sex and physical condition. Motivating factors can be ritual or economic. In some societies twins are considered unpropitious and one or both may be killed. Other societies consider them a highly favourable omen." (Benedict 1970: p.173)

2. "Infanticide for economic reasons seems to be closely linked with the food supply. It occurs among peoples living in very harsh environments, in very restricted environments such as small islands, or among those living in great poverty. In such societies it is usually female infants who are killed, and this factor is closely linked to other aspects of the social structure." (Benedict 1970: p.174).

3. "A high rate of investment is a necessary condition for a strong rate of economic growth, especially when capital-widening investment is as demanding of resources as it is in Australia. But a high rate of investment is not a sufficient condition for strong growth; capital has to be directed to economic and efficient uses." (McMahon 1972: p.295).

4. "The huge investments in the mining industry have yet to fully pay off in terms of national growth. The same could be said of much investment in manufacturing industry in recent years, or the large development investments the Government has undertaken — beef roads are an example." (McMahon 1972: p.294).

Frequency of Reference to Human/Environmental Interaction

Rating

0	No explicit interaction references.
1	References rare or widely dispersed.
2	Occasional or intermittent references.
3	Consistent or recurrent references.

II. Qualitative Classification of Environmental Interaction

This index reflects an overall impression of explicit attitude or approach toward the nature of human/environmental interaction, together with a measure of intensity related to the presentation of this position. The object of interaction is irrelevant to this index and may either be man, some part of the biophysical environment exclusive of man, or both, since these are in practice inextricably related, and thus the effects of interaction in the long term mutually determined.

The assessment is based on a five-point classification scheme, as shown in the table below. Although the qualitative divisions are not mutually exclusive, in that a single unit may present both positive (concordant or beneficial) and negative (conflicting or detrimental) aspects of human/environmental interaction (example 1), each of these divisions is limited to a single rating of overall intensity. These ratings, however, need not be the same, and may be used to reflect a differential emphasis on positive or negative aspects of environmental interaction. It should be emphasised, on the other hand, that a neutral rating (example 2) logically precludes any additional qualitative classification. It should also be stressed that the assessment of predominant emphasis prevents the classification of any isolated reference to concordant or conflicting interaction, unless substantiated with additional qualification or appropriately positioned to maximise the overall impression.

EXAMPLES

1. "D.D.T., used with care, and at minimum doses, has successfully controlled insects in many countries, and has usually done so without any recognisable harmful side effects. Used in ways which we now consider unwise it has done a great deal of damage to wild life." (Mellanby 1970: p.123).

2. "It is possible to treat considerable areas of woodland with D.D.T. without necessarily causing havoc to wild life, if suitable precautions are taken, minimum doses are used, and the operation is not too often repeated." (Mellanby 1970: p.124).

Classification	Intensity	Interpretation
++	Strong	Concordant or consilient relationship, expressing beneficial aspects of human/environmental interaction.
+	Moderate	
0	-	No explicit stance on qualitative relationship, or no expression of beneficial or detrimental influence associated with human/environmental interaction.
-	Moderate	Discordant or conflicting relationship, expressing detrimental aspects of human/environmental interaction.
- -	Strong	

III. Conservational Approach

This index is concerned with explicit considerations of environmental conservation, which in general terms refers to the long-term management and beneficial use of both natural and artificial resources (Barwick 1971, Wright 1970). The classification of conservational approach, which is based on three independent divisions or categories defined in the table below, is determined by an overall assessment or impression of predominant emphasis, and involves no rating of degree. Thus an isolated reference to conservational implications (example 1) would probably not produce an impression of general concern, or positive classification, unless substantiated with additional information (example 2) or appropriately positioned to maximise the overall emphasis or impact.

EXAMPLES

1. "The annual harvest (of kangaroos) should not exceed the equivalent of the natural increment, that is the number of young produced that can be expected to reach maturity." (Ratcliffe 1970: p.5).

2. "An industry based on the cropping of a wild animal population which does not accept the restraint implicit in this principle is doomed to self-destruction because it will eat into the living resource on which it depends, and usually at a progressively increased rate." (Ratcliffe 1970: p.5).

Classification	Interpretation
-	Negative or anti-conservational approach, involving explicit rejection of long-term considerations, or denial of concern for possible resource limitations and associated problems of management.
0	Indifferent or non-conservational approach, involving neglect of conservational implications and suggesting no explicit reference to long-term considerations of resource limitation and management.

\+ Positive conservational approach, involving recognition of possible resource limitations, and explicit consideration for problems of longterm management.

IV. Emotive Intensity

This index involves a subjective assessment of general emotive impact or intensity, and is based on a four-point rating scale as shown in the table below. The assessment of emotive intensity is not necessarily associated with any particular environmental reference or theme, and should instead take general account of the frequency, position and intensity of all emotive statements. Although influenced by recurrence or repetition, it should be emphasised that the assessment of overall intensity involves more than a basic frequency count of emotive words and phrases, since these, if taken out of context, may lose their emotive impact or present a false impression of strong emotionality. For example the terms "destroy", "wipe out", and "eradicate", although apparently synonymous, may all be used in a more or less emotive sense, depending on the general context in which they occur. Moreover, with respect to position, a single emotive statement is likely to be stronger in terms of general impact if placed near the start of a passage rather than inserted elsewhere since this may increase symphathetically the impact of subsequent statements, and in fact the emotive intensity of any individual statement may similarly increase the impact of others around. Thus the examples outlined in the table below are suggested only as possible indications of appropriate rating divisions, and in this sense assume some consistency of presentation style.

Rating	Emotive Intensity	Example
0	None	"It is well established that pesticides present in small amounts in water can be concentrated many-fold by aquatic organisms, including algae, and that the degree of concentration may thereby increase from link to link in the food chain." (World Health Organisation 1968: p.55).
1	Weak	"Besides being direct threats to human health, synthetic insecticides are among man's most potent tools for simplifying ecosystems. The increase in concentration of the most persistent of these compounds with each upward step in a food chain exposes the populations least able to survive poisoning to the highest concentrations." (Ehrlich and Ehrlich 1970: p.167).
2	Moderate	"We are poisoning plants on land and in the sea in various ways at an alarming rate. An estimated three thousand chemical compounds have been added to the atmosphere by man, and we are dumping as many as half a million pollutant substances into the oceans." (Ehrlich and Ehrlich 1970: p.190).
3	Strong	"The history of the recent centuries has its black passages

the slaughter of the buffalo on the Western plains, the massacre of the shore-birds by the market gunners, the near-extermination of the egrets for their plumage. Now, to these and others like them, we are adding a new chapter and a new kind of havoc — the direct killing of birds, mammals, fishes, and indeed practically every form of wildlife by chemical insecticides indiscriminately sprayed on the land." (Carson 1971, p.67).

V. Quantitative Emphasis

This index reflects the degree of reference to various forms of statistical information. The assessment is based on a four-point rating scale, as shown in table 1 below, and is determined by the degree of recurrence, or frequency distribution, of reference to numerical or statistical data. This data may either be presented in the text, or given separately in tables, diagrams, graphs or pictorial legends. Since the index is intended to measure the frequency of reference rather than the absolute number of figures, tables of multiple data should be classified comprehensively as single examples, unless further reference is made to the information throughout the text. Cases of multiple reference to the same individual item of data should also be classified as single examples, although an individual statement incorporating different numerical examples may justify multiple classification. General comparative statements such as "more" or "less" should not be classified as examples of statistical reference unless qualified by more specific quantitative terms, although these may include approximations, such as fractional estimates or rounded percentages.

The general rating of quantitative emphasis (table 1) is independent of any reference to environmental interaction, and should account for every instance of statistical information. The conditional rating, however, outlined in table 2, involves an assessment of the overall proportion of references to statistical information which are directly or explicitly associated with some aspect of the humna/environmental interaction theme. This rating is not limited by the first, provided that the general rating is positive, since a unit which refers infrequently to statistical information may still have a high proportion of environmental data. The only limiting condition for this index is therefore one of relevance to the environmental theme outlined above.

(1)	General Rating	Frequency of Reference to Statistical Information
	0	No reference to statistical data.
	1	Rare or scant reference to data.
	2	Occasional or intermittent reference to data.
	3	Consistent or recurrent reference to data.
(2)	**Environmental Rating**	**Proportion of Statistical References**
	0	0 – 25%
	1	25 – 50%
	2	50 – 75%
	3	75 – 100%

VI Pictorial Emphasis

This index reflects the degree of reference to various forms of pictorial representation. The assessment is based on a four-point rating scale, as shown in table 1 below, and is determined by the overall frequency of reference to photographs, drawings, maps, charts, graphs and other pictorial forms. Composite pictures or graphs should be classified as single examples, unless involving multiple reference in the text to different aspects of information. In this case each reference should be classified as a separate example. It should be emphasised, perhaps, that this index is independent of pictorial area or size, although the overall rating may be influenced by position or distribution. As explained for index V, the general rating of pictorial emphasis (table 1) is independent of any reference to environmental interaction, and should account for every instance of pictorial representation, while the conditional rating of proportion (outlined in table 2) involves some form of explicit association with the human environmental interaction theme. This relationship should either be explained in the text or outlined in the appropriate legend, and constitutes an essential condition for environmental classification.

(1)	General Rating	Frequency of Reference to Pictorial Representation
	0	No reference to pictorial representation.
	1	Rare or scant pictorial reference.
	2	Occasional or intermittent pictorial reference.
	3	Consistent or recurrent pictorial reference.
(2)	Environmental Rating	Proportion of Pictorial References
	0	0 – 25%
	1	25 – 50%
	2	50 – 75%
	3	75 – 100%

VII. Personal Involvement – (A) Intellectual Activities

This index reflects, for materials with an educational role, the degree of personal or student involvement initiated through various intellectual activities. The general assessment is based on a four-point rating scale, as shown in table 1, and is determined by the overall frequency of reference to all intellectual activities, irrespective of environmental relevance. These activities may include questions, directions for further reading or literary research, written assignments involving recall or rational argument, and a range of similar intellectual problems. Composite questions or activities should be classified as single examples, and rhetorical questions, for which answers are subsequently given in the text, should not be included at all in this assessment. The presentation of activities in terminal groups, while probably justifying multiple classification, need not involve separate consideration for each example, unless additional reference is made to these activities in the text. The conditional rating (outlined in table 2) involves an assessment of the overall proportion of intellectual activities explicitly associated with some aspect of the human/environmental interaction theme.

(1)	General Rating	Frequency of Reference to Intellectual Activities
	0	No reference to intellectual activities.
	1	Rare or scant reference.
	2	Occasional or intermittent reference
	3	Consistent or recurrent reference

(2)	Environmental Rating	Proportion of Intellectual Activities
	0	0 – 25%
	1	25 – 50%
	2	50 – 75%
	3	75 – 100%

VIII Personal Involvement (B) Practical Activities.

This index reflects the degree of personal involvement initiated through practical or experimental activities. The general assessment is based on a four-point rating scale, as shown in table 1, and is determined by the overall frequency of reference to all experimental activities, irrespective of environmental relevance. These activities must involve some form of physical manipulation or psychomotor skill, apart from any cognitive or intellectual component, and verbal (oral or written) assignments are specifically excluded from assessment. Composite experiments or activities should be classified as single examples, while those presented in terminal groups, although probably justifying multiple classification, need not involve separate consideration for each activity, unless additional reference is made to these activities in the text. The conditional rating (outlined in table 2) involves an assessment of the overall proportion of practical activities explicitly associated with some aspect of the human environmental interaction theme.

(1)	General Rating	Frequency of Reference to Practical Activities
	0	No reference to practical activities.
	1	Rare or scant reference.
	2	Occasional or intermittent reference.
	3	Consistent or recurrent reference.

(2)	Environmental Rating	Proportion of Practical Activities
	0	0 – 25%
	1	25 – 50%
	2	50 – 75%
	3	75 – 100%

REFERENCES

Barwick, G. *Economic Growth and the Environment*. (Occasional Publication No. 7 Australian Conservation Foundation, Melbourne, Victoria, 1971).

Benedict, B. "Population Regulation in Primitive Societies" in *Population Control* ed. A. Allison (Penguin, U.K., 1970).

Carson, R. *Silent Spring* (Penguin, U.K., 1971).

Ehrlich, P.R. and Ehrlich, A.H. *Population, Resources, Environment — Issues in Human Ecology* (Freeman & Co., San Francisco, 1970).

McMahon, W. "Investment and Economic Growth" in *Applied Economics: Readings for Australian Students* ed. K. Hancock, B. Hughes and R. Wallace (McGraw-Hill, Sydney, 1972).

Mellanby, K. *Pesticides and Pollution* (Fontana, U.K., 1970).

Ratcliffe, F. *The Commercial Hunting of Kangaroos*. (Occasional Publication No. 4 Australian Conservation Foundation, Melbourne, Victoria, 1970).

World Health Organisation. *Research into Environmental Pollution*. (Technical Report Series, No. 406, W.H.O., Geneva, 1968).

Wright, J. *Conservation as an Emerging Concept*. (Occasional Publication No. 2 Australian Conservation Foundation, Melbourne, Victoria, 1970).

TITLE . CHAPTER UNIT ....

INDEX	RATING					COMMENTS
I Environmental Emphasis	0	1	2	3		
II Qualitative Classification	- -	-	0	3	+ +	
III Conservational Approach	-	0	+			
IV Emotive Intensity	0	1	2	3		
V Quantitative Emphasis	0	1	2	3		
(1) General						
(2) Environmental						
VI Pictorial Emphasis	0	1	2	3		
(1) General						
(2) Environmental						
VII Intellectual Activities	0	1	2	3		
(1) General						
(2) Environmental						
VIII Practical Activities	0	1	2	3		
(1) General						
(2) Environmental						

ASSESSMENT BY .... DATE: ....

PART IV

RESEARCH TECHNIQUES

SOME ASPECTS OF METHODS FOR PLACING PEOPLE IN GROUPS

R. T. White

On occasion in research it may be necessary to divide a group of people on whom measurements have been made into smaller groups of more similar or homogeneous types. An example is provided by Theobald's* on-going research into the effect of teaching styles on various student outcomes. Since it is not really possible to direct teachers to use one style or another Theobald is using a number of teachers, and is classifying them into one style or another. The classification involves making audio-tapes of several lessons given by each teacher throughout the year, and then analysing their behaviour by a Flanders-type system which gives them scores on some fifty variables. The teachers are then to be split so that those whose scores on the fifty-odd variables are similar are placed in the same group. The groups will then be compared with respect to their pupil's scores on tests of various outcomes.

Although it is simple to say that people will be grouped on the basis of how similar their scores are on a set of tests, the problem of how to go about the grouping is more subtle than may first appear. Some of the subtleties, and two solutions, are described briefly below. Another introductory treatment of the purposes and problems of clustering has been given by Tatsuoka (1973: 282-284).

Measures of similarity

Figure 1 shows a hypothetical set of scores for three people. How can these scores be reduced to single indices which will

		1	2	3	4
	Joh	0	-2	6	2
Person	Robert	2	3	-4	0
	Dick	8	10	-4	4

Figure 1. Hypothetical set of scores for three people.

indicate the degree of similarity between each pair of the three? Perhaps the most obvious is the Pythagorean distance between them. In this example the value for the distance between Joh and Robert is

$$\sqrt{(0-2)^2 + (-2-3)^2 + (6-(-4))^2 + (2-0)^2} = 11.5$$

* J H Theobald. Unfinished Ph D thesis, Monash University.

The greater the distance, the less the similarity. Pythagorean distance is used as the similarity measure in the grouping technique known as H group (Veldman 1967). It can also be used in the grouping technique invented by Cattell and Coulter, known as Taxonome (Cattell and Coulter 1966). Cattell and Coulter argue, however, for the use instead of a transformation of the distance into a similarity measure known as r_p

$$r_p = \frac{E - d^2}{E + d^2}$$

Where E is the expected value of d^2 for a pair of people chosen at random. For scores which are distributed normally with standard deviation 1.0, E is equal to twice the median (50%) chi-square value for number of degrees of freedom equal to the number of scores for each person. d is the Pythagorean distance between the pair. It can be seen that when d = 0, r = 1; when d = E, r = 0, and when d approaches infinity, r approaches -1

A quite different similarity measure is the correlation between one person's set of scores and another's. This measure is used in Stephenson's (1936) clustering method, and is also known as Q technique. It suffers from the disadvantage that it overlooks differences in degree with which attributes are held. The correlation between Dick's and Robert's scores in figure 1 is, for example, 1.0, indicating that Dick and Robert are identical when they clearly are not.

Meaning of similarity

Once a measure of degree of similarity is selected, it must be determined what really is meant by "similar" and by "groups of similar people."

In defining "similar", the choice lies between definitions in continuous or dichotomous terms. If similarity is seen as a continuous thing, then the measure of similarity that has been selected will do quite well as an operational definition of similarity. However, one might want to say that there is a threshold of similarity: if two people are closer together than this threshold value, they are counted as similar; if further apart, as dissimilar, totally and absolutely, not as a matter of degree. Which choice is made is up to the researcher. Some techniques, such as H-group, are based on a continuous definition of similarity, while Cattell and Coulter's Taxonome is based on a dichotomous definition. It will be seen that the dichotomous definition introduces a further point of arbitrariness, the placing of the threshold between similar and dissimilar. The final form of the groups of people is heavily dependent on just what is accepted as similar. To expand this briefly, if too broad a threshold is specified, everyone will be found to fall in one large group, while if too narrow a one is used no one will be found to be similar at all to anyone else.

In defining what is meant by "groups of similar people", there is a choice between what may be called "clumping" and "contiguous chains". Consider the distribution of people, who have been measured on two attributes, which is shown in figure 2.

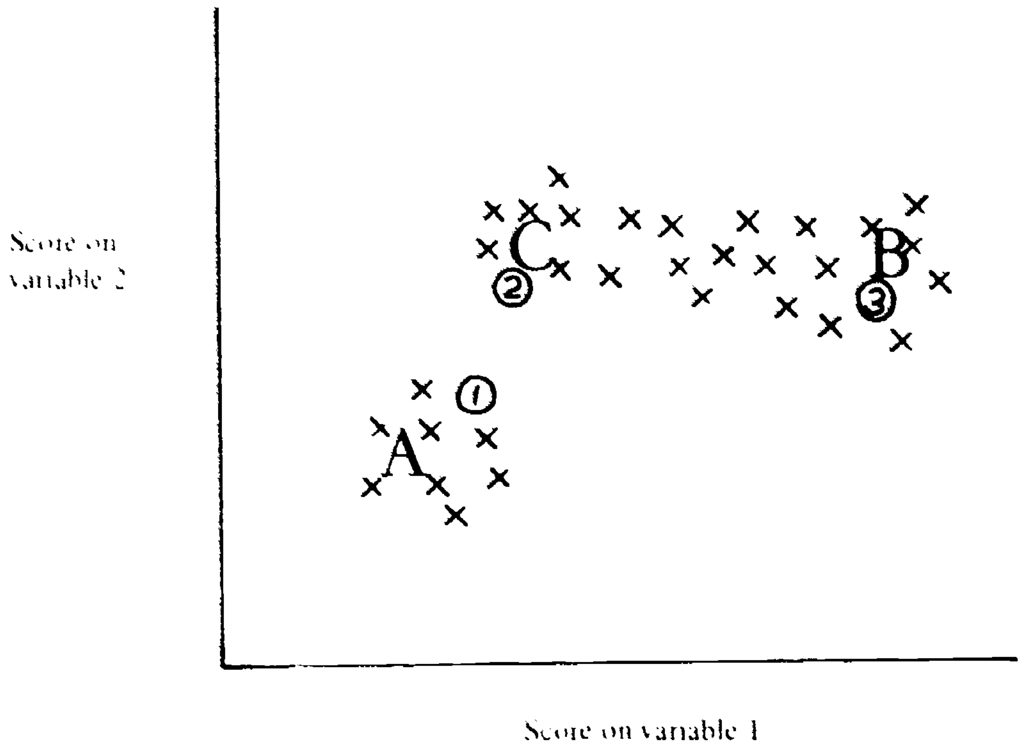


Figure 2. Hypothetical distribution of people on two variables.

In a clumping scheme it would be probable that, if the people were forced into two groups, the clusters A and C would be in one group, and B in the other. In a contiguous chain scheme A would be in one group, and clusters B and C in the other. Consider the three people marked 1, 2, and 3. The clumping scheme would put 1 and 2 together, because they have a small separation. Person 3 would not be put with 1 or 2 because it is far from them. The chaining scheme would not link 1 and 2 because there is a chasm between them, a chasm that is not bridged by a chain of "similar" people. Person 2 is connected to 3 by a chain of people, so they would be grouped together. Which definition one chooses is a matter for consideration. H group is a clumping scheme, Taxonome a contiguous chain scheme.

Methods of forming groups

Although the distinction between clumping and contiguous chain schemes was outlined in the previous section, more needs to be said about how these procedures operate to bring out the second type of difference between H group and Taxonome. This is the distinction between what may be called hierarchical, dendritic, or agglomerative methods and instantaneous or singular methods.

One form of agglomerative method is to start with all the N people in N groups of one person each. Two people are then placed in a single group, so that there are now (N-2) groups of one, and one group of two, a total of (N-1) groups. This process of merging groups, one pair at a time, is continued until everyone is in one large group of N. The criterion for which pair of groups is to be merged is another matter of choice. It could be to join the groups with the closest centroids, or it could be to join those which cause the smallest addition to the sum of squares of scores within groups, as H group does, or it could be some other criterion.

Other dendritic methods are possible. One could, for example, start with everyone in one group, ~~and cleave this~~ group into two, then each of the new groups into two, and so on until N groups of one are ~~attained. The~~ general characteristic of these methods of the first type is that they provide N solutions, for the people divided into 1, 2, . . . , N groups.

In contrast, instantaneous or singular methods give only one grouping. In Taxonome, the first step is to calculate a similarity matrix, which contains values of r_p, or Pythagorean distance, or correlation coefficient, or whatever similarity measure has been selected. This matrix is then replaced by a corresponding one which indicates which pairs of people come within the threshold of similarity. Linked chains of people are then searched for, and finally specified in terms of their members. How many such chains there are cannot be foretold — there is no guarantee that there will be two groups, or one, or N. The number can be manipulated, however, by altering the definition of the similarity threshold, or by altering the definition of the strength of a link — instead of saying that two mutually similar people constitute a link between two others, one can break links and so increase the number of groups by saying that for two people to be linked they must have at least three, or four, or five, mutually similar partners.

Summary

The distinctions between H group and Taxonome are summarized in figure 3. There are other grouping schemes, but these two were chosen to illustrate some of the techniques and problems in grouping people.

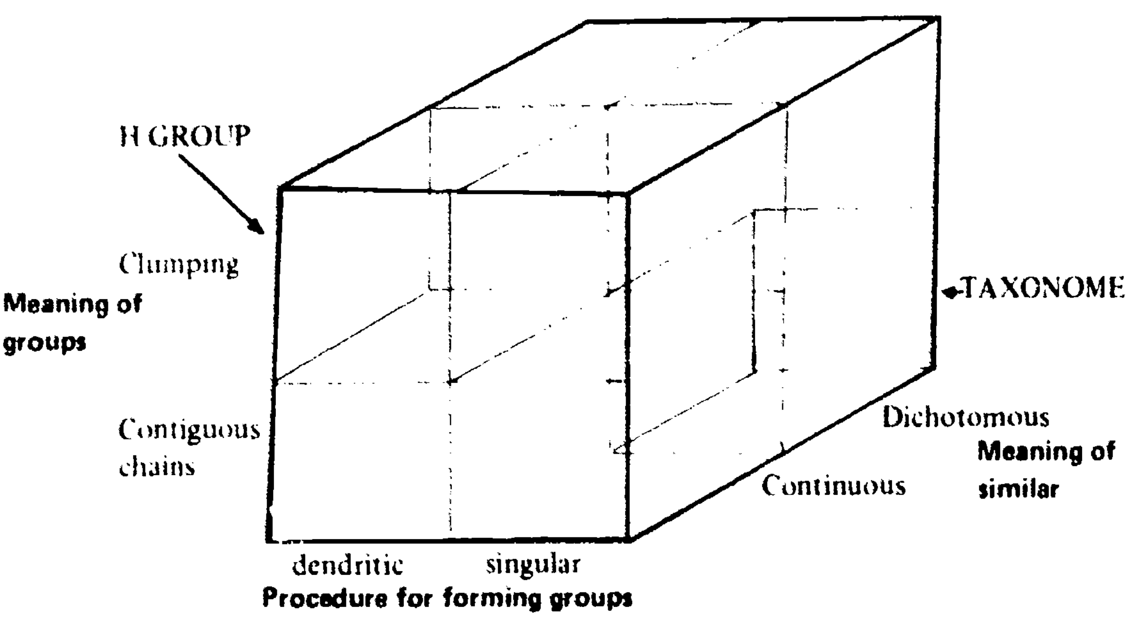


Figure 3. Types of grouping schemes.

BEST COPY AVAILABLE

It has been shown that there are several more-or-less arbitrary decisions to be taken in carrying out grouping. A measure of similarity has to be selected; one has to decide between a continuous and a dichotomous definition for similarity itself; one has to choose between a clumped and a chained meaning for groups; a dendritic or a singular solutions has to be adopted. With dendritic methods such as H group the main problem is where to stop. Which of the groupings is the best? When there are two, three, four, or how many groups? With Taxonome, the problems are at what values to set the threshold of similarity — that is, how similar is similar? — and how many people are needed to constitute a link between two other people. The presence of these problems, while they perhaps annoy the researcher who wants quick, unarguable answers, makes the process of grouping richer in interest and more fascinating than would otherwise be the case.

REFERENCES

Cattell, R.B., and Coulter, M.A. Principles of behavioural taxonomy and the mathematical basis of the taxonome computer program. *The British Journal of Mathematical and Statistical Psychology*, 1966, *19*, 237-269.

Stephenson, W. The inverted factor technique. *British Journal of Psychology*, 1936, *26*, 344-361.

Tatsuoka, M.M. Multivariate analysis in educational research. In *Review of Research in Education*, 1. F.N. Kerlinger (ed.). Itasca, Illinois: AERA, 1973.

Veldman, D.J. *Fortran Programming for the Behavioral Sciences*. New York: Holt, Rinehart and Winston, 1967.

ANALYSIS OF COVARIANCE FOR COMPARISONS OF CHANGE[1]

Barry McGaw

In educational and psychological research it is often very difficult to make valid comparisons of group performances at the end of the experiment, because the groups used are seldom equivalent beforehand. The usual procedure in such cases has been to base the comparisons on measures of change.

Techniques for analysing changes in group performance can be separated into two categories. In one category are those techniques which compare groups on the total change from initial to final status i.e. *absolute change*. In the other category are those techniques which compare groups on only that part of the final status which cannot be predicted from the initial status i.e. *unpredictable change* or residual change.

With either absolute change or unpredictable change, the change itself may be calculated from *observed scores* on pretest and posttest or from estimates of *true scores*.

Analyses of Absolute Change

The simplest measure of absolute change is the difference between the observed scores on the pretest and posttest. There are two precisely equivalent ways of analysing such change scores obtained for different groups. One way is to use an analysis of variance (ANOVA) with the change scores as the dependent variable. The hypothesis of no differences among the groups in the extent of their change can then be tested with the F-ratio for the Between groups effect.

The other way is to use a repeated measures ANOVA with the observed pretest and posttest scores themselves serving as the dependent variables. The hypothesis of no differences among the groups in the extent of their change can then be tested with the F-ratio for the groups x occasions interaction effect. The two F-ratios will be identical (Werts & Linn, 1971).

A major weakness in these analyses with observed scores is that the estimates of change are usually quite unreliable. Real changes in status are confounded with changes due to regression to the mean, a phenomenon arising from the unreliability of the measures. These difficulties can be overcome to some extent by using estimates of true scores rather than observed scores. Techniques for deriving both simple and multiple regression estimates of absolute true change are described by Cronbach and Furby (1970) and O'Connor (1970, 1971).

Analyses of Unpredictable Change

Scores on a posttest can be partitioned into two components — one linearly predictable from the pretest, the other (residual) unpredictable from the pretest. Analyses of

---

1 Summary of issues discussed in seminar at the Annual Conference of the Australian Science Education Research Association, Brisbane, May 1972. Copies of a fuller version of this paper are available from the author.

unpredictable changes may be performed in two ways.

One way is to partition each individual's posttest score as described and to use the unpredictable portions as the dependent variable in an ANOVA. A more direct way is provided by analysis of covariance (ANCOVA) with the posttest as dependent variable and the pretest as covariate. In ANCOVA, the analysis is performed on the posttest adjusted for differences in the pretest.

These two procedures, while conceptually the same, are computationally different. The sum of squares between groups in the ANOVA with unpredictable change scores will be greater than the adjusted sum of squares between groups in ANCOVA (Cochran, 1957). ANCOVA provides the better test of the hypothesis that there are no differences among the groups in the extent of their unpredictable change.

Unreliability of observed scores adversely affects analyses of unpredictable changes, just as it does those of absolute changes. Various procedures for estimating unpredictable true change are described by Cronbach and Furby (1970). Adaptations to ANCOVA, based on these estimates, are described in detail in McGaw (1972a) and are outlined in McGaw (1972b). They are summarized below.

Absolute vs Unpred' :able Change

The choice of whether to focus on absolute or unpredictable change in the comparison of groups is fundamental. In the case of unpredictable change, removal from the posttest score of that portion predictable from the pretest is equivalent to removing the pretest score itself *plus* any component of the absolute change predictable from the pretest. In the case of absolute change, of course, only the pretest score itself is removed from the posttest score.

In an experiment with groups differing in initial status, the most appropriate way to make comparisions of post-experiment performance would be to remove from the posttest scores the effects of both differences in pretest level and differences in change predictable from pretest level. Only those differences which remained would then be attributed to differences among the experimental treatments. The best procedure, therefore, is to analyse unpredictable change and ANCOVA provides the best means of doing this.

There are, however, some serious problems in the use of ANCOVA (or any method of analysing unpredictable change). The remainder of the paper describes these problems and proposals for minimizing their effects.

Problems with Analysis of Covariance

In the ANCOVA model, derivations of estimates of parameter values depend upon the assumption that the covariate is measured without error. The effect of any error in the covariate is twofold. First, it increases the estimate of the within groups sum of squares and thus gives a less powerful test of the test of significance of differences among the groups.

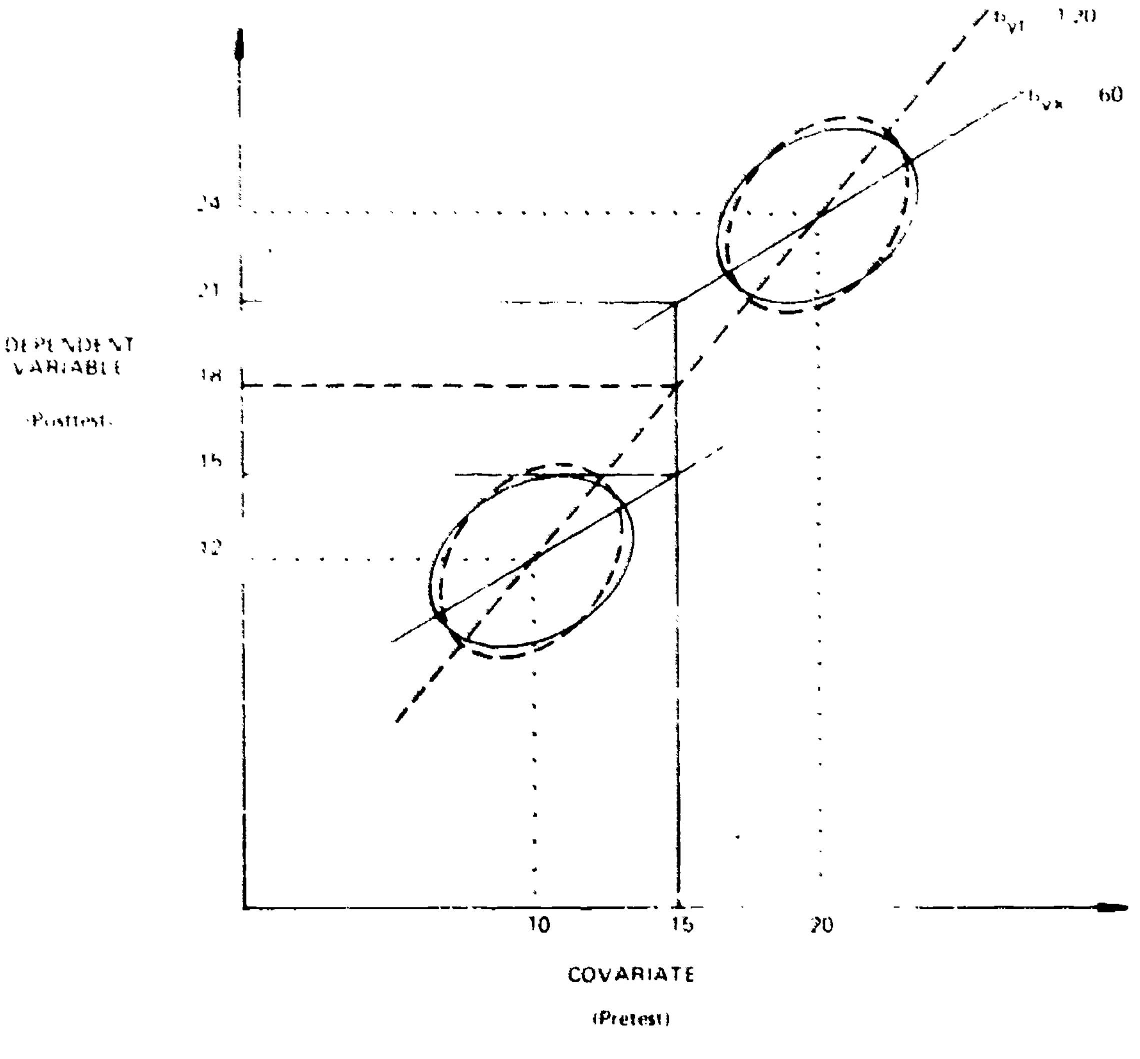


Fig. 1 Covariance Adjustments Based on Observed and True Covariate Scores

Second, it results in an underestimate of the slope of the within-groups regression line for posttest scores (Y) on pretest scores (X). If the regression slope for posttest scores (Y) on true pretest scores (T) were b_{YT} the regression slope for observed scores would be:

$$b_{YX} = b_{YT} r_{XX} .$$

To whatever extent the reliability of the pretest (r_{XX}) is less than 1.00, the regression slope for the observed scores will be an underestimate of the slope involving the true pretest scores. And it is this latter slope which is assumed to be known in making covariance adjustments.

The effect of underestimating the regression slope is an underadjustment of group means on the posttest. That is, the full effect of differences among the groups on the pretest is not removed. This is illustrated in Figure 1 in which the regression slope for true pretest scores is 1.20 but that for observed scores is only .60, since the pretest reliability is only .50. Had the slope for true scores been known the posttest means for both groups would have been adjusted to 18 and it would have been concluded that there was no difference between the groups on the posttest which was not predictable from pretest differences.

Use of the regression slope for observed scores would have produced adjusted posttest means of 17 and 21 and would have led to the erroneous conclusion that there was an unpredictable difference in posttest scores which was attributable to differences in the experimental treatments.

Adjustments to Analysis of Covariance

There are essentially two approaches to the adjustment of ANCOVA procedures to take account of unreliability of the covariate. These approaches differ in the point at which estimates of population parameters are derived. In one, estimates of true pretest scores are derived and ANCOVA is performed with them. (In fact, the estimates themselves need not be derived, since the algebraic consequences on the ANCOVA computations can be determined and the adjustments can be made directly upon them.) In the other, adjustments to ANCOVA for true covariate scores can be derived and estimates of these adjustments obtained.

Empirical comparisons (McGaw 197 , 1972b) suggest that the former approach is the superior. With this approach, the estimates of true scores may be obtained from either simple or multiple regression techniques. One readily available computer program[2] (Hall *et al*, 1972) now provides at least for simple regression estimates of true scores rather than observed scores for the covariate.

The effect of the adjustments to ANCOVA based on estimated true covariate scores is to alter the within groups sums of squares and products based on observed scores in the manner shown in Table 1. The between groups sums of squares and products are unaltered.

2 The author has installed this program on the UCCompunet network in eastern Australia and can provide details for its use to anyone interested.

TABLE 1

Within-groups Sums of Squares and Products with Estimates True Covariate Scores

	Simple Regression Estimate	Multiple Regression Estimate
Sum of Squares	$W_{tt} = r^2_{xx'} W_{xx}$	$W_{tt} = \dfrac{r^2_{xy}(1 - 2r_{xx'}) + r^2_{xx'}}{1 - r^2_{xy}} W_{xx}$
Sum of Products	$W_{ty} = r_{xx'} W_{xy}$	$W_{ty} = W_{xy}$

Estimates of Reliability

In the adjustments summarized in Table 1 an important variable is the pretest reliability, $r_{xx'}$.

The most appropriate estimate for the purpose described here is a test-retest reliability. In a case where the same test is used for pretest and posttest, a pooled within-groups correlation between pretest and posttest could serve as the reliability estimate.

Conclusions and Recommendations

From the empirical comparisons reported in McGaw (1972a) the following conclusions and recommendations can be offered.

Identical Pretest and Posttest

1. If pretest reliability > .80 and prior differences among group means < 1.25 times error variance, the observed covariate is adequate.

2. In general, simple regression estimates of true covariate scores provide sufficiently unbiased adjustments for appropriate inferences to be made.

Non-identical Pretest and Posttest

1. Unless prior differences among group means < error variance, observed covariate is inadequate, no matter how reliable.

2. Simple regression estimates of true covariate scores produce underadjustment unless prior differences among group means < 1.25 times error variance.

3. In general, multiple regression estimates of true scores provide sufficiently unbiased adjustments for appropriate inferences to be made.

REFERENCES

Cochran, W.G. Analysis of covariance: its nature and uses. *Biometrics*, 1957, *13*, 261-281.

Cronbach, L.J. and Furby, L. How we should measure "change" — or should we? *Psychological Bulletin*, 1970, *74*, 68-80. See also Errata, *ibid.*, 1970, *74*, 218.

Hall, C.E., Kornhauser, K. and Thayer, D. VARAN: A Linear model variance analysis program. *Research Memorandum 72-12*. Princeton, New Jersey: Educational Testing Service, 1972.

McGaw, B. *Measurement of Change with fallible data.* Center for Instructional Research and Curriculum Evaluation, University of Illinois at Urbana-Champaign, 1972(a).

McGaw, B. *Inferring group differences on measures of change.* Paper read at the Annual Conference of the Australian Association for Research in Education, Canberra, November 1972(b).

O'Connor, E.F. Extending classical test theory to the measurement of change. *CSE Report No. 60*. Los Angeles: Center for the Study of Evaluation, 1970.

O'Connor, E.F. Response to Cronbach and Furby's "How we should measure 'change' or should we?" *Psychological Bulletin*, 1971, 78, 159-160.

Werts, C.E. and Linn, R.L. Problems with inferring treatment effects from repeated measures. *Educational and Psychological Measurement*, 1971, *31*, 857-866.

THE CONTRIBUTORS

BRIAN CARSS is Reader in Education at the University of Queensland where he has a particular interest in the application of computers and computing technology to educational research problems.

He was educated in England and carried out all his postgraduate work at the University of Illinois. After receiving his doctorate, he worked for three years in the research laboratory of the Gulf Oil Corporation. In 1966 he returned to the University of Illinois to become Director of the ERIC ECE Clearinghouse. In 1970 he was appointed Reader in Education at the University of Queensland.

JOHN CLARKE is Lecturer in Educational Psychology at the Kelvin Grove College of Teacher Education. He has taught in Queensland secondary schools and prior to his present appointment was a Science subject master. He is a graduate in Science and Education from the University of Queensland and is at present undertaking a Master of Educational Studies at that University. His main interests are in science curriculum evaluation and science classroom dynamics.

PETER FENSHAM is currently Professor of Science Education at Monash University after an earlier career that included work in physical chemistry and the social sciences. Before coming to Monash he was Reader in Physical Chemistry at Melbourne University. At present he is engaged in innovating new patterns of science teacher education.

R.B. FLEGG is a teacher of science in New South Wales secondary schools. He recently completed the Bachelor of Science (Education) course at the University of New South Wales where his special area of interest in the honours programme was the measurement of scientific attitudes.

BARRY FRASER is a Senior Teaching Fellow in Education at Monash University. Previously he has taught in Victorian schools and has been a part-time Research Officer at ASEP. He is a graduate in science and in education and is currently completing a Ph.D. in education. His main teaching and research interests are science education, test construction and curriculum evaluation.

MAJORIE GARDNER is Professor of Science Education in the Department of Chemistry and the Department of Secondary Education at the University of Maryland. She has helped to implement the CHEM Study, to develop the Earth Science Curriculum Project (ESCP) and the new Interdisciplinary Approaches to Chemistry (IAC) programme. Experimenting with self-pacing in large lecture sections at the University level, she recently received the O Haus Award for innovation in the teaching of College chemistry.

WILLIAM HALL is Director of the Schools Council Integrated Science Project which is housed at the Centre for Science Education, London. He has had teaching experience, has been in business (as a chief editor of a publishing company), and has done educational research (his doctorate was in science education). The past four years have been spent in curriculum development.

AUSTIN HUKINS is Professor of Science Education at the University of New South Wales. After some years as a science teacher in secondary schools in New South Wales, he was appointed to Sydney Teachers College and later became head of the Department of Chemistry there. In 1963 he completed his Ph.D. at the University of Alberta with a special interest in evaluation in science education. He was appointed to the University of New South Wales in 1970 where he is Director of the Bachelor of Science (Education) course as well as Professor of Science Education.

DONALD HUTCHINGS is a Senior Lecturer at Oxford University Department of Educational Studies where he directs a research unit concerned with the academic motivation and career orientation of secondary school pupils. He has taught science in the United States, Switzerland, and in Britain. Before taking up his appointment at Oxford in 1959, he was Vice-Principal of Bournville College of Further Education. He has been a Convener of the Schools Council *Project Technology* and a Consultant to a Royal Society working party on the relevance of university physics to industrial needs.

RUSSELL LINKE is a Research Fellow in Education at Monash University and is currently conducting a national survey of environmental education in Australia. The project is supported by a grant from the Australian Advisory Committee for Research and Development in Education. He is also completing a Ph.D. in Science Education. Russell Linke came to Victoria from South Australia where he completed a B.Sc. (Hons.) in Developmental Biology at Flinders University.

LINDSAY MACKAY is a Senior Lecturer in the Faculty of Education, Monash University. His main research interests have been in the areas of evaluation of the PSSC physics course in Victoria, the evaluation of secondary schools science curricula in Papua New Guinea, cognitive preferences in science and mathematics, and the effects of optional questions on examinations.

BARRY McGAW is Head of the Research and Curriculum Branch, Department of Education, Queensland. He was for some time a science and mathematics teacher in that State and in 1967 he joined the staff of Kedron Park Teachers College. In 1968 he became a Senior Research Officer with the Research and Curriculum Branch. During

1969-1972 he completed an M.Ed. and Ph.D. at the University of Illinois, U.S.A. He also has a B.Sc. and a B.Ed. with first class honours from the University of Queensland.

SCOTT McKENZIE is presently Science Master at Wavell High School in Brisbane, and A.S.E.P. National Trials Co-ordinator for Queensland. He was a member of a team of Queensland teachers which wrote the first draft of the A.S.E.P. unit "Sticking Together". His main interest lies in the field of curriculum development in Science.

COLIN POWER Senior Lecturer in Education at the University of Queensland, has completed a Ph.D. study on the effects of communication patterns and feedback in science lessons. Dr. Power has served as a science teacher in various state high schools in Queensland, as a Research Officer for the Research and Curriculum Branch, State Education Department, Queensland, and as a Lecturer in Education, University of Queensland.

MALCOLM ROSIER is a Chief Research Officer at the Australian Council for Educational Research. Prior to joining ACER he was a science teacher in various secondary schools in Western Australia. His main task at ACER is as National Technical Officer in Australia for IEA (the International Association for Evaluation of Educational Achievement). He has been responsible for the conduct in Australia of the IEA Science Project. He is currently Secretary of the Australian Science Teachers Association.

RICHARD TISHER Reader in Education at the University of Queensland, was, for some time, a science teacher in secondary schools in New South Wales, and a lecturer in Teachers Colleges. He joined the staff of the University of Queensland in 1963. Dr. Tisher's research activities have been associated with science education, the effects of teaching strategies and teacher education.

LEO WEST is Research Scholar in Education at Monash University currently involved in Ph.D. research project on the role of prior knowledge in the learning of science. He taught for nine years in Victorian state and independent schools prior to completion of B.Sc. (Hons.).

R. T. WHITE Senior Lecturer in Education at Monash University. Dr. White has specialized in research into learning hierarchies. As well as his interest in Science Education, he has wide contact with educational research in general through his work as consultant on experimental design and statistics for his Faculty.