DOCUMENT RESUME

ED 059 890 SE 013 338

TITLE	Council of Europe Information Bulletin 3/1971.
INSTITUTION	Council of Europe, Strasbourg (France). Documentation Cente. for Education in Europe.
PUB DATE	Dec 71
NOTE	75p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Achievement Tests; *Aptitude Tests; Grading; Higher Education; *Objective Tests; Research; *Student Evaluation; Test Reliability; *Tests; Test Validity
IDENTIFIERS	Council of Europe

ABSTRACT

The major part of this bulletin consists of two studies on the place of examinations in the school system. In a long paper (28 pages) on "New Techniques for Assessment of Pupils' Work," A.D.C. Peterson discusses: (1) efforts to improve reliability by the use of objective tests, improved marking procedures, and better standardization; (2) efforts to improve validity by the use of continuous assessment and taxonomies of objectives; (3) scholastic aptitude tests; and (4) oral examinations. The second study consists of three shorter contributions relevant to the question of secondary school leaving examinations: "Examination Research: Results Thus Far and Outlook for the Future," by M. Reuchlin; "Objective Testing and Educational Assessment," by W. D. Halls; and "From Point-in-time Examination to General Assessment," by J. Capelle. The remainder of this bulletin is made up of reports of meetings of the Consultative Assembly of the Council of Europe, the Council for Cultural Cooperation, and the Committees on Higher Education and Research, General and Technical Education, Out-of-School Education and Cultural Development, and Educational Documentation and Research. (MM)

December 1971

Contents

The Information Bulletin *which is distributed free of charge three times a year in an English and a French edition, informs on the educational, cultural and scientific activities of the Council of Europe and reprints important policy documents of European interest in these fields.*

First Part

Consultative Assembly

During its twenty-first session, which took place in Strasbourg from 4th to 9th October 1971, the Consultative Assembly of the Council of Europe held an important debate on culture and education. The following reports were presented : "European Co-operation in the field of Culture and Education" by A. Borel ; "Present trends in educational reform and further prospects with a view to permanent education" by J. Capelle ; "The setting up of a tele-university" by G. Vedovato. At the close of the debate the Assembly unanimously adopted a number of resolutions and recommendations.

EUROPEAN CO-OPERATION IN THE FIELD OF CULTURE AND EDUCATION

In the framework of the Consultative Assembly's annual debate devoted to culture and education, Mr. A. Borel, Chairman of the Committee on Culture and Education, presented a report on the situation of European co-operation in this field. The report, which traditionally deals with recent developments and future prospects, emphasizes this year the importance of a more pragmatic approach to European co-operation and solutions liable to reactivate this co-operation in the short term. After recommending that a European Office of Education should be set up [Rec. 567 (1969)], the Assembly, wishing to come to an arrangement with the Committee of Ministers, agreed that the CCC should perform the functions allocated to such an office. This could bring about certain transformations, which the Assembly feels are inescapable.

The Assembly unanimously adopted the Recommendation and the Resolution which underline the main ideas of the report. Extracts of them are given below.

RECOMMENDATION 649 (1971

on European co-operation in the field of culture and education

The Assembly,

— Recalling its proposals for restructuring and reinforcing European cultural co-operation, and in particular its Recommendation 567 (1969) on "Twenty years of European cultural co-operation" ;

— Confirming that its proposals are designed to re-establish European educational and cultural co-operation on fresh foundations on the eve of the enlargement of the European Communities ;

— Conscious that with this in mind governments will find it necessary to review the terms of reference and functions of European intergovernmental organisations, and that it is therefore more important than ever to insist on the Council of Europe's special task in the field of culture and education, and in particular with regard to the definition and application of a European policy for permanent education and cultural development ;

— Increasingly concerned by the fact that in the educational field Europe is lagging behind what has been accomplished in the economic sphere because it has been unable to avail itself of a complete, integrated and coherent system of co-operation, and convinced that the desire for the widest possible cultural unity in Europe should lead to

the search for such systems not in a community restricted to a small number of countries, but in the wider framework of the States parties to the European Cultural Convention ;

— Observing that by their very nature problems concerned with education and culture cannot suitably be dealt with by a community as such, and emphasising on the other hand the admirable flexibility of the system instituted in the Council of Europe whereby a certain number of governments are able to co-operate in the framework of so-called "partial or limited" agreements with a view to carrying out priority projects over a number of years, by means of which the governments concerned are enabled to intensify their co-operation in a given field in a way which enables all the member States to benefit from the results obtained ;

— Realising that, though the establishment of a European Office of Education as advocated in Recommendation 567 (1969) must be regarded as a long-term objective, there is an immediate need to find a practical solution by conferring at once on the CCC the task of performing the functions of such an Office on an experimental basis ;

— Aware that in this case it would be essential to review if not the terms of reference at least the composition of the delegations to the CCC as well as the vital problem concerning the relations between that body and the Conference of European Ministers of Education ;

— Considering in this context that the CCC should not limit itself to mere study and research, but assume certain political responsibilities which alone would enable it to pass beyond the stage of mere international co-operation and reach that of common redefinition of national policies ;

— Believing that, in order to facilitate such a development, it is necessary to provide the system of co-operation with "political leadership" and put the CCC under the technical control of the Conference of European Ministers of Education and a similar Conference of European Ministers of Culture the establishment of which is becoming increasingly indispensable if we are to encourage a long-term policy of cultural development ;

— Recommends the Committee of Ministers :

- to call upon the Conference of European Ministers of Education :
  - — to ensure to the fullest possible extent the co-ordination from the planning stage onwards of the activities of the various international organisations concerned with the field of education ;
  - — to exercice, in accordance with Recommendation 567 (1969), a technical control over the CCC with regard to the development of education ;
- to establish a Conference of European Ministers of Culture whose principal task would be to lay down, for the guidance of the CCC, in association with representatives of any other ministries which might be concerned, the priorities for a European programme of cultural development ;
- to instruct the CCC to perform, for an experimental period of five years, the functions allocated to a European Office of Information in accordance with the letter and spirit of Recommendation 567 (1969), and for this purpose :
  - — to revise the composition of the delegations to the CCC by ensuring the predominance of the educational and cultural elements through the presence at the head of these delegations of officials from the immediate entourage of the European Ministers of Education and of the Ministers responsible for culture ;
  - — to establish a plan with a view to at least tripling over a period of five years the governmental contributions to the Cultural Fund, so as to permit that body to provide adequate finance for the harmonious expansion of a European programme for permanent education and long-term cultural development in consonance with the aims of the Council of Europe.

RESOLUTION 499 (1971)

on European co-operation in the field of culture and education

The Assembly,

— Having regard to the work of the 7th Conference of European Ministers of Education, held in Brussels from 8 - 10 June 1971, and noting with satisfaction the resolutions adopted by that Conference ;

— Noting with special satisfaction that, in accordance with the spirit of Assembly Recommendation 567 (1969), the Conference adopted a permanent statute, thus fulfilling one of the conditions essential to enable it to play to the full its role at the head of a system of European co-operation in urgent need of reform ;

— Fully approving the decisions taken by the Conference which, with a view to ensuring the continuity of its work, extended the terms of reference of its Committee of Senior Officials whose task in future will consist not merely in the preparation of future conferences, but in observing the development of the situation in Europe in the field of education, in maintaining closer contact with the international organisations concerned and in the practical implementation of resolutions of the Conference, thus in fact making this Committee an organ capable of taking technical decisions ;

— Instructs its Committee on Culture and Education to establish close contact with the Conference of European Ministers of Education, and in particular its Committee of Senior Officials ;

— Calls on its members to take all necessary steps to ensure that the Ministers of Education will be present in person at forthcoming conferences or are represented by other ministers, and in all cases by political personages capable of undertaking responsibilities on behalf of their governments.

PRESENT TRENDS IN EDUCATIONAL REFORM AND FURTHER PROSPECTS WITH A VIEW TO PERMANENT EDUCATION

In presenting his report on "Present trends in educational reform and further prospects with a view to permanent education", Rector J. Capelle, Vice-Chairman of the Committee on Culture and Education, informed the Assembly about the programme and the results of the Symposium on Basic education held in Salerno (Italy) in July 1971. He emphasized that his report endeavours to put out, on the basis of the Symposium documents, the questions affecting new trends in education. Commenting on the four topics of the Salerno meeting, the rapporteur gave detailed information on each of these : relations between parents, teachers and pupils, acquisition of means of expression at the various levels of basic education, study of the attitude towards knowledge and finally the place of technical education at the basic levels of education.

In concluding, Rector Capelle observed that each of these themes brought out another aspect of the need to rethink the traditional educational system, and consequently teacher training. He expressed his regrets that very often in the past, people were contented with small reforms which superficial as they were, frequently served to disguise a fundamentally conservative approach. The Resolution, which was unanimously adopted, is given below.

RESOLUTION 500 (1971)

on present trends in educational reform and further prospects with a view to permanent education

The Assembly,

— Recalling its Recommendation 611 (1970) and Resolution 463 (1970) on permanent education in Europe ;

— Having regard to the report by its Committee on Culture and Education on present

trends in educational reform and future prospects with a view to permanent education, and noting especially the results of the Symposium on Basic Education held in accordance with the aforementioned resolution on 28 and 29 June 1971 at Salerno (Italy) ;

— Noting that the aim of this Symposium was to study :

- in what way education, and more generally educational activities open to young people from birth to 18 years of age, should be devised so as to meet the demands of permanent education ;
- in support of this study, to put forward a set of specific measures for consideration by governments ;

— Noting in this context that the Symposium dealt with the following subjects, on which the problems of school reform in Europe are concentrated :

- the three groups of participants in the educational process : parents, teachers, pupils ;
- the acquisition of means of expression at the various levels of basic education ;
- the reappraisal of the attitude towards knowledge ;
- the place of technical education at the basic levels of education ;

— Invites the Conference of European Ministers of Education and the Council for Cultural Co-operation, in their work aimed at reforms in the school system from the viewpoint of permanent education to be proposed to the States adhering to the European Cultural Convention, to be guided by the principles and measures set out hereunder.

General principles and specific measures

In regard to the three participating groups (parents, teachers, pupils)

There is a crisis in the relationships among parents, teachers and pupils. Participation should be neither a dilution of responsibilities nor a misapprehension as to where competence lies, nor a rejection of authority. Thus it shall be organised on a realistic footing, by :

— preparing the pupils to assume a measure of responsibility according to their degree of maturity, while satisfying their need for security and guidance ;

— encouraging the "school for parents" by bringing in large-scale educational assistance and integrating this into an educational insurance scheme on the lines of existing social insurance schemes as part of a coherent system of methods designed to train children for their full responsibility as adults ;

— in an endeavour to correct the inequality of social opportunities, developing nursery schools and taking steps to bring them within closer reach of families, particularly in sparsely populated rural areas ;

— encouraging co-operation according to the age of pupils from the viewpoint of continuing education ;

— studying the mutual responsibilities and the deontology of the teaching profession ;

— making the relations between parents and teachers more reciprocal and more functional ;

— developing appropriate methods of compensation in favour of children whose family background is culturally inferior ;

— developing a critical spirit among the under-eighteens and defining the limits of protest action that represents a constructive preparation for maturity.

In regard to the acquisition of means of expression

One of the main objects of basic education is to equip the individual with means of expression so that he may then develop his personality fully and establish relations with his surroundings.

This, then, is a matter of developing normative expression (mathematics, languages, conventional drawing) and at the same time spontaneous expression (artistic, poetic) to encourage individuality.

Measures proposed :

— to stimulate artistic expression from pre-school days, primarily through games and the child's freedom of choice, and to continue such action by appropriate means at all stages so that it can be dovetailed with subsequent artistic activities when adult age is reached ;

— to introduce the practical and direct teaching of a foreign language at the pre-school stage, and to this end :

- to see that teachers are trained ;
- to define a teaching system geared to the objectives peculiar to each of the stages in compulsory education ;
- to plan and make the best use of the teaching aids made available by modern technology ;
- from the outset of compulsory schooling, to teach pupils to speak, particularly in discussions or at symposia and in front of an audience, beginning before their own class.

In regard to the reappraisal of the attitude towards knowledge

In view of the rapid increase in knowledge, its changing pattern and frequently its obsolescence, the basic question is how to approach and use knowledge. It is a matter of reshaping education, giving the methodology of *access* to knowledge priority over the *acquisition* of knowledge.

There is need for a reappraisal of the concept of "discipline" in the sense of the multi-disciplinary approach to social and to scientific thought in the contemporary world. Basic education should prepare the pupil for integration into three environments :

- the human environment, through the study of civilisation ;
- the natural environment, through knowledge of ecology ;
- the technological environment.

Measures proposed :

— to introduce into basic education practical training in the "processing" of information so as to prepare young people, through their own active participation, to "store" information or to "select" it. For this it is necessary to pass from the stage of the static library, designed for the scholarly, to the data-processing laboratory ;

— to encourage docimological research which should be developed with the aim of defining what qualities of the pupil's personality require analysis, particularly by reference to an ideal "profile" where basic qualities would be precisely defined and capable of quantitative assessment ;

— to redefine the teacher's function, which should aim at developing in the pupil a dynamic and responsible attitude towards knowledge and in face of the situations he will meet, in particular the pressures of information.

In regard to the place of technical education at the level of basic education

The aim is to give to technical training, an essential factor of economic and social progress, the position and prestige consonant with its mission.

Specific measures :

To improve the rewards for "production" duties, at present lagging behind "service" duties :

— to help persons with trade qualifications (craft workers, skilled workers, technicians) to gain easier access to jobs as representatives and to positions of responsibility ;

— to ensure that operational skills and the corresponding diplomas rank as high in public esteem as degrees testifying to academic excellence ;

— to expand technical education in two directions : better interchangeability between specialisations and better training in human terms, particularly in the mastery of expression and knowledge of social patterns ;

— to introduce technical subjects into the general education preceding initial technical education, principally in the lower secondary stage ;

— to link up the technical educational establishment and professional circles with a view to permanent education, from several viewpoints :

- by developing vocational training through combined efforts on the part of teachers and firms ;
- by arranging training periods during school studies ;
- by arranging for supervision and sponsorship of adaptation periods for those who have just left school ;
- by arranging for people already at work to undergo training for different jobs, to bring themselves up to date, or to qualify for promotion ;

— to combine political action with suitable co-operation between technical education and industry on behalf of young people who after their national service have not undertaken any further studies and have not yet secured a job.

EUROPEAN TELEVISION UNIVERSITY

The rapporteur, Mr. G. Vedovato, recalling the Order No. 308 (1970) adopted and the terms of reference given by the Assembly to work out a project of European Television University, stressed that the Institute would be acting as a medium to stimulate universities to produce television teaching programmes or else it would produce these programmes in co-ordination with them, or any other specialised institute. Its activity would be extended to all levels : pre-university, university, post-university, and would respond to the ever-growing educational needs. The rapporteur submitted to the Assembly a Recommendation and a European Television University Project conceived as a European inter-university institute for the development of multi-media distant study systems.

Council for Cultural Co-operation

The twentieth session of the CCC was held in Strasbourg from 17th to 23rd September 1971. It was attended by delegates from twenty member States, representatives of the Consultative Assembly, the Chairmen of the CCC's permanent Committees, as well as observers from UNESCO, OECD, the Commission of European Communities and the European Cultural Foundation.

After having heard the customary progress reports of each of the chairmen of the three permanent Committees and the statements by the representatives of the Consultative Assembly, the CCC examined the various items placed on its agenda and adopted its programme/budget for 1972 in its new form.

Below is given a summary of the conclusions concerning three fields, namely : European cultural co-operation, permanent education, requirements of European higher education for satellite communication services and frequency band allocations.

Twenty years of cultural co-operation — European Office of Education

Pursuing the discussion on Recommendation 567 of the Consultative Assembly on Twenty Years of European Cultural Co-operation and in the light of proposals made and decisions taken at the Seventh Conference of European Ministers of Education, the CCC adopted an Opinion for the attention of the Committee of Ministers. Excerpts of it are given below :

"Invited by the Committee of Ministers to 'study the long-term aspects of the Assembly's proposal for the creation of a European Office of Education' and to report to it thereon, the CCC has noted that the Conference of European Ministers of Education, in deciding to give itself a permanent character while maintaining its independent status, has placed great stress on the development of collaboration between the international organisations already active in the field of education in Europe, and has envisaged for the CCC important tasks involving the promotion of new and intensified forms of co-operation between the countries of Europe. Following the thought underlying Resolution No. 3 of the Brussels Conference, the CCC shares the opinion already expressed by the Committee of Ministers that it would be premature at this stage to establish a European Office of Education as a separate institution.

It has now decided to set up a Working party which will examine further the practical means whereby the functions which had been envisaged for a European Office of Education can be progressively carried out within the CCC itself.

One of the first tasks for the Working Party will be to consider the possibility of setting on foot, in selected cases, projects called the 'priority subjects' which might be supported and financed by those member governments most directly interested.

It is not intended that the Working Party should consider afresh the general operations or programmes of the CCC in the educational field, since these are already the subject of continuous study by the CCC itself. It is envisaged, however that the Working Party, subject to the results of its consideration of the principle of 'priority projects' as defined above, should provisionally select one or two such projects with a view to examining their implications and potentialities in greater depth. Projects which would offer possibilities of economies to individual member governments through increased European co-operation would merit particular attention in this context.

The Working Party will also consider, having regard to the decisions of the Brussels Conference, means of strengthening relations between the CCC and the Committee of Senior Officials responsible for the preparation and follow-up of the Conferences of European Ministers of Education."

Permanent education

The CCC was informed of the work and proposals of two meetings, one held in Paris on 7th - 9th June 1971, the other in Strasbourg on 15th - 16th September.

The September meeting attended by representatives of the three Permanent Committees and by experts, approved the report prepared under the responsibility of Mr. B. Schwartz, Project Director, on "Fundamentals for an integrated educational policy" as well as the working plan suggested by the Second Round Table on Permanent Education which met in June.

After discussion, the CCC adopted the document on the fundamentals for an integrated educational policy and decided to set up a Steering Group, the functions and working procedures of which would be :

— to select, for study and evaluation, on the basis of criteria established by the CCC, pilot experiments in progress in member States ;

— to act as a body available to the three Permanent Committees for purposes of consultation ;

— to examine, once a year, with representatives chosen from each of these Committees, the Committee programmes in the light of the concept of permanent education and to review its own work and the criteria for the selection of pilot experiments.

As regards these criteria, the participants were of the unanimous opinion that the pilot experiments should have an important bearing on the work of at least two Permanent Committees of the CCC and should exemplify the practical application of one or more of the main principles of permanent education. Among the most important of these are :

— the promotion of the process of learning throughout life, whether for vocational or non-vocational reasons ;

— the promotion of the means of continuous review of education systems, with the active participation of the teachers and with particular reference to curriculum reform ;

— the promotion of participation in the educational process by those taught.

The study and evaluation of experiments in progress in member States geared to the concept of permanent education would be the second operational phase of work in this field.

Requirements of European higher education for satellite communication services and frequency band allocations

The CCC was informed that the Committee for Higher Education and Research had approved the final report adopted by the Steering Group on "Requirements of European higher education for satellite communication services and frequency band allocations" by Mr. R. L. Jankovich, consultant expert.

The aim of this project is threefold :

— Short-term objective : to make sure that the World Administrative Conference of the ITU (International Telecommunications Union) reserved certain frequency bands for educational purposes ;

— Medium-term objective : to set out in detail, on the basis of Mr. Jankovich's scientific study, the nature of these needs and the technical means for meeting them ;

— Longer-term objective : progressively to implement the ideas and suggestions set out in the Jankovich report. This work will have to be carried out by national bodies, institutes and departments, as well as by international organisations, and in particular by the Working Party on Educational Technology.

In conclusion, the CCC decided to refer the Jankovich Report to :

— The Steering Group on educational technology for study and follow-up action ;

— the Assembly Committee on Science and Technology for its opinion ;

— the following organisations for information :

- International Telecommunications Union (ITU)
- European Space Research Organisation (ESRO)
- European Space Vehicle Launcher Development Organisation (ELDO)
- European Conference on Satellite Communications (CETS)
- European Space Conference (ESC).

Documents : CCC (71) 36 ; DECS/Inf. (71) 8.

Higher Education and Research

Strasbourg *27th - 29th October 1971*

Twenty-fourth meeting of the Committee

The autumn meeting of the Committee, attended by delegates of nineteen member States, by observers from UNESCO, the European Communities and the Consultative Assembly, was devoted, in particular, to the discussion of the future programme and working methods.

In examining the draft programme for 1973, the Committee decided to continue to concentrate its future activities around some major fields.

Two UNESCO projects were also discussed : one concerning the possible creation of a European University, and the other the setting up of a European Centre for Higher Education.

Moreover, the Committee gave its approval to the concept of creating a European Inter-University Institute for the Promotion of Distant Study Systems. It emphasised, however, that the use of the term "Tele-University" should be avoided. With regard to the founding of a European "Open University", the participants were unanimously in agreement that such a scheme was still premature.

Documents : CCC/ESR (71) 87.

Florence *30th - 31st August 1971*

The creation of a European tele-university

(Ad hoc Sub-Committee)

Parliamentarians from the Consultative Assembly's Committee on Culture and Education and university representatives from the CCC's Committee for Higher Education and Research examined together various aspects of a possible European "tele-university" to be set up in Florence. They discussed in particular the objectives and the functions, the terms

of reference, the status and the organisation, as well as the staff categories of such an institution. A summary of conclusions is printed below.

Objectives and functions

The tele-university must be neither a "super university" nor a "counter university". Its main aim would be to promote multi-media distant study systems in member States and to help the national universities to produce software packages and to set up such teaching and learning systems. Its activities will therefore cover not only television but also media like films, video cassettes, video tapes, correspondence material, programmed text books. It would help to meet the problems of student influx, to make higher education accessible to a wider public. Likewise, the teaching of outstanding specialists would be available to students of other universities. It would address students in higher education, graduates wishing to polish up their knowledge and adults who are neither students in higher education, nor graduates.

As regards functions, the "tele-university" would have the following tasks :

— Collection of information on national experiments with multi-media distant study systems and available software in higher education ; confrontation and evaluation of these experiences ;

— Supply of technical assistance by creating an exchange of existing material so that the software in one country would become available to universities in other countries ;

— Organisation of meetings of national administrators responsible for the possible introduction of multi-media systems in higher education ;

— Assistance in the preparation of multi-media software packages, e.g. by way of convening university teachers in selected disciplines in order to reach agreement on the possible content of such material ;

— Research into all aspects of multi-media distant study systems ;

— Promotion of multi-media distant study systems ;

— Organisation of training courses for university teachers to introduce them to the new methods and techniques.

Terms of reference

Apart from training courses for university teachers, the "tele-university" would not provide direct teaching. It would also refrain from producing the necessary software packages itself. With the exception of refresher courses, the content of the teaching would always be at higher education level and it would be so conceived that formal integration into a study course in higher education would always be possible.

Statuts and organisation

The "tele-university" would legally be an independent institution to be created under the auspices of the Council of Europe by way of convention open to signature to all member States of the CCC. Its bodies would be composed of a director (or a body of directors), a scientific council and an administrative council.

Staff

The staff would consist of the three following categories :

— permanent and temporary academic staff ;

— professional staff experienced in the new media ;

— technical staff.

During the meeting it was generally emphasised that the tele-university was not to be considered as a proper university, as it was not going to have students of its own, nor was it going to grant degress and diplomas. For this reason and at least by some participants, it was recommended to avoid using the term "university". Once set up, the institution, as proposed by some participants, might be called "European Inter-University Institute for Tele-Teaching" or a variation of this title.

Document : CCC/ESR (71) 62.

Strasbourg *4th - 5th November 1971*

Mobility of higher education staff and research workers

(Meeting of experts)

Greater mobility of university staff and research workers is of vital importance for progress in research and for the restructuring of higher education at European level.

On the basis of reports and documents presented to the meeting, the participants from nine member States examined the present situation and tried to distinguish the priority needs in this field.

The most important factors of a concerted policy on mobility were defined by Mr. H. Lesguillons, President of the Association "Europe Université". Linking closely the theme of mobility with structural reforms and current trends, he stressed both in his report and in his statement the harmonisation not only of initiatives but also of university regulations and career structures, as well as the removal of legal and statutory obstacles. Finally, he pointed out that four types of stimulus should be developed :

— the liberalisation and systematic diffusion of information ;
— the extension of the right of teachers and research workers to permanent training ;
— the setting up of machinery for equivalences ;
— the development of facilities to promote integration into the host country of foreign academic staff.

After discussion, participants agreed that the national policies, pursued in recent years, have gradually removed some of the main obstacles to mobility. However, it is still too early to talk about their complete removal in the near future.

On the other hand, abolition of the legal requirement whereby teaching or research posts in higher education must be held by nationals of the country concerned could have little practical effect, if State regulations require national diplomas for access to the teaching profession.

In the interest of mobility within Europe, it is also necessary to remove the differences between the national structures as well as between staff structures of higher education and to break down compartmentalisation.

Furthermore, the meeting decided that highly specialised seminars in the natural sciences, such as the EUCHEM, EUREMECH, etc. Conferences should continue to be organised.

After having examined the present situation, the meeting dealt with its future work programme. It was aware of the fact that complete freedom of movement within Europe for university staff and research workers could not be reached merely through changes in the legal requirements but that it presupposed a long slow evolution. It was felt necessary to proceed gradually by means of short and long-term stages.

Lastly, concrete proposals were put forward for further action. The following priority

areas were chosen : student mobility, short-term mobility of staff and improvement of the systematic diffusion of information.

As regards a European status for staff in higher education and research, as envisaged in Resolution No. 2 adopted by the Seventh Conference of European Ministers of Education, the participants recommended the definition of certain basic principles. It was felt that a European partial agreement between member States with comparable higher education systems might contribute to the formulation of such principles.

Documents : CCC/ESR (70) 18 ; 19 ; 28.
CCC/ESR (71) 11 ; 47 rev. ; 84.

Strasbourg *9th - 10th November 1971*

Ethics of science

(Meeting of experts)

Experts from thirteen member States together with observers from UNESCO discussed various problems connected with the responsibility of scientists.

Advances in science and the applications of science have given rise to a host of problems, with ethical and moral implications, in which scientists must feel specially involved. These matters are of concern to many other people and in particular to politicians, doctors, religious leaders and educators. The participants were aware that no particular group is likely to find solutions alone, and see value in a common approach. There is a need for continuing interaction among these groups, interaction that will involve new forms of co-operation and some changes of attitude.

The role of scientists may often be not to offer a solution but rather to provide "a warning of risks and proclamation of benefits and discussion of quandaries".

Nevertheless, the present attitudes of scientists need reassessment, particularly the norm and value commitments implicit in their activities. The group was impressed by the suggestion that a code of ethics should be accepted by scientists — similar to that adopted by the medical profession.

The group then turned to problems of education and particularly to the introduction in the education of scientists of a better understanding of their role in society. This work will extend over several disciplines including social and behavioural sciences and will call for the co-operation of several groups of specialists such as sociologists. A problem-oriented treatment is likely to be the most effective. An attempt should be made to introduce studies of this kind into university curricula ; this will provide a new field for university research.

The following recommendations were made :

— The Committee for Higher Education and Research shouldadvise its members to take such steps as are necessary to ensure that these proposals are considered by groups in participating countries. It would be important that recommendations from individual countries should come back to the Committee for Higher Education and Research for further consideration by this group or its successor.

— Further study should be made of the many-sided problems that face politicians and particularly of the mechanisms that should be developed to aid the co-operation of scientists and politicians and others in tackling these problems. In some countries this will mean bringing together parliamentarians and scientists ; in others it will involve strengthening and widening existing arrangements.

— Some group should be asked to do the preparatory work associated with the formulation of a code of ethics for scientists.

Documents : CCC/ESR (71) 53 ; 59 ; 64 ; 65 ; 66 ; 68 ; 69 ; 75 ; 77 ; 80 ; 81 ; 91.

General and Technical Education

Strasbourg *25th - 29th October 1971*

Tenth meeting of the Committee

The Committee held its meeting under the chairmanship of Mr. J. de Bruyn (Netherlands). It was attended by delegates of twenty member States, representatives of the Consultative Assembly, together with observers from the Commission of European Communities and the European Schools' Day.

The Committee discussed the various items on its agenda, in particular : structure and organisation of basic education ; teachers ; curricula as well as past, present and future activities and conclusions of important meetings which took place during the period of 1970-71.

In examining its future activities and in approving the concentration of the programme around a limited number of priority objectives, the Committee agreed in principle on the general approach of the programme, which is evenly based on five fields of equal importance :

— structure and organisation of education ;

— teachers ;

— curricula ;

— media and methods ;

— assessment and guidance.

As regards the methods and the planning of the programme, the Committee decided to set up co-ordinating groups for the following sectors : pre-school education and primary education ; secondary education ; technical and vocational education ; curricula ; assessment and guidance.

Furthermore, the Committee discussed the documentation to be presented to the next Conference of European Ministers of Education, which will be held in Switzerland in 1973 and will have as the main theme : "The needs of the 16-19 age group, both in full-time and part-time education".

Document : CCC/EGT (71) 47.

Vienna *21st - 25th June 1971*

Road safety education in schools
(Conference)

The Second Conference of Governmental Experts on Road Safety Education in Schools, organised jointly by the Council of Europe and the European Conference of Ministers of Transport (ECMT) in co-operation with the Austrian Federal Ministry of foreign Affairs, was also attended by observers from Austriaustrian Government. Delegates from twenty States parties to the European Cultural Convention and twenty-seven member States of the European Conference of Ministers of Transport took part in the Conference, which was also attended by observers from Austria and representatives from OECD, the European Communities, the United Nations, the IFSPO (International Federation of Senior

Police Officers), the OTA (World Touring and Automobile Organisation), the PRI (International Prevention of Road Accidents), and the IFP (International Federation of Pedestrians).

The Conference emphasised the importance of road safety education for children from the age of two. It pointed out that the present situation was highly disturbing : statistics on the number of children killed or injured in road accidents showed that the casualty rate had risen faster amongst young people than in the population as a whole. There was therefore an urgent need for action by goverments and local education authorities as well as by parents and teachers.

The Conference requested the member States of the Council of Europe and the European Conference of Ministers of Transport to increase their expenditure on road safety arrangements and asked the two organisations to urge that the necessary political decisions be taken to this end.

The two main themes of the Conference were "The education of children in road safety" and "The training of teachers for road safety education". Reports were submitted on each theme.

Road safety education for children

Research into road safety education has shown that children behave very differently from adults ; hence the need to adapt children's environment according to the various psychological and physical factors which condition their road behaviour.

Alongside the various measures which need to be taken by national authorities to deal with the problems raised by children (town planning, layout of roads, revision of highway codes, attention to the design of vehicles and school buildings, etc.), parents and teachers must make every effort to provide children with road safety education that is more effective and better suited to the different stages of their development.

The objectives of road safety education and its place in the curriculum were precisely defined by the Conference. It was agreed that road safety education should be dispensed as a compulsory subject, systematically and continuously in kindergartens and in primary and secondary schools. To obtain its full educational value, it should not be treated as an isolated element, but should be fully integrated with the curriculum, being linked up in particular with technical subjects, natural science, ethics, social sciences, physical education and hygiene.

The aim of road safety education should be to make children behave responsibly as both pedestrians and vehicle-users. At least twenty hours should be set aside for road safety teaching every school year, the length of a lesson depending on the class.

School crossing patrols are an excellent device for substantially improving the safety of children, as well as of adults, on their way to and from school. In view of the very good results achieved in countries which have already instituted this arrangement, the Conference adopted a considerable number of basic principles on the subject and recommended that they be applied in all countries.

Training of teachers for road safety education

The Conference unanimously agreed that parents were primarily responsible for the safety of their children on roads but that teachers should be required to co-operate with parents, the police and others in a safety campaign.

Teachers should, it was felt, be given thorough training in road safety education, including the relevant aspects of child psychology. Teachers already in service should be provided with introductory courses in the subject and kept in touch with the improvements that are constantly being made to road safety promotion methods.

International seminars for teachers would also be highly desirable.

Teachers should establish close co-operation with the various authorities and groups concerned with road safety, such as the police, motoring organisations and pedestrians' associations, both national and international.

Also, the results of research into road safety education should be made available to teachers. At international level this research should be carried out in co-operation with OECD ; at the same level national efforts should be co-ordinated, information pooled and priority subjects selected.

In conclusion, the Conference stressed the importance of international action to ensure the continuation of work on road safety education in schools. For this purpose, it called on the Council of Europe and the European Conference of Ministers of Transport to set up an ad hoc committee of educational and road safety experts.

Its terms of reference would be to follow up and co-ordinate the application of the Conference's proposals and recommendations, and it would also serve as an appropriate forum for the exchange of experience on all sectors of road safety education, including research.

Documents : CCC/EGT (71) 13 ;
EC/Conference (71) 14 and 15 + Addendum ;
Conference (71) 2, 3 and 4.

Palma de Mallorca

21st - 26th June 1971

The contribution of audio-visual media to the training and further training of teachers

(Symposium)

The Symposium, attended by delegates from seventeen member States as well as observers from national and international organisations, was organised by the Spanish Government.

Its aims were :

- to sum up the experiments carried out by the experts of the Committee for General and Technical Education on the contribution of audio-visual media to the training and further training of teachers ;
- to define the present arrangements for the production, distribution and use of such aids for this purpose, and also the methods used for promoting their use and assessing their effectiveness ;
- to identify the main trends in this field and establish a programme for European co-operation.

In recent years, teacher trainers have come to realise that audio-visual techniques afford them new information, learning and practice possibilities. Research has made it possible to establish the broad lines of a methodology for the use of audio-visual aids in this field. The methods vary according to the type of training and the type of trainee. On the one hand, there is the need to improve the training of future teachers, and hence of school methods and techniques. On the other hand, provision has to be made for in-service training, refresher training and re-training. In both cases audio-visual techniques have proved effective, provided they are used for their proper purpose and in the proper manner.

Conclusions

The great advantage of audio-visual techniques in teacher training is that, intelligently combined, they can at the same time help to reach a mass audience and provide sophisti-

cated analysis instruments for group-work and self-teaching. The active encouragement given to the setting up of learning resources centres in many training establishments will greatly assist in the rational incorporation of new methods and techniques both in class teaching in such establishments and in separate or complementary self-teaching systems.

Careful and systematic combination of audio-visual aids can help further the cause of education on three fronts : the proper training of future teachers, the accelerated training of teachers in sectors where there is a staff shortage because of the school population explosion and lastly in-service and refresher training. Furthermore, there is reason to expect that teachers trained by audio-visual methods will naturally tend to use them in their own teaching. Having been taught by them, they will have practical knowledge and personal experience of their strong and weak points.

With the aid of audio-visual media it is now possible to :

— manipulate time by recording the sounds of all kinds of pedagogical situations, repeating them at will and watching or listening to them individually or in groups ;

— provide a pedagogical mirror in which the teacher can see himself at work and so criticise and correct his performance ;

— make diachronic comparisons in order to measure better the students' progress and the effectiveness of their training ;

— create test situations and teach the students how to deal with them.

Future teachers are capable of a better performance if they are so motivated and activated as to be themselves involved in the educating process. Learning situations should be highly individualised as there are considerable differences in the experience, ability and knowledge of the students.

Audio-visual media help to improve communications between teacher and pupil, as together they learn a new language, a new mode of expression designed to modify pupils' behaviour and overcome their feeling that school and life are two different things.

As regards in-service and refresher training, it is important for courses to allow opportunities to criticise and to make changes. In the case of practising teachers, audio-visual media can be used for a variety of purposes : to inform, to arouse awareness, to change behaviour patterns and attitudes, to bring teaching up to date, to encourage innovation. It seems to be essential not only that all refresher training techniques be used in combination but also that they be co-ordinated in a comprehensive system (television teaching, attendance at courses, conferences, seminars, correspondence courses).

As for the use of audio-visual media at home, it should be noted that it is not very effective, unless followed up by group discussion. This group use is more effective, since discussion automatically ensues, but it raises other problems : if organised during school hours, the pupils may suffer ; if organised outside school hours, the teachers have to work over-time.

As regards the production of material, four levels must be distinguished : local (training centres), regional, national and international. At each of these levels production meets certain types of need and displays a certain complexity and a certain degree of technical perfection. There is a place — but a strictly limited one — for school production : its function is to satisfy the requirements of perhaps a single class, to personalise professional production, as it were, with made-to-measure material or supplementary detail. The function claim to compete with commercial mveil the mysteries of audio-visual techniques and to train the students in non-verbal expression. In no circumstances can local production claim to compete with commercial material. Regional, national and international production must be adapted to the available production facilities, users' needs and general objectives of education. At these three levels production should concentrate more on multi-media material rather than on isolated documents.

The participants in the Symposium agreed that international exchanges and co-production were essential.

Recommendations

On the basis of these conclusions, the delegates approved the following recommendations. The Council of Europe should :

— set up a working party to investigate information, documentation and research in teacher training and further training, and the production and distribution of suitable audio-visual material ;

— compile a glossary of terms used in connection with the training and further training of teachers ;

— promote and co-ordinate the use of audio-visual media in that field ;

— study the problem of copyright in relation to the increasing possibilities of reproduction ;

— take stock of the situation in member Sta with regard to the use of learning laboratories ;

— encourage the various countries to instal closed circuits and organise a symposium of specialists to work out methods for the training of users ;

— invite member States to study the problem of standardising audio-visual equipment, with particular reference to the international compatibility of video-tape recorders ;

— facilitate access by training establishments to the archives of radio and television organisations.

Documents : CCC/EGT (71) 9 ; 29 ;
DECS/EGT (71) 29 ; 39 ; 44.

Vienna *22nd - 23rd June 1971*

Films on road safety education in schools

(Meeting of experts)

Every year the Committee for General and Technical Education arranges for a small group of experts to view and select films and other audio-visual material in conjunction with one of the activities in the Council of Europe's programme. This year, in connection with the Second Conference of Governmental Experts on Road Safety Education in Schools, material relating to that subject was viewed and selected by a group of experts from six member States.

Fourteen countries were represented by the material, which included forty-four films. In making their selection, the experts had regard to the material's educational value, technical quality and European nature (i.e. its suitability for international exchanges). They judged the material as a whole and, rather than select best items, chose some examples of current trends in the production of educational films. They accordingly distinguished between films for the classroom (8 mm, usually short), motivational or introductory films (16 mm lasting 15-20 minutes) and films equally suitable for use in schools, on television and in adult education (35 or 16 mm lasting 30 minutes or so).

The following selections were made :

Category one : Material for classroom teaching

Films :

— " Jeux et circulation " (France)

— "Chercher l'erreur n° 4" (France)

— "Joupi n° 2" (Belgium)

— "FWU-Information : Pamfi Medienkombination" (Federal Republic of Germany).

Sets of slides :

— "Slides on traffic education" (Spain)

— "Trafik Güveni" (Turkey).

Category two : Motivational or introductory films

— "Der Radfahrer" (Austria)

— "Zoo Logic" (Ireland).

Category three : Films for schools, television and adult education

— "Mit voller Wucht" (Federal Republic of Germany)

— "Peut-il s'arrêter pile" (France).

Technical details of this material are to be found in the catalogue, document DECS/EGT (71) 67. The material selected was also shown to the participants in the Second Conference of Governmental Experts on Road Safety Education in Schools.

There is a possibility of grants being provided to enable the films to be dubbed in the Council of Europe's official languages.

Documents : DECS/EGT (71) 67 ;
CCC/EGT (71) 28.

Brussels *4th - 8th October 1971*

Creativity and artistic activities in school
(Symposium)

The Symposium was devoted to the discussion of questions relating to the development of creative powers among pupils in primary and secondary schools, in particular in those subjects which offer special possibilities for creative expression : drawing, painting, the plastic arts, music, dance, mime and drama. It was attended by delegates from eighteen member States of the CCC, teachers and inspectors in all branches of the arts, teacher trainers as well as university lecturers and professors.

The programme of the Symposium included ntroductory speeches, lectures, direct contacts with pupils and teahcers, visits to schools and teacher training colleges. Participants were also shown films and slides on creative work in schools.
In a final general statement, the Symposium :

— Emphasised that creativity was one of the most decisive elements in the process of the development of the individual. It is essential to foster creativity in school and adult life and to encourage every individual to develop his creative potential to the full in a democratic social context ;

— Stressed that it was indispensable to develop and perfect techniques for the promotion of creativity in all subjects in the curriculum ;

— Recommended that all teachers should be made aware of the importance of creativity. They should be provided with the appropriate training and equipment for the successful promotion of creativity. Account should be taken of the necessity of incorporating conditions favourable to creativity in the building of schools ;

— Claimed that the arts (language, music, drama, movement, and the visual and plastic arts) had a unique contribution to make in the development of creativity and that, consequently, they should be given a more central place in the school curriculum than in the past. They should be formally recognised as a fundamental element in education.

In their recommendations to the Council of Europe, the participants proposed that the Organisation should :

— Study in detail the full implications for the school of an education based on creativity. Particular attention should be paid to curriculum development, teacher training, methodology, evalution and assessment, resources for learning and the design of school buildings ;

— Examine the application to education of the techniques for promoting creativity already in use in industry and scientific research ;

— Collect and distribute to curriculum development centres in the member States of the CCC, information and documentation on projects aimed at stimulating creativity in all subjects in the curriculum ;

— Organise a symposium on methods, content and trends in the basic training of teachers of the arts ;

— Prepare and distribute a survey of in-service training facilities for teachers of the arts in the member States of the CCC ;

— Prepare and distribute a study on methods of developing creativity in artistic subjects in the member States of the CCC.

Documents : DECS/EGT (71) 82 ; 90 ; 95 ; 101 ; 102 ; 120 ;
CCC/EGT (71) 15 ; 45.

Venice *11th - 16th October 1971*

Pre-school education — Aims, methods and problems
(Symposium)

This Symposium, organised by the Italian Government under the auspices of the Council of Europe, was attended by delegates from member States and observers from UNESCO and the European Communities.

The purpose of the Symposium was to examine the aims, forms and content of pre-school education.

Conclusions and recommendations

Pre-school education has three main functions : education, compensation and therapy, and detection.

Its educational role involves not only the child but also its parents and its background as a whole. The psycho-analytical school has stressed the decisive importance of the first years of life for future psychological development, and recognition of the deep impact that its child's first experiences are likely to have on its personality has been a vital factor in drawing attention to the importance of pre-school education from birth onwards.

Although the compensatory and therapeutic role of pre-school education has only been understood for a few decades, many studies have already shown divergences in the development of children living in environments that differ economically and culturally.

At the pre-school stage particularly, the well-trained teacher is able to play a major part in detecting backwardness in young children.

These general factors, which were discussed in detail by the working groups, formed the basis for a number of recommendations, the most important of which are set out below :

— Governments ought to recognise the importance of pre-school education both for the individual development of each child and for the general good of the community. All children, of no matter what social class, should be given the opportunity of attending pre-school establishments, by the age of three at the latest, and consequently it will be necessary to set up and expand such establishments.

— Whenever local conditions make it possible, pre-school education should be brought under the authority of a single ministry grouping all educational, administrative and social services.

— Pre-school education should be ~~accepted~~ accepted as an independent branch, without becoming a preserve cut off from all other forms of education ; its autonomy should be acknowledged by setting up a team of specifically qualified inspectors.

— Pre-school ~~education~~ establishments should ~~be accepted nsery~~ make nursery facilities available for adequate periods outside school hours so that parents may rest assured that their children are safe when they are unable, for valid reasons, to look after them themselves.

— Pre-school teachers should have the same educational standard as that required of teachers at the elementary level and they should enjoy the same professional status and salary conditions. Student teachers should be capable of helping to educate parents and be introduced to group discussion and behaviour techniques and in the problems of group dynamics so as to improve their relations with parents.

— Member States should ensure that children are prepared for the transition from a pre-school establishment to the primary school during the last year of pre-school education (visits, meetings, . . .).

— Research should be carried out into specific aspects, e.g. vocabulary development, and it should have a multi-disciplinary character.

— A special meeting should be arranged to discuss the possibilities, advisability and ways of preparing children to learn reading, writing and arithmetic at pre-school level, in the light of modern scientific findings.

Document : CCC/EGT (71) 46.

Out-of-School Education and Cultural Development

Strasbourg *18th - 22nd October 1971*

Second meeting of the Committee

The activities of the Committee composed of two main parts, the educational and the cultural sections, were discussed by delegates from eighteen member States. Observers from UNESCO, and the European Communities also attended the meeting chaired by Mr. M. Hicter (Belgium).

In connection with concentrating the CCC programme around a limited number of main themes, the Committee noted that an important work had since been done in this field. In examining the activities of the period 1970-71 and the draft programme for 1973, it acknowledged the methodical development and the gradual completion of projects in the two main fields of its programme.

STRUCTURES AND ORGANISATION OF THE EDUCATION SYSTEM

Permanent education — plan for co-ordinating and evaluation projects

The conclusions of the meeting of the representatives of the three Committees and those of the CCC were communicated to the Committee. In supporting these decisions, it considered that the setting up of a Steering Group on permanent education and its future work would make a two-way feedback possible : revision of the concept based on the results of pilot-experiments and better orientation of such experiments at national level. This could inaugurate a new form of European co-operation with the emphasis no longer on comparisons between national experiences but on a comparison of the experiments in the light of the common concept.

Organisation and future structures of adult education

The present situation and possible developments in adult education as well as the legislation and planning in this field were discussed by the Committee. After having been informed about the conclusions of the Rüschlikon Symposium and the proposals of the group of experts responsible for working out a European unit/credit system for the teaching of modern languages to adults, the Committee approved the proposed plan of work.

Educational technology — means and methods

The Committee took note of the formation of the Steering Group on educational technology, which is concerned not only with the training and retraining of adults but also with all the educational techniques that are innovating education. This Group is also required to take over or reorientate the on-going work of the Steering Group on new types of Out-of-school education whose mandate was out at the end of 1971.

Activities concerning the documentation, the studies and the co-production projects in this field is in progress. The Committee showed particular interest in European co-production projects for multi-media programmes.

CULTURAL DEVELOPMENT

Management of cultural affairs

The need for governments to have a cultural policy next to economic and social policies was explained by Mr. A. Girard (France), Project Director. He also reported on the state of the work and indicated possible or desirable issues. The Committee gave its approval to the various plans and propositions.

As regard the experimental study of cultural development in European towns, the Committee recommended the CCC to seek via the Committee of Ministers the support to the project of the governments of the eleven countries in which the towns are situated. It was regarded as very desirable that these towns have at their disposal the necessary financial resources to enable them fully to participate in this European project.

Cultural enrichment

Mr. A. J. Simpson (United Kingdom), Project Director, explained the general purposes and trends of the programme which consists of two parts : research and experiment.

The varied and complex activities in this field bear on :

— Socio-cultural facilities (facilities and innovations ; animation methods ; European system for the exchange of information ; training and status of cultural managers) ;
— New audio-visual media ;
— Contents of cultural advancement (programmes) ;
— Other cultural activities not comprised in the new programme (European art exhibitions in their previous form ; film week ; cultural identity card).

Sport for All

The work and the results of various meetings during the period 1970-71 were approved by the Committee. It was suggested that the CCC give its support to the Consultative Assembly proposal to prepare a draft European "Sport for All" Charter and invite the European Conference of Local Authorities to study ways and means of closer co-operation in this field.

Youth

The Committee was informed about various activities in this field, namely research into youth questions, the European Youth Centre and the European Youth Foundation.

European Youth Centre

The statutes being adopted, the Advisory Committee having already met, the courses could in principle be organised in the building of the Centre. The budget of the Centre which is now a subsidiary of the general budget of the Council of Europe is no longer dependent on that of the CCC. Its Governing Board will meet in December 1971.

Three intensive European language courses and eight information courses were proposed for 1972.

Document : CCC/EES (71) 130.

Strasbourg *29th - 30th June 1971*

European co-operation in Sport for All

(Co-ordinating Group)

The aim of this meeting, attended by six governmental and non-governmental experts, was to study the latest developments in Sport for All and to make proposals as regards objectives and contents of European Co-operation in this field.

After having examined the different suggestions put forward to set up a rational planning system, the participants agreed on the following guidelines :

- the needs and the proposals as expressed periodically by governments and NGOs should be the starting point of a planning system ;
- the selected priorities should be integrated into a flexible overall long-term plan ;
- within the framework of the long-term plan, short- and medium-term working plans should be developed.

Conclusions

In summarising its discussions of the previous and the present meetings, the group out-

lined five principles for a European co-operation system, aiming to develop Sport for All. These are :

— Close co-operation between the public and the private section is necessary at European, national, regional and local level ;

— The initiative of the Consultative Assembly to draft a European Charter on Sport for All is of the greatest importance : any common action requires a consensus on certain common principles ;

— A certain amount of common planning at European level is necessary in order to promote the best possible development of Sport for All : the establishment of a long-term plan is therefore a top priority ;

— A certain degree of division of labour between member countries is necessary in order to implement a common European policy under a common plan ;

— The Clearing-House is an important instrument both for the planning and implementation of such a common policy.

Document : CCC/EES (71) 83.

La Chaux-de-Fonds
(Switzerland)

30th September - 1st October 1971

Experimental study of the cultural development of European towns

(Meeting of experts)

Representatives from the ten towns taking part in this study, whose main aim is to awaken local authorities to the difficulties facing municipalities in regard to cultural policy, were able for the first time to consider jointly the various aspects involved in implementation of this Council of Europe project. They discussed the most effective means of inculcating a logical and forward-looking approach but were forced to concede that the definition of the ultimate aims and objectives of the project was far from easy.

Each town made a statement on the basic features of its current cultural policy.

The coherence and effectiveness of the project would largely depend on three factors : methods of observing cultural life would have to be rationalised, a co-ordinating body set up and the results of the experiment analysed.

To publicise the project and obtain a more exact idea of its objectives, it was agreed that the Secretariat should commission a representative from each town to draw up a report along the following guidelines :

— An *introduction* summarising the general socio-economic features of the town, and possibly of the urban area or region, its cultural policy to date and its current position with regard to the Council of Europe project.

— A *definition of the programme,* setting out the general aims of local cultural policy, its future policy and definition of possible guide-lines for the fields selected, at the three levels specified in the Council of Europe memoranda : final objectives, intermediate objectives and resources.

— An analysis of the *content of the programme,* in particular financing systems and sociological studies, with special emphasis on the cultural aspirations of the various social categories, the relations between cultural life and the general activities of the population and the results of the experiment.

— Another section of the report, on the *implementation of the programme,* was to deal

with the bodies responsible for putting the project into effect, administrative, technical and financial resources and the part (active/passive) to be played by the population.

At present the following towns have been selected for the project: Annecy (France), Apeldoorn (Netherlands), Bologna (Italy), Exeter (United Kingdom), Krems (Austria), La Chaux-de-Fonds (Switzerland), Namur (Belgium), Örebro (Sweden), Stavanger (Norway) and Turnhout (Belgium).

Documents : CCC/EES (71) 64 ; 107 ; 112 ; 124.

Strasbourg *30th September - 1st October 1971*

Adult language learning — A European unit-credit system

(Meeting of experts)

As a follow-up to the Symposium held in Rüschlikon, May 1971, the Secretariat convened in Strasbourg a meeting of experts to prepare on the basis of recommendations made in Rüschlikon a phased plan for the implementation of a European unit/credit system for the learning of modern languages in adult education. Experts from five member States and an observer from the British Broadcasting Corporation took part in the meeting.

The main tasks concerning the implementation of this plan were defined by the participants as follows :

- to break down the global concept of language into units and sub-units based upon an analysis of particular groups of adult learners, corresponding to their personal and typical communication situations. This analysis should lead to a precise articulation of the notion of "common core", with specialist extensions at different proficiency levels ;
- to set up on the basis of this analysis an operational specification for learning objectives ;
- to formulate, in consultation with the Steering Group on educational technology, a metasystem defining the structure of multi-media learning systems to achieve these objectives in terms of the unit/cridit concept.

As for the preparation of a plan of work, it was decided that in the *first phase* of a development and research programme (1971-72), the following preliminary studies should be carried out :

— A methodological analysis of adult learners groups in terms of their communication situations, with a view to establishing a model for the definition of language needs of adults learning a modern language.

— An investigation into the linguistic and situational content of the "common core" in a unit/credit system.

— A definition of a level of basic competence (threshold-levels) in each of the four skills (listening, speaking, reading and writing).

This work will help to map out, in a *second phase* (1972-73), an integrated units/credits scheme. The results of this research, intended to provide an adequate specification of language content, should then, in a *third phase* (1973-74), be combined with media taxonomy, with a view to producing multi-media materials on the basis of the proposed units/credits system.

Documents : EES/Symposium 53, 12 ;
CCC/EES (71) 55 ; 102 ; 120.

Educational Documentation and Research

Neuchâtel *21th - 24th September 1971*

Research into the acquisition of reading skills
(Symposium)

A Symposium on research into the acquisition of reading skills was arranged under the auspices of the Council of Europe by the Institut Romand de Recherches et de Documentation pédagogiques (I.R.D.P.), whose Director, Professor S. Roller, acted as Chairman.

This Symposium was in line with the CCC's efforts to encourage member States to promote meetings between research workers and administrators in this new form so that the two groups might have the opportunity of studying their problems, exchanging experience, co-ordinating both their projects and their requirements and, wherever possible, reaching conclusions with regard to new teaching methods and research on a European scale.

The Neuchâtel Symposium was attended by experts from Belgium, Canada, France, Luxembourg, Switzerland and Sweden.

The discsusions were primarily centred on two aspects :

— Progress made in psychopedagogic work concerning the acquisition of reading skills, especially in the following five fields : perception, spoken and written language, learning, affectivity, vehicular thinking and vocabulary.

— Progress made by educational research workers in the assessment of short, medium and long-term methods of acquiring reading skills.

The Symposium included a number of talks on :

— the state of reading instruction in four wholly or partly French-speaking countries : Belgium, Canada, France, Switzerland ;

— the present state of research by psychologists, psycho-linguists and psycho-educationalists into the learning of written language and particularly of reading skills.

In addition, three working groups considered problems related to the preparation, acquisition and consolidation of reading skills and drew up reports providing the basis for a summary of the discussions.

This summary was intended to encourage the establishment of machinery for mutual exchange of information compiled by psychologists, linguists and psycho-educationalists, on the one hand, and by teachers, on the other.

A report on the proceedings and findings of the Symposium will be drawn up by Professor Roller and submitted to the Council of Europe for comment and approval.

Paris *11th - 12th October 1971*

EUDISED project
(Meeting of the Steering Group)

The second stage of the EUDISED project was concluded at the Paris meeting of its Steering Group. A draft report, prepared by the Group's rapporteur Mr. J. Viet, and eight studies commissioned from experts of various member States were discussed in great detail. The final versions as they emerged from the meeting will now be published by the Secretariat and submitted for further deliberation and action to the next plenary meeting of the ad hoc Committee for Educational Documentation and Information.

Whereas the first stage of the project led to the publication of a feasibility study supported by national reports and technical studies (EUDISED Report, 3 vols., Strasbourg,

1969), the second stage concentrated on examining the technical agreements which have to be reached to implement the project. How can computerised national and international projects concerning educational documentation and information be co-ordinated ? What are the requirements which a network for information retrieval concerning educational research and development, planning and policy, technological media, subject matter instruction — to name only these fields — has to fulfil ? How can a multi-lingual thesaurus, which is to be used by all centres and projects co-operating within the system, be built up ? What agreements on common formats and standards are necessary to enable a direct exchange of tapes or disks ? These are the questions which the second EUDISED Report seeks to answer.

Documents : DECS/Doc (71) 1 ; 6 ; 8 ; 9.

London *10th - 11th November 1971*

Colloquium of Directors of Educational Reserach Organisations

Forty-one directors from eighteen member States met for three days to discuss common problems. Observers from UNESCO, OECD, the European Commission, the European Cultural Foundation, the United Kingdom Social Science Research Council and the Canadian Ontario Institute for Studies in Education attended the meeting which was organised by the Educational Research Committee in collaboration with the National Foundation for Educational Research in England and Wales. It was the first time that such a Colloquium had been held. When at the end of the meeting participants unanimously recommended that such colloquia should be repeated at two-year intervals, it became clear that also in this field a European community has come into being, requiring its own channels of communication.

The Colloquium was introduced by addresses from Mr. W. van Straubenzee, Parliamentary Under-Secretary of State, Department of Education and Science and Mr. N. Borch-Jacobsen, Director of Education and of Cultural and Scientific Affairs of the Council of Europe. The Chairman of the Colloquium, Professor W. Taylor, Bristol University, gave the introductory lecture on "Prospects and problems in educational research co-operation in Europe" which was followed by a panel discussion. In the afternoon of the first day participants paid a visit to the headquarters of the National Foundation for Educational Research in England and Wales where they had the opportunity to discuss the ongoing projects in which they were most interested, with the project directors.

The second and third days were devoted to discussing the two main themes. The present chairman of the Educational Research Committee, Mr. L. Legrand, summarised his paper on "Policy of Educational Research Organisations" and Professor K. Härnqvist, Göteborg University, reported on his study on "Training and Career Structures of Educational Researchers". Both lectures were followed by panel discussions and thereupon the meeting split into small groups to discuss the main themes on the basis of simulation papers prepared by Professor W. Taylor. Finally, spokesmen of each group reported to the plenary on the results of the discussion.

In the concluding session Professor G. de Landsheere, Liège University, summed up the results of the discussions and the recommendations made.
The two main recommendations were :

— to examine subjects and methods for co-operative educational research projects on a European scale and, eventually, the creation of a European Foundation for the Promotion of Educational Research and Development to be structured similarly to the European Youth Foundation at present under discussion, and
— to study the possibilities for reforming and harmonising the training and career structures of educational researchers in member States.

The results of the Coloquium and in particular the recommendations will be discussed at the next meeting of the Educational Research Committee.

Documents : Simulation Papers ; DECS/Rech (71) 19 ; 20.

Second Part

COUNCIL OF EUROPE STUDIES

The Council of Europe, and in particular its Committee for General and Technical Education, has devoted special attention to the problems inherent in traditional examination methods and the introduction of new techniques for assessment. The ad hoc Conference of European Ministers of Education, held at Strasbourg in 1967, emphasised the importance of these questions and, following its Resolution No. 4/1967 on the place of examinations in the school system, the Committee for General and Technical Education commissioned the following studies : "New techniques for assessment of pupils' work", by A. D. C. Peterson (Oxford), and "Secondary school leaving examinations", by E. Egger (Geneva). Mr. Peterson's study, which is set out in full below, will be supplemented by extracts from Mr. Egger's report.

NEW TECHNIQUES FOR ASSESSMENT OF PUPILS WORK

by A. D. C. PETERSON

1.0. GENERAL INTRODUCTION

This study is concerned with the formal assessment of pupils' work resulting in a publicly available grade, rank order or orientation. Of course teachers are continuously assessing their pupils' progress as a matter of personal judgement, but the techniques by which they do this are too subjective and the results too rarely formulated to be the subject of a study such as this. Moreover it is the formal assessments and their published outcome which are of immediate concern to educators in all countries.

When Professor F. Hotyat (1) published in 1962 his magisterial survey of examinations he was concerned very largely with examinations for the selection of pupils at the point of entry to secondary education, the age of eleven or twelve. The fact that selective procedures at this stage dominated research work in the first two decades after the second world war can be seen from the list of references which Hotyat quotes. It was natural that this should be so. In terms of human destinies, and therefore politically, it was selection at this stage which was of crucial importance. A study which concentrates today on methods of assessment at the point where they most vitally affect both the pupil and the educational system must move forward to the assessment of achievement at the end of the secondary stage, to selection for entry to the tertiary stage and to the continuing process of orientation (2).

In doing so we should be wise not to neglect the lesson of the immediate past. Events overtook those docimologists (students of examinations) who laboured with such devotion to improve the reliability of selection at 'eleven plus'. As the democratisation of education at the secondary stage becomes a reality and secondary education throughout Europe is genuinely opened to all, the selection procedures whose reliability once seemed — and indeed was — so crucially important have begun to vanish before their eyes.

All the evidence seems to confirm that it will. Yet the procedure may well be slower than in the case of selection for secondary education. Even on the general theories of the sociologists, which

(1) Hotyat, F. (1962) : *Les Examens.* Paris. Editions Bourrelier for the Unesco Institute, Hamburg.

(2) Piéron, H. (1963) : *Examens et Docimologie.* Paris. Presses Universitaires de France, p. 4.

predict that the achievement of universal education at any one stage in education leads to open entry at the next within a generation, Europe will still be concerned for many years to improve its methods of assessment and selection at the end of the secondary stage : and the economic barriers to the achievement of universal tertiary education may prove more formidable than did those blocking the way to universal secondary education. Moreover, it is clear that as continued education of some sort is increasingly enjoyed by all and selection, in the sense of acceptance or rejection. dies out, so orientation to different types of extended education becomes more important. For orientation to the appropriate type of course assessment both of potential and achievement continues to be of the greatest importance. The fact that it now contributes to guidance rather than to allocation, means simply that the factor of the pupil's desire and commitment now enters into the process to a degree which was not possible when the decision had to be made by an external judge, that is by the examiners. It will be one of the contentions of this study that this combination of external and internal factors does not in fact happen so long as the purpose of assessment is the distribution of a limited number of opportunities for extended education among a larger number of applicants, even though it would in fact improve the validity of the process if it did. The reason for this is clear. So long as extended education is, in economic terms, a 'good' in limited supply, made available to some and not to others and provided by the State out of the common resources of society, the method by which those who are to enjoy it are selected is of paramount importance to social justice. This means that it must at least seem to be impartial as between rich and poor, free from favouritism of any kind and conducted with open and scrupulous accuracy. Hence the great concentration on improving the reliability of the assessments on which the allocation of such important 'life-chances' is made.

Throughout Europe there has developed in the years since the Carnegie Commission's report of 1934 ([3]) a growing amount of dissatisfaction with traditional examinations. This dissatisfaction extends from the protest of the libertarian student against any form of judgement or classification of human beings, at one end of the scale, to the scepticism of the professional educator about the reliability of current examining practices at the other. Many examinations which were once the object of vehement criticism, the 'eleven-plus' in England, the studentexamen in Sweden or the first part of the baccalauréat in France, have either disappeared or are in the process of disappearing.

Nevertheless, so long as society demands certificates of competence from the entrants to such professions as law, medicine, engineering and accountancy, certification examinations will be necessary : and as long as there are more young people seeking to enter a particular stage in the educational process, whether higher secondary or tertiary, than the institutions in this stage can accommodate, selection devices, including different types of examination, will persist. The more education is democratised, in the sense that educational opportunity, however limited, is equally open to the rich and the poor, the more crucial becomes the role of the selection procedure. Before the second world war most European countries controlled the proportion of the age group entering tertiary education by eliminating most of the poor. If ressources do not allow the whole age group to be admitted and poverty is not to be the barrier, then someone else must be eliminated. In a democratic society this can only be those who fail selection tests. The purpose for which selection examinations are actually used, therefore, is often to provide something which people can fail. It is on the assumption that examinations either for certification or selection will continue to be necessary for many years that many educators have been turning their attention to improvements in the technique of assessing pupils' work. This study is concerned with the search for these improvements and in it no distinction will be made between the 'examination' and the 'test'. Even so short a time ago as 1958 S. Wiseman, later Director of the National Foundation for Educational Research, wrote : "Teachers, on the whole, are responsible for 'examinations', psychologists for tests" ([4]). This may have been a necessary distinction then, but in the last ten years the rapprochement which he advocated has taken place, and 'objective testing' now plays a large part in many school examinations.

Many factors enter into the judgement of what is or is not, in any particular case, a good techni-

(3) International Institute Examination Inquiry (1936) : *La Correction des Epreuves Ecrites dans les Examens — Enquête Expérimentale sur le Baccalauréat.* Paris. A la Maison du Livre.
and
Hartog and Rhodes (1935) : *An Examination of Examinations.* London. Macmillan.

(4) Wiseman, S. (ed.) (1961) : *Examinations and English Education.* Manchester University Press, p. 134.

que of assessment. Of these the most important are : 'backwash', validity, reliability, cost and speed. Attempts to improve assessment techniques usually concentrate on one or other of these and it often turns out that an improvement in one factor can only be achieved at the cost of a deterioration in another. In England, for instance, attempts to improve the reliability of GCE examinations since the report of the Carnegie Commission have had considerable success : but they have been paid for in cost, speed and backwash effect. Hence the importance of clearly evaluating the relative importance of the factors.

By 'backwash' is meant the effect which any particular examining technique will have on the teaching and learning which goes on in the period devoted to preparing for the examination. For instance, if an examination in Chemistry rewards with success the capacity to reproduce from memory a large number of formulae, teachers will tend to spend a great deal of time on memory drill and neglect laboratory work : if on the other hand it rewards practical manipulative skill, but provides the pupil with any formulae which he will need in his answers, teachers will not bother to make their pupils learn them, but will drill them intensively in manipulative skills. This whole important area was analysed for the Council of Europe in Professor A. Agazzi's Report on "The Educational Aspects of Examinations" (5).

The concept of validity is simple. An examination is valid to the extent that it measures what it purports to measure and not something else. A geography examination which gives great weight to the beauty and neatness of the maps drawn by the candidates is to that extent measuring draughtsmanship, not geographical understanding. It may be that draughtsmanship is an important skill for the geographer, but in that case it should be made clear to candidates that this skill, as well as their geographical knowledge and understanding, is being tested, and examiners must be accurately briefed as to the weight to be given to it. One of the great difficulties in assessing validity is, in fact, that educators have been very slow to state in clear operational terms what are the objectives of their courses. Unless we know what the pupil is expected to have gained from the course, how can we judge whether the examination is validly measuring that gain ?

Another difficulty often arises from the near impossibility of finding another measure of the stated objectives, when they are stated, against which the examination, as a measuring device, can be calibrated. Let us suppose that one of the objectives of a course in literature is the development and refinement of the moral sense. What other measure have we got of the extent to which this has happened than the examination itself ? And if we have no other at least equally good measure, how can we judge to what extent the examination is a valid measuring instrument ? A common process for estimating the validity of a new type of examination is to correlate it with the teacher's careful estimate of the pupil's ability, but this suffers from a high degree of subjectivity and if we were really satisfied that the teacher's estimate was the most valid measurement, the examination might be unnecessary. Nevertheless attempts to improve the validity of examinations are well worth making. The most reliable examination in the world will do nothing but harm if it has little or no validity. We could select candidates for tertiary education with almost complete reliability by weighing them.

There is one special form of validity which has much concerned docimologists. This is 'predictive validity'. Here it is assumed that the purpose of the course was not to bring about changes of any intrinsic and specified nature in the behaviour of the pupil, but merely to prepare him for the successful completion of the next stage in the educational process. The examination is therefore not intended to measure any existing qualities of the candidate but to forecast his future performance. Neither of the problems outlined above arise in this case and the predictive validity of any examination is often measured simply by the extent to which it predicts success in the next examination.

It was on the reliability of commonly used examining techniques that the report of the Carnegie Commission cast serious and alarming doubt. If the competence of a professional man or the future career of a student is to depend on an examination result, then it is desirable that this result should represent a consistent judgement and not a fluctuating assessment dependent on chance factors. Yet the report of the commission showed that in such examinations different examiners would mark very differently the same paper and even that the same examiner would mark it substantially differently at six month intervals. H. Piéron describing the conclusions of the French commission on the baccalauréat goes so far as to conclude that it showed that "to predict the mark of a candidate it was more important to know the

(5) Agazzi, A. (1967) : *Les Aspects Pédagogiques des Examens*. Strasbourg. Council of Europe.

examiner than the candidate" (6). It is not surprising therefore that much of the work which has subsequently been done on the improvement of examining techniques has been devoted to improving their reliability.

It is natural that work on reliability should lead on to cost. To mark questions of the 'essay' type with any high degree of reliability it seems necessary to resort to multiple marking, the average of a number of even fairly superficial markings by different examiners proving more reliable than careful marking by a single examiner. But examiners in most examination systems have to be paid. Here it should be sufficient to insist that the true cost of any system of examining should be calculated in terms of opportunities foregone — opportunities for teachers to teach and students to learn.

Finally speed is an important factor in any examination. Here I am referring no longer to the opportunity foregone in terms of the time occupied during the examination period which might have been more educationally spent in the class-room or library, but of the importance of rapid publication of results, particularly where these determine entry to the next stage of education. In England, for instance, the time gap between the taking of the first paper in the General Certificate of Education at Advanced Level and the publication of results may be as much as twelve weeks. Examinations start early in June and results are published half way through August. In France the results of the baccalauréat are published within three weeks and in Germany, as in Europe generally, the results of equivalent examinations are available equally quickly and before the schools disperse in late June or early July for the summer holidays. The disadvantages of the longer period devoted to assessment, both in terms of student and family anxiety and of inconvenience for the 'receiving' institution, which does not know until almost the last moment which of its candidates have reached the required standard, have to be weighed against the increased reliability which may be achieved by more prolonged scrutiny.

The value of any new technique of assessment, therefore, will depend on the balance of advantage along these five parameters.

2.0. EFFORTS TO IMPROVE RELIABILITY

Recent attempts to improve the reliability of examinations by introducing new techniques of assessment have mainly concentrated on diminishing the unreliability due to the examiner's subjective judgement. This unreliability arises not only from genuine differences of opinion between different examiners about the worth of a particular answer, but because examiners tend to be affected, and affected to different degrees, by factors which may or may not be intended to enter into the assessment. Thus a pupil who excels in literary style may secure not only a high mark in the literature examination but also a higher mark in history than another pupil whose actual understanding of history is greater, but whose powers of expression are inferior. Professor W. D. Furneaux (7) has shown in England that with conventional 'essay type' examinations there is a common element of "general examination-passing ability" which enters into the assessment of all subjects and which may therefore carry undue weight in what purports to be a balanced assessment based on a number of tests in different areas of the curriculum. Professor K. Ingenkamp reports similar evidence of the influence of extraneous factors from studies in the 1960s on the German Abitur (8).

It has long been recognised that in extended examination answers at the upper secondary or tertiary level this so-called 'halo' effect will operate. An examiner, judging on general impression of an extended piece of writing or of an extended interview, will be influenced by unusually good or bad performance at one stage of the written or oral answer to form a general, but subjective, opinion of the candidate's work and to extrapolate from that opinion in judging the whole of the rest of his performance. We all know how prone we are to assess anything, whether it be a new acquaintance, a painting or an examination performance, in accordance with our preconceived expectations. It would seem that, unless fully recognised and deliberately intended as part of the assessment procedure this could be particularly dangerous in systems where the final decision depends on the views of a 'jury'.

2.1. *Objective tests*

One way to avoid this halo effect and also to avoid the subjective differences inevitably arising from the judgements of different examiners, some severe

(6) Piéron, H. : *Op. Cit.* p. 24.

(7) Furneaux, W.D. (1962) : The Psychologist and the University. *Universities Quarterly*. Vol. 17 No. 1.

(8) Ingenkamp, K. (1969) : Möglichkeiten und Grenzen des Lehrerurteils und des Schultests. Deutscher Bildungsrat : *Begabung und Lernen*. Stuttgart. pp. 409-410.

and others lenient, some tired at the end of the day and others optimistic at the beginning, is to break down the qualities and knowledge which it is intended to examine into measurable units which can be assessed separately and objectively. It is the search for this kind of 'objective' reliability which has given rise to the so-called 'objective test'.

The objective test does not ask the candidate to develop in his own terms, either written or oral, the full answer to a general question, but to respond to a very specific question or to select from a number of possible answers the one which seems to him the most appropriate.

Since the justification of the technique depends upon this prior analysis into measurable units of the skills or information which it is intended to assess, the starting point in the construction of objective tests is always the definition of the outcome required, the truism that accurate measurement only becomes possible when one is clear about what is being measured. Here the various classifications of different kinds of learning which will be discussed later under the heading of 'validity' have made it possible to use tests to measure the ability, not merely to recall facts but to understand and apply general principles induced from facts. Thus it is claimed that in this fashion it is possible to assess a whole range of learning. In the IEA study, for example, eight objectives were listed in the testing of physical sciences (9) :

- Obtaining scientific information
- Interpreting scientific information
- Theorisation ; Construction
- Theorisation ; Utilisation
- Comprehension
- Application of scientific knowledge
- Personal and social objectives
- Philosophical aspects

Questions were then devised to test each of these.

Once the objectives of the course have been classified under this or some similar system it is possible to consider which of them can best be tested by this method, what balance struck between them and what weight given to each in the final global assessment. When it has been decided to make use of this type of testing to assess performance either in a whole course or in some part of it, the next step in the generally accepted procedure is to construct a 'grid', a simplified illustration of which is given in Fig. 1. drawn from a grid designed to test the outcome of courses in civic education.

(9) U.S. Department of Health, Education and Welfare : *Cross-National Study of Educational Attainment. Stage 1. Investigation in Six Subject Areas. Final Report.* Feb. 1969. p. D-3.

Figure 1

Content of syllabus (civics)	Learning outcome		
	Knowledge of facts and principles	Understanding of facts and principles	Application of facts and principles
Constitutions			
Historical development			
Local/central control			
The legislative			
The executive			
The judiciary			

The text constructor must then consider what sort of questions, usually known as 'items' to include in each box of the grid. The main difference between them is between those which require the candidate to supply an answer and those which require him merely to select the most appropriate from a list of answers presented in the question paper. Both have the advantage that they test the candidate's knowledge or ability over a much wider range of the course which he is supposed to have studied than can a limited number of essay questions, and so diminish the unreliability due either to 'halo' effect or to the 'luck of the draw' in a candidate's finding questions in the examination which relate either to his strongest or to his weakest points.

The supply type, however, which is often known as the 'short answer' question cannot be marked completely objectively, either by clerks or machine since answers will differ and someone competent to do so must judge whether or to what extent they are acceptable. It is possible that in very large scale examinations it may become possible to overcome this objection by programming a machine to accept a large number of specified variations and so to mark automatically a wide range of free responses, but even this procedure would not be able to reward adequately the brilliant and totally unexpected response. A typical 'short-answer' item, drawn from the International Baccalaureate examination in Physical Science for 1969 is the following :

> "At 273 K and atmospheric pressure (101.3 kPa) 10^{-3} m^3 of a certain gas are found to have a mass of 0.18 g.
>
> (a) Calculate the number of moles of gas present
>
> (b) Calculate the molecular weight of the gas
>
> (c) Calculate the r.m.s. velocity of the gas molecules at 273 K."

Beneath each question an adequate amount of space is left for the candidate to write an answer of the length required.

The 'selection' type of time requires no more of the candidate than to choose between alternative answers presented to him. He writes nothing, but simply ticks or underlines the correct response. Items of this type may take several forms. The candidate may merely be required to say whether certain statements presented to him are true or false. Yet even this apparently simple type is capable of considerable sophistication as is shown by the following example, quoted from *Objective Testing* by H. G. Mackintosh and R. B. Morrison (10) :

> *"Directions*
>
> Each of the following items may be true without qualification, true with qualification, or false. If it is true without qualification, circle the T and mark a 3 in the space provided. If it is true with one of the listed qualifications, circle T and mark the number of the appropriate qualification in the space. If the item is false, circle the F.
>
> *Statements*
>
> T F..... The total resistance in an electrical circuit is equal to the sum of the individual resistances.
>
> T F..... The total current in an electrical circuit is equal to the sum of the currents in the individual parts of the circuit.
>
> T F..... The total current in an electrical circuit is equal to the electromotive force in the circuit divided by the resistance of the circuit.
>
> T F..... The power supplied to a circuit is equal to the product of the total resistance and the amount of current in the circuit."
>
> *Qualifications*
>
> 1 if the resistances are connected in parallel.
>
> 2 if the resistances are connected in series.
>
> 3 no qualification.

He may on the other hand be asked to select from a number of possible responses, usually five, the one which is appropriate. This is the typical 'multiple choice' question. It is capable of testing more than the mere recall of factual information as will be seen from the examples given below, but it requires considerable skill in composition.

> (a) "Which of the following had as their major purpose the achievement of national independence ?
>
> I The Octobrists
> II The Sinn Feiners
> III The Young Turks
> IV The Carbonari
> V The One Thousand
>
> (A) I and III only
> (B) II and V only

(10) Mackintosh, H.G. and Morrison, R.B. (1969) : *Objective Testing*. London. University of London Press.

(C) I, III and IV only
(D) II, IV and V only
(E) I, II, III, IV and V"

(quoted from *College Entrance Examinations Board Achievement Tests — 1965)* (11).

(b) "500 houseflies were kept in a cage. One cubic centimetre of 1 % DDT solution was discharged into the cage and as a result 95 % of the flies died. The survivors were transferred to a fresh cage and allowed to reproduce. The resulting adults were then submitted to the same DDT treatment. The survivors were transferred to a fresh cage and the procedure was repeated. The mortality rate declined with each generation until it was 34 % for the 14th generation.

The most acceptable scientific explanation for these results is

( ) a. repeated DDT treatment causes houseflies to become resistant
( ) b. DDT treatment causes mutations in the genetic material
( ) c. a few flies get a sub-lethal dose and then become immune
( ) d. interbreeding causes mutation leading to DDT resistance
( ) e. DDT resistance occurs naturally and is inheritable."

(quoted from *International Baccalaureate Higher Level Biology 1970).*

The essentials of a good multiple choice item are :

(a) That the preliminary information, known as the 'stem' should contain all the information that a candidate ought to need, but no guidance.
(b) That there should be one and only one answer that is acceptable.
(c) That the other answers, known as the 'distractors' should be sufficiently plausible not to rule themselves out of court, but not so nearly correct as to allow of genuine difference of opinion.
(d) That the degree of skill or information required to pick the right answer should be sufficient to distinguish between the more and the less well prepared candidates.

Clearly this latter criterion will vary from item to item and it is usually considered wise to begin a test with a number of easy items, which almost all candidates will get right and proceed gradually to more difficult items.

The extreme difficulty of composing fifty or a hundred items of this kind makes it desirable that multiple choice tests should be composed by a team of experienced examiners working together. Even so the first version of the test is likely to contain some items which in practice do not call forth the responses that the team expected, and it is usual to administer the test in advance to a population similar to, but different from, those pupils for whom it is being designed. In this way the faulty items, particularly those which fail to discriminate, in the sense that they are answered correctly by all or by none, can be identified and rejected. A commonly accepted rule is that the limits of acceptable discrimination lie between correct responses from thirty and from seventy per cent of the candidates taking the test.

The advantages claimed for the multiple-choice type of examination are many and it will be convenient to tabulate them and then to discuss them separately.

(a) It is extremely reliable in that it eliminates from the examination the subjectivity of the examiner. Each examination paper will score the same, whoever marks it and at whatever time it is marked. Indeed it can be and often is marked by machine. In view of what we know about the unreliability of examinations of the 'essay' type when marked by a single examiner this is a very significant advantage. It is sometimes argued that the multiple choice type of examination merely removes the element of subjectivity from the marking of the answers to the composition of the questions. This is, of course, true, but the number of people involved in the construction of a multiple choice paper, whether as test constructors, item writers or pre-test participants, is so great that the final result cannot be regarded as representing the judgement of one man, as can the assessment of an 'essay' type paper set and marked by a single examiner. Moreover, even if an element of subjectivity enters into the construction of the paper and so favours a whole group of candidates with one cast of mind rather than another, at least the performance of all candidates on that paper is assessed equally, reliably and without the fluctuations due to the subjectivity of different examiners.

(b) Provided that it is really true that the qualities, skills and information which it is intended to assess can be analysed and broken down into a large number of separate and independent parts,

(11) College Entrance Examination Board (1963) : *Achievement Tests.* New York. CEEB p. 75.

then the multiple choice paper can sample the whole area to be tested much more fairly than can the traditional 'essay' type examination.

Let us illustrate this from the examining of foreign languages and let us assume that one purpose of our examination is to discover whether the candidate has a wide knowledge of the syntactical and idiomatic conventions of the language. If we ask him to write an essay, or even to make a translation into the language we shall learn little about his competence over the whole of this field. The subject set for the essay may or may not demand the full range of his knowledge. If he uses what appears to be a restricted range, it may be because he has a genuine purity of style which seeks always the simplest forms. If, on the other hand, his prose exemplifies a wide variety of structures, he may be artificially dragging in constructions and idioms which, while representing the extreme range of his own capacity, still fall short of the best we could expect.

On the other hand if we ask him to answer fifty to a hundred multiple choice questions, covering the full range of all that we might expect him to know, we can discover with a great deal of accuracy how far his range of syntactical and idiomatic usage extends. We shall not, of course, discover how far he is capable of using that range in sensitive and meaningful prose, but that is not what, in this particular part of the examination, we were setting out to discover.

(c) As soon as the number of candidates passes a certain figure, which will vary with the subject being examined, this method becomes progressively cheaper than the conventional 'essay' type examination or the oral. For multiple choice papers the main cost lies in the construction and pre-testing of the examination : the actual papers can be very cheaply marked by clerks or a machine. For 'essay' type papers the actual construction of the examination is not very expensive and is sometimes done by a single examiner in a few hours. What matters is not so much the questions he asks as the answers the candidates give. But this means that the answers must be marked by highly qualified academics, perhaps as highly qualified as the examiner who has set the paper : and the time of highly qualified examiners is valuable and expensive. We must remember here the point made earlier that the costs of an examination should be measured not in cash expended but in opportunities foregone. One of the questions which all industrialised countries have to ask themselves is whether too high a proportion of the time of their limited supply of highly qualified academics is being used in examining rather than teaching.

(d) This method is quick. With machine scoring the results of tens of thousands of examinations can be provided within a few days. It was pointed out in the introduction that there is a certain importance in this criterion, if the results of examinations are to be used to determine transition from one stage in education to another. To achieve the same speed in the provision of results on 'essay' type examinations for a comparably large group of candidates, it would be necessary to employ hundreds of different examiners. And the greater the number of different examiners employed the greater the unreliability of the results.

It seems therefore that objective tests have considerable advantages from the point of view of reliability, cost and speed. They are criticised on grounds of validity and backwash effect.

Early criticisms of their validity, made when they first appeared were based on the view that they could test no more than the factual recall of information and that they favoured random guessing. This may have been true of the earliest type of true/false item (e.g. "Bergen is the capital of Norway" : True : False), but it is manifestly not true of some of the more sophisticated items now in use. Moreover a certain amount of retention in the memory of factual information is an essential part of any academic study or, indeed, of any effective intellectual activity at all. It is, therefore, *one* of the factors which any process of assessment should take into account, provided it is not made the dominating factor. A more sophisticated criticism is that many of the more recently developed items which purport to test judgement or interpretation of evidence are really testing not these activities themselves, but the recollection of them, as previously undertaken during the course either by the pupil or his teacher. This may well be true : but it is a criticism which can be equally well made of examination questions of the conventional essay type. Other criticisms of validity have perhaps arisen from inflated expectations. Few tests are perfect in terms of the criteria proposed above. Professor B. Hoffman has shown in *The Tyranny of Testing* (12) that some items can be so ill-designed as positively to discriminate against the subtle or more inventive thinker and this is undoubtedly true. Even in the most scrupulously designed tests it is still possible that individual bad items will appear, but the influence

(12) Hoffman, B. (1964) : *The Tyranny of Testing*. New York. Collier-Macmillan.

of these on scores, when they are one or two out of fifty is slight, and the test, as a whole, may still be sampling the whole syllabus in a more comprehensive and reliable way than a conventional examination. 'Objective testing' has acquired a mystique and perfectionism of its own and it is sometimes forgotten that, for certain purposes, even a faulty multiple choice test may be a better measuring instrument than a traditional examination.

The objection that tests of this kind favour random guessing is part of a wider criticism of their validity arising from the fact that they do rely on a number of separate and unrelated responses. It is of course true that a candidate who neither knew nor thought anything could expect to score twenty-five on a hundred item test where there were four choices or fifty on a true/false test. Although this can be corrected for by the use of a simple formula many test constructors are in fact avoiding the true/false type. Where four, or more commonly five, alternatives are offered in a selection test, they argue that candidates do not, in fact, guess at random and that a high percentage of right 'guesses' simply indicates a range of competence which, although not complete, is very considerable and should be assessed as such. They therefore make no statistical adjustment for guessing at all (13). It may, however, be significant that the American College Entrance Examination Board which has greater experience of these tests than any other body, continues to arrive at its final score by applying the formula suggested by the Educational Testing Service :

$$R - \frac{W}{n-1}$$

Where R is the number of right answers, W the number of wrong answers and n the number of choices offered.

The justification for this procedure is clearly described in the Board's pamphlet on the Scholastic Aptitude Tests : "When the SAT is scored, a percentage of the wrong answers is subtracted from the number of right answers as a correction for haphazard guessing. Mere guessing, therefore, is as likely to lower your scores as to raise them. If, however, you are not sure of the correct answer but have some knowledge of a question and are able to eliminate one or more of the answer choices as wrong, your chance of getting the right answer is improved, and it will be to your advantage to answer such a question " (14).

The real objection to using objective tests simply to replace conventional examinations was given with great eloquence by M. Desclos, President of the French Commission at the time of the Carnegie enquiry :

> "Si l'on se bornait à considérer les examens du seul point de vue de leur utilité en tant qu'instruments de sélection d'après les connaissances acquises, on pourrait être amené à envisager des épreuves quasi mécaniques capables d'inventorier rapidement le stock des notions que possèdent les candidats et de fournir des résultats parfaitement exacts ou objectifs. Mais si notre ambition va au-delà, si nous estimons que l'objet des études est moins d'accumuler des connaissances que de les coordonner pour en tirer une philosophie et pour les mettre en œuvre, moins de meubler l'esprit que de l'assouplir, l'affiner et le vivifier, si nous attachons du prix à la liberté de jugement, au sens critique, à l'effort personnel, à l'imagination créatrice, au goût et à la mesure, à toutes les richesses spirituelles, subtiles et fugitives qui constituent la culture, il faut nous en tenir à des épreuves d'examen moins exactes sans doute, mais capables d'en déceler l'existence, de les évaluer sinon de les mesurer, et qui permettent en même temps d'utiliser pour leur préparation des exercices qu'une expérience pédagogique séculaire a mis au point et dont l'efficacité est incontestable " (15).

Refinements of the technique of objective testing may have enabled us to extend its use to assessing more than 'the stock of ideas', but the truth remains that essential qualities involved in the ordering and marshalling of thought, in creativity, in imagination and style, to say nothing of what is now known as 'the affective domain' are difficult if not impossible to assess by a method which sums the responses to a number of isolated stimuli. This can easily be seen from the example quoted earlier of assessment of command of a foreign language. The multiple choice test will tell us better than any essay how wide and accurate is the candidate's command of lexis. It will not tell us whether he is capable of using this command to develop a coherent argument in the language, whether his imagination is sterile or rich, or whether he has

(13) Ebel, R. (1965) : *Measuring Educational Achievement.* New York. Prentice-Hall. pp. 98-99.

(14) College Entrance Examination Board (1965) : *Scholastic Aptitude Test.* New York CEEB.

(15) International Institute Examination Inquiry : *Op. Cit.* p. 385.

any 'style' in his use of language. Even the attempts which have begun in America to test literary appreciation by multiple choice questions, although interesting in themselves, have so far proved of limited value.

When we come to the most important question of backwash effect objective tests are heavily criticised. The main objection is that which Desclos hints at in his last lines, that if this is known to be the sole method of assessment, teachers will be tempted to neglect the development of their pupils' capacity to write continuous prose. It has been said of these tests that they require the candidate to read and think, where traditional examinations require him to think and write : and writing is a very important accomplishment. So strong was the reaction in the United States that the College Board tried to counteract it by requiring a 'writing sample' as well as the completion of the tests from each candidate. Unfortunately it proved impossible to achieve the degree of reliability in the assessment of writing samples that had come to be expected and the requirement was dropped. This negative backwash effect remains, however, a very serious one in any country where teaching methods are strongly influenced by examining procedures.

A positively harmful backwash effect has been seen in the danger that constant practice in completing this type of test will condition pupils to the belief that intellectual activity consists in a series of responses to a set of problems, for each of which there is one and only one completely right answer — in fact that the manifold diversity of our experience is best conceived as a programme fed to us by the heavenly computer. M. Trow, the American sociologist is recorded as having found "that at one institute in America students had so many objective tests during their first year that their approach to their studies became dominated by the need to pass exams. The next two years were spent curing them" ([16]).

2.2. *Multiple marking*

One technique for securing greater reliability in the grading of 'essay' type questions is to have them marked by more than one marker and to adopt as the final grade the average grade awarded. There is little doubt that this does improve reliability. Its disadvantages lie in cost and speed.

If a very quick superficial reading by three separate examiners is substituted for a careful assessment by a single examiner taking three times as long over the work, the cost in terms of examiners time need not be any greater ; and it is claimed that this technique did give improved reliability in the assessment of English essays in the 'eleven-plus' examination in England ([17]). Even here however there is an extra cost in photo-copying and postage or an extra delay and administrative risk in requiring one reader to post scripts on to the next or in bringing together readers to a central point. Moreover a technique which gave improved results in the assessment of a more or less free composition written by pupils aged between ten and eleven cannot be assumed to be of equal value in assessing other forms of work.

At the upper secondary level multiple marking is at present employed in Luxemburg. Each written paper is marked by three examiners drawn from the 'commissions' for three different schools who send their grades, independently and without consultation, to the government assessor. If there is a significant difference between the grades the assessor calls the three markers together to re-examine the script. If this confrontation does not produce agreement the assessor submits the script to the commission for the school from which the candidate comes, for a final decision. Presumably this brings in at the crucial stage an element of continuous assessment since the school's judgement is likely to be affected by what they know of the candidate's previous work.

It may well be that this system is possible only in a small and homogeneous system where sufficient examiners with comparable standards and easy access to each other are available. In France experiments were started with double-marking after the report of the Carnegie Commission but were subsequently abandoned on grounds of cost and speed.

In England double-marking of essays was introduced in the first part of the GCE (Ordinary level) which is normally taken between 15 and 16 by the Northern Board in 1967 and extended in 1969 to over 40,000 candidates.

Multiple marking is, of course, much easier to achieve in an examination system which is largely internal and marking of written scripts by two examiners, one from the school itself and one from a neighbouring Gymnasium, with an official assessor giving the final adjudication where the

(16) (Reported by) Cox, R. : *New Society* 24. 5. 1966.

(17) Wiseman, S. (1949) : The Marking of English Compositions in Grammar School Selection. *British Journal of Educational Studies* XIX. pp. 200-209.

grades differ, has been practiced in some Länder for the German Abitur. In terms of real costs this rather more than doubles the cost per candidate of actual marking, but this does not appear in money terms where those engaged in the process are not paid separately for their examining work but undertake it in lieu of teaching or administrative duties. That there is little loss of speed in such a system is probably due to the comparatively coarse grading system of the Abitur on a six point scale. If a twenty point scale similar to that of the French baccalauréat or the new proposals for the English GCE were used, differences of opinion between the two markers and the consequent necessity for the assessor to read the scripts, thus increasing both real cost and delay, might be much increased.

2.3. *Team-marking*

Another result of the Carnegie Commission and the publication of Hartog and Rhodes *Examination of Examinations*, which was based on its findings, has been much more elaborate arrangements for the setting and marking of essay type papers. The most exhaustive attempts to improve reliability by this type of sophistication have probably been made by the English GCE examining boards. These have tried to avoid the unreliability disclosed by the Commission's finding of very poor correlations between a number of independent markers by organising the markers into a 'team'.

The first step in this process is the preparation by the Chief Examiners of extremely detailed marking schemes for each paper, to which his assistant examiners are expected to adhere very closely. To quote the 1969 report of the Northern Universities Joint Matriculation Board : "When the examiners first draft their questions they provide 'notes for the answers' indicating what in their opinion candidates might reasonably be expected to include in their answers... Much more detail is however required in the marking scheme since this is intended to ensure that all members of the panel of examiners responsible for marking a paper (or a section of a paper) use the same method of assessment and adopt a similar standard" ([18]).

This marking scheme breaks down the expected response to a long piece of translation or an 'essay' type question into comparatively small elements and fixes the allocation of marks to be given for each one. We can see already therefore an approach to the sort of reliability to be expected from a number of short answer or even multiple choice questions. On the other hand with every step in this direction we are getting further from Desclos' justification of the 'essay' type questions as flexible and humane instruments encouraging creativity, imagination and flexibility of thought.

The provision of a detailed marking scheme is only a beginning. The first stage after the actual examination is for each examiner on the panel to send a batch of papers which he has provisionally marked to the Chief Examiner as a sample of that year's performance. This is followed by a meeting attended by all the examiners at which the marking scheme is revised and standardised in the light of the performance of this sample of the candidates who have taken it. The examiners then return home and begin their definitive marking During this period they continue to send samples to the chief examiner who can thus assess whether a particular examiner is proving 'severe' or 'lenient' — or worst of all inconsistent, in which case all his papers have to be redistributed for remarking by other examiners. When all the marking is completed the Chief Examiner and his assistants proceed to adjust the marks by increasing those of 'severe' examiners or diminishing those of 'lenient', until they are satisfied that a common standard has been reached. Only at this point do they decide the mark which is to be accepted for that year as a 'pass mark', and then the panel proceeds to review all scripts whose marks fall around this borderline. This final review procedure is in some cases carried out by the Chief Examiner but in others involves a final general meeting of examiners which is called the 'Award' and which may take several days. The document quoted above, for instance, records that in 1969 the review of borderline scripts in Paper B of English Language at Ordinary Level (the first part of the GCE) occupied ten examiners for ten days ([19]).

The improved reliability of 'essay' type examinations marked with this meticulous care is paid for in terms of cost and, above all, speed. The delays of postage and travelling, added to the time occupied in standardising and reviewing, mean that the English GCE candidate sometimes does not know the results of the examination he took at the beginning of June until late August and does not know whether he will be admitted to a university until two or three weeks before the session starts. The human cost here in prolonged anxiety, which is known to many English families, and the

(18) Northern Universities Joint Matriculation Board (1969) : *The Work of the Joint Matriculation Board.* Manchester. J.M.B. p. 12.

(19) Northern Universities Joint Matriculation Board : *Op. Cit.* p. 16.

administrative cost to the universities and other institutes of tertiary education, which do not know the examination results of their candidates until the new academic year is almost beginning, has already been referred to.

The financial cost may be estimated from the fact that the fee charged for each single examination in each subject at the 'Advanced Level' is now two pounds sterling. For the average university candidate taking six subjects at Ordinary level and three at Advanced level the cost would be £ 13.

How far then has this meticulous procedure succeeded in reducing the unreliability of such examinations exposed by the Carnegie Commission? It is remarkable how little research has been done on this topic since the publication of Hartog and Rhodes report in 1935. The English Examining Boards have some justification for complaining that most docimological studies are still quoting, as their main evidence for the low reliability of 'essay' type questions, studies which are thirty-five years out of date. On the other hand their own research has been mainly concentrated on establishing, or seeking to establish, the comparability of standards between one Board and another, rather than the reliability of marking within each Board. It would be extremely interesting, for instance, to set up controlled experiments to establish whether or to what extent, the elaborate procedures and long review period have given a higher reliability index to the marking of 'essay' type question in the GCE 'A' level than in the much more quickly assessed French baccalauréat, or in the triple-marked Luxemburg system.

One study, that of E.L. Black of the University of Manchester, though carried out on a small scale, reproduced as nearly as possible the conditions and safeguards now built into the GCE system ([20]). Nineteen different examiners, who had been 'briefed' in the standard way and were part of a team, were given the same script in English Language to assess ten days after the briefing and while they were in the middle of marking their official scripts. The experiment was repeated in four separate years. The marks of the different examiners varied from 53 % to 31 % in year one, from 61 % to 40 % in year two, from 65 % to 49 % in year three, from 61 % to 43 % in year four. When it is remembered that a difference of a few percentage marks can make the difference between a 'pass' and a 'fail' it is clear that, even though the correlations found are considerably better than the extremes disclosed by the Carnegie Commission, the subjectivity of the examiners' judgement has been by no means eliminated from this type of assessment. The Schools Council in England has now proposed that the examining boards should change from their present seven point scale to a twenty point scale and publish some guidance to those who interpret examination results about the degree of accuracy which can be expected. Their conclusion is that "Enough is known about the reliability of examinations of the types at present commonly set at 'A' level to establish that on the twenty point scale a standard error of *less* than two scale points would be very difficult to obtain."

One of the factors which may contribute to this degree of unreliability is the technique employed for translating the raw marks of different examiners on to a common scale of grades for publication and use in selection or orientation.

2.4. *Procedures for expressing raw marks as grades*

Various procedures may be used to transform raw marks into grades and these may affect the reliability or validity of the assessment. There are four types of unreliability which such procedures may be intended to mitigate: that which arises from examiners being called upon to assess the work of schools with widely different mean levels of ability ; that which arises from an unequal spread of marks in different subjects ; that which arises from candidates choosing different questions to answer from a paper where, for instance, the candidate is asked to answer three out of twelve questions ; and that which arises from different standards adopted from one year to the next. In all cases there is a possible conflict between reliability and validity and in all cases the correct decision in this conflict will depend upon the purpose for which the assessment is being made.

If it is assumed that there really is such a thing as an objective standard of achievement or ability and that it is the purpose of the examination to measure as nearly as possible the pupil's performance against this standard, then it is clearly important that the standard should not vary from school to school, question to question or year to year. Where global assessments are based on the summing of results in different subjects, as in Sweden, it is important that it should not vary from subject to subject.

(20) Black, E.L. (1962) : The Marking of G.C.E. Scripts. *British Journal of Educational Studies*. Vol. 11. pp. 61-71.

To assure a reliable comparability of standards between school and school it would be desirable that each examiner, each 'team' of examiners or each jury should receive scripts to correct from a representative sample of the population as a whole. It would be difficult, for example, for an examiner who saw scripts only from the Lycée Henri Quatre or Manchester Grammar School to form a just appreciation of the national standard. J. Petch claims that in the General Certificate of Education in England this is achieved by allocating schools to examiners 'alphabetically by names of towns' and thus achieving something approaching a random sample.

The danger arises, however, in any system where the examining is entrusted to a number of autonomous boards, commissions or juries, that the whole of the sample being assessed by one board may differ substantially in academic potential from the whole of the sample being assessed by another. H. B. Miles and G. E. Shipworth have shown that over a five year period the mean IQ of entrants for the different General Certificate of Education Boards in England varied very substantially (e.g. between 125/7 in one board and 109/11 in another) and yet that "it is difficult to extract from the figures any consistent relationship between IQ of entrants and grades obtained. Board 5 for example with lowest IQ's on entry is more generous in its grades on mathematics over the years than Board 2 with candidates of much higher IQ on average" ([21]).

The most carefully designed procedure to ensure the maintenance of a common standard is that which has recently been introduced in Sweden and which is described in section 3 of this study, in relation to the even greater problems of comparability involved in the use of continuous assessment by teachers.

The need for some sort of statistical adjustment to produce comparable gradings as between different subjects arises from the well established fact that examiners, except perhaps in mathematics, tend to bunch their marks about the mean and make little use of the extreme ends of the scale, reserving, as is said in France, the top mark for 'le bon Dieu'. Thus the Northern Universities Board in England found that in a typical year the marks, out of a nominal 200, ranged from 200 to 0 in mathematics with few below 48, while those in geography ranged from 144 to 11 with few below 54 ([22]). Both sets of marks had, at that time, to be translated for publication on to a five point scale roughly categorised as 'excellent, very good, acceptable, insufficient, failed'.

If the grades are to be published separately and neither global assessment nor compensation between subjects is concerned, it can be argued that nothing should be done to standardise the spread of marks between subjects, that pupils at this stage do differ much more widely in their mathematical ability than in the battery of skills and knowledge required to write a geography examination, and that the spread of marks therefore represents a true judgement and not just the inability of the examiners to perceive distinctions. If this is so, then to adjust the marks statistically so as to produce a comparable pattern of spread between subjects will be to distort the validity of the assessment.

If, on the other hand, results in different subjects are to be regarded as comparable or summed to produce a global assessment, then it is important that the spread of marks in different subjects should also be comparable. Otherwise a subject like mathematics which normally has a wide spread will dominate the process of arriving at a rank order on which, in practice, selection depends. The best way to normalise raw scores for this purpose seems to be to translate them into 'T scores', i.e. scores expressed in terms of a mean of 50 and a standard deviation of 10, giving effective upper and lower limits of 80 and 20. T scores express the raw score not in terms of a fixed scale, invariable as between subject and subject, but in terms of deviation above or below the average of all candidates for each particular subject. Thus they ensure that marks in a subject where there is an unusually wide spread of marks or where a high proportion of candidates score highly (e.g. Theology in the Abitur) do not carry undue weight in the global assessment. The procedure for calculating T scores can be found in standard works on educational and psychological measurement ([23]).

In many examinations of the conventional essay type, e.g. History examinations in the English system, a candidate may be given as many as twelve questions from which he is required to answer no more than three. The purpose of this pattern is clear. It ensures that the whole syllabus

(21) Miles, H.B. and Shipworth, G.E. : *The Times Educational Supplement* 2. 10. 1970.

(22) Petch, J. (1966) : *Marks and Marking*. Manchester. Northern Universities Joint Matriculation Board.

(23) (See, for instance) Guilford, J.P. (1956) : *Fundamental Statistics in Education and Psychology*. New York. McGraw Hill.

is covered in the examination and gives the candidate an opportunity to show what he can do well. To set three compulsory questions only introduces so great an element of chance in favour of the candidate who happens to have concentrated on the areas chosen for the examination that it often appears more as a test of what the candidate can *not*, do, or of his teachers expertise in 'spotting' the questions likely to appear.

Yet the wide choice of questions means in effect that a candidate choosing questions 1, 2 and 3 is answering a totally different examination paper from one choosing 10, 11 and 12 : and a simple calculation of the permutations makes clear how many different examinations are in fact being offered. How then are we to approach comparability between the students taking them ? Research on the reliability of examinations where such a choice of questions is permitted has been carried out at Oxford University and by the National Foundation for Educational Research ([24]).

There are two possible ways of achieving the maintenance of comparable standards from year to year, dependence on the experience and subjective judgement of the examiners and the assumption that with sufficiently large numbers of comparable populations the distribution and level of the qualities which it is intended to assess will not vary significantly from one year to the next. In practice most systems rely on a combination of the two, the examiners having a rough idea of the proportion of candidates to be placed in each grade, but checking this against their subjective impression that standards have risen or fallen.

There are a number of points worth noting here. While it is true that in each succeeding pair of years a marked change in the statistical distribution of grades would be most unlikely to be justified, this might well not be the case over a sequence of a number of years and the change, for better or for worse, might well be so gradual as not to be noticed even by the most sensitive and experienced examiners. This could happen particularly where, through the democratisation of education, the population taking the examination was significantly changed. It is also worth remembering that where the purpose of an assessment is selection rather than evaluation the 'pass mark' will vary from year to year according to the policy of the accepting institutions. Thus a considerably higher mark is now required to secure entry to an English university to read Arts and a considerably lower mark to read Science than was the case ten years ago. The proportion of entrants who passed the French baccalauréat was for a long time more or less static around 60 %. The drop to 50 % in 1966 and the rise to over 80 % in 1968 were the result of outside events rather than of any significant change in the quality of the pupils being assessed.

3.0. EFFORTS TO IMPROVE VALIDITY

Three main criticisms of the validity of traditional techniques of assessment have been made by docimologists in recent years. Of these the most important, because the most fruitful, has been that the assessment is falsified because it relies on the performance of the pupil on a single occasion which is governed by what have come to be known as 'examination conditions'. This has led to a growing interest in techniques of 'continuous assessment'. The other main criticisms have been that the skills and knowledge tested in conventional examinations are not, in reality, the skills and knowledge which those examinations purport to test and that, even if they were, there is some doubt whether these are really the skills and knowledge which are most valuable at a later stage.

One obvious flaw in the single terminal test carried out under examination conditions, is that the pupil may do himself more or less than justice because he, or even more probably she, is affected on that particular day by accidental physiological or psychological factors. These may work either to the advantage or disadvantage of the candidate. There are some pupils who are stimulated by the dramatic atmosphere of the examination room and the single examination, which they treat as a challenge, and who therefore consistently perform better in the examination than they have done throughout the course. It is they, presumably, who cling to all the formalities, including special forms of dress, which have been associated with the occasion. There are others, notably perhaps slow, hesitant and scrupulous thinkers, who find it feverish or claustrophobic and who perform worse. And there are those who have, on the day of the examination, a migraine, or a menstrual period.

Moreover examinations of this kind by their very nature, favour the pupil with the ready pen, for whom the rate of nervous response between brain and fingers is rapid and unimpeded. They favour, in some cases almost demand, the type of memory which can retain not so much facts — we have got beyond that — but prestructured interpretations of facts, long enough to reproduce them in

(24) Backhouse, J. (1971) : *Report to the Schools Council.*

the examination room in legible and coherent form. The pupil who is asked, in the French baccalauréat to write an essay on such a subject as "Un critique a défini Alfred de Vigny 'un cornélien mélancolique' ; qu'en pensez-vous ?" or in the English GCE at Advanced Level to discuss, in twenty-five minutes : "Milton's Satan moves us because he alone is able to convey dramatically what goodness is" has no recourse but to his memory of what either he or his teacher thought about the subject on a previous occasion and to the facility of his pen. He has, in the English system at least, no time to think, and if he had he cannot refer to the text, as would anyone seriously considering such a question. It is not surprising that in such circumstances the examiners regularly complain that candidates answer, not the question which has been set, but a more or less close approximation to it, which they have prepared before entering the examination room. The skill lies in disguising this memorised reasoning as a response to the question on the examination paper. One device intended to mitigate the reliance on memory which has been tried recently in a number of systems is to allow pupils to bring reference books into the examination, but whatever value this may have it does not eliminate the effects of requiring pupils to work to a strict time limit on a single occasion.

The harmful backwash effects of such examinations in promoting the tacit acceptance of received ideas and opinions, rather than the exercise of thought, has been as much criticised as their failure to assess the qualities which they purport to assess. Before we condemn them altogether, however, in favour of some more continuous form of assessment it is worth considering why they have enjoyed the long ascendancy which, in some countries, is only now being challenged.

There is some justification for the view that the skills and qualities which enable the first type of pupil described, the 'good examinee', to perform well in such examinations are important and should play their part in any assessment. Many pupils would themselves pay tribute to the value of such examinations as incentives. The case for this view was well stated as long ago as 1911 in the Report of the Consultative Committee on Examinations in England and Wales, which includes among the good effects of terminal examinations on the pupil "that they train the power of getting up a subject for a definite purpose, even though it may not appear necessary to remember it afterwards — a training which is useful for parts of the professional duty of the lawyer, the administrator, the journalist and the man of business" (25).

There seems to be no justification, however, for the view that the encouragement and measurement of these abilities should play the dominant part in assessment, and it is partly the realisation of the undue importance which they have assumed, and in many systems still assume, that has led to the demand for more continuous assessment. After all, educators are not engaged simply in training and selecting lawyers, administrators, journalists and men of business and even in the academic world there is a growing recognition that there is no very high correlation between success in terminal examinations of this type and capacity for continued research.

The real reason why these examinations continue to play so large a part in our process of assessment at the most decisive point in the educational process is surely their reliability and their demonstrable impartiality. Impartiality, which is demonstrable to parents and students, is of crucial importance at any point where life chances depend on acceptance or rejection, and the reliability of this type of examination, though low as we have seen, is still higher than that of most techniques of continuous assessment. If we are to replace them with some method of continuous assessment which does not suffer from their inherent weaknesses, we must try to preserve within the new form of assessment as much as possible of the virtues of the old.

3.1. *Continuous assessment*

Clearly one way to avoid the weaknesses of the single terminal examination as a method of assessment is to spread the process over a considerable period of time. The term 'continuous assessment' has been used somewhat loosely, however, and it will probably be convenient to distinguish three main usages.

The first, which we may call periodic assessment, substitutes for the single terminal examination a number of tests taken on different occasions throughout the course. This is not unlike the 'course credit' system in American universities, with its regular tests and grades at the end of each semester contributing to the final award of a degree. In Europe the German Abitur and the Swiss Cantonal Maturité have long given weight to this type of continuous assessment.

(25) *Report of the Consultative Committee on Examinations in Secondary Schools* (1911). London. HMSO.

In Germany the procedure may differ slightly from one Land to another, but in a typical system the assessment of 'Klassenarbeit' will be based on a series of tests spread over the last two years of the secondary course (e.g. one in the middle of the penultimate year, one at its end and one in the middle of the last year). These tests are given by the teacher in each subject and their administration is closely controlled. In some Länder, for instance, the number of tests to be given is specifically related to the number of hours per week for which the subject is studied, or the tests may not be given on a Monday, nor two tests on the same day. The cumulative grades for these tests establish the 'Vornote', which may carry as much as half the weight in establishing the final grade in the Abitur. In the Swiss Cantons the standard procedure is to give half the weight in establishing the grade in the Maturité to this periodic assessment.

Periodic assessment of this type almost certainly improves validity by eliminating or mitigating some of the physiological or psychological hazards of a single terminal examination ; but it is still open to the criticisms which can be justly levelled against the traditional written or oral examination as a measuring instrument. The fact that this instrument may be used on as many as five occasions eliminates some of the reasons which lead us to question its validity when used on a single occasion, it may even improve slightly its reliability ; the problem remains however that what is measured may be recall rather than either understanding or creativity. And because it is administered and marked by the teachers in each school or college questions immediately begin to arise as to its impartiality and comparability between one school and the next. It is perhaps significant that whereas in Switzerland the Cantonal Maturité relies to this extent on continuous assessment because the teachers making it will be employed by the Cantonal authorities and known to the parents, the Federal Maturité, which is required from pupils in independent schools, does not.

The second type of continuous assessment, which seeks to break away from the restrictions of single occasion tests altogether, relies upon a general assessment of the student's work in each subject over the whole of the course. This we will call cumulative assessment. Cumulative assessment makes it possible to avoid all the pitfalls of the single occasion test, whether used terminally or periodically, but it increases still further the subjective element in the assessment, since in most systems the teacher is the only judge who has been in contact with the pupil throughout the course and must therefore accept responsibility for the assessment.

An interesting variant of cumulative assessment which spreads the responsibility more widely and also introduces a certain element of self-assessment has been developed in connection with the assessment of projects in which a number of pupils have collaborated over an extended period. Here the problem is not merely to assess the success of the project as a whole, but to assess how much each individual in the group has contributed to it. The technique is to ask each pupil and each teacher who has taken part in the project to allot a grade to each participant, including, in the case of the pupils, himself. In order to avoid over or under-marking based upon prejudice, a device used in International Games, when panels of national judges assess diving or skating, is adopted : from each series of grades the top and the bottom grade are eliminated and the final grade is the mean of all the others. This method has been tried experimentally in the assessment of architecture students, both in England and Germany, and reports upon it are encouraging (26).

Periodic and cumulative assessment are commonly used together as in the German Abitur, where the teacher's assessment based on the Klassenarbeit not only contributes to the final Abitur grade but is also decisive in the important question of whether a pupil is ready for promotion to the next class or must repeat the year's work. Whereas in the English and French systems approximately 35 % of all candidates fail the single terminal examination, in Germany failure is minimal because a process of continuous assessment has ensured that candidates do not take the examination until they are ready to pass it. Thus many of the tensions and injustices of the baccalauréat and the GCE are avoided.

It has been noted above, however, that improvements in one feature of assessment are often paid for by concomitant disadvantages in another. The price of the long period of continuous assessment in the years leading up to the Abitur has been a steady rise in the average age at which it is taken, until it has now reached 20.5, with a very high rate of drop-out on the way. H. Peisert and R. Dahrendorf, working with a total cohort of 5383 children entering the first year of the Gymnasium course, found that this had been reduced to 1579

(26) Oxford Polytechnic (1970) : Department of Town Planning. *Occasional Papers No. 2.*

on entry to the penultimate year, of whom 1236 finally passed the Abitur ([27]).

Combinations of periodic and cumulative assessment naturally vary according to circumstances but the following regulations for assessment in the Diploma in Education of Bristol University give some idea of the process where examination is wholly internal :

> "Assessment of course work is used instead of traditional final examinations. Normally this means that for each course studied, a written assignment will be required. Assignments take various forms : longer essays, a series of short papers, seminar papers, case-study reports, etc., at the discretion of individual tutors. It should be noted that course work assessment does not preclude the use of tests of either the traditional essay or the short answer type. Such tests, if used, are given during the session.
>
> Assignments are so phased as to spread the work load throughout the session. They must be completed by the set date, and when returned by tutors, must be kept together in a folder, for submission to the examiners at the conclusion of the course.
>
> Systems of assessment are of necessity comparative, and it is not always easy for an institution to cater with any real refinement for individual differences. In addition, complete objectivity is not possible ; in marking work, tutors attempt to take account of the following factors :
>
> — The quality of the work submitted in relation to the rest of the group ;
> — The quality of the work in relation to the tutors' assessment of general standards in the subject.
>
> Inherent in the evaluation problem are factors of an affective nature which are complex and variable; and an attempt is made to take account of these factors in the assessment of students' work.
>
> Obviously, different tutors have different marking styles and produce different distributions of scores. This is covered very largely by a final assessment meeting at which not only grades but also ranking orders are considered in reaching a final overall grade for each candidate.
>
> Two further points should be made :
>
> - It is unfair to both tutors and students if work schedules are not adhered to. Only in special circumstances may work completion dates be postponed.
> - Attendance below the 70 % minimum for *any of the courses* is not acceptable and will be considered as constituting failure of the total course" ([28]).

The examples quoted so far have illustrated the combination of periodic or cumulative assessments with each other or with terminal examinations, when the same examiners are essentially responsible for both. A typical example of this process where the terminal examination is in the hands of external examiners is the use of the 'livret scolaire' in the French baccalauréat. This cumulative record book, in which are recorded the teachers assessments of course work over the year, is made available to the examiners, who take it into account, both in confirming the immediate acceptance of those who have scored a mark of 12 or above on the 20 point scale in the first group of two written and two oral examinations, and to decide borderline cases both for admission to the second group of examinations and for final acceptance or rejection after their completion.

In the International Baccalaureate teachers' assessments are made available to the examiners for consultation after the preliminary marking, as an additional check on the grading, a practice also used by some of the English GCE Boards.

In the examinations for the Higher National Diploma (a professional qualification for engineering and kindred professions) in England 30 % of the weight in the final grading is given to periodic assessments, carried out by the teachers sometimes on as many as twelve occasions, and the remainder to terminal examinations on syllabuses submitted by the teachers, but controlled by assessors, as in the German Abitur. It is noteworthy that in this, as in the Bristol example quoted above, or in American universities, a minimum attendance rate, in this case 67 % of all possible attendances, is required. Such a proviso is of course unnecessary in secondary schools, but it is difficult to see how cumulative assessment, at least, could be justified without such a requirement, which might be hard to enforce, in many areas of tertiary education.

(27) Peisert, H. and Dahrendorf, R. (1967) : *Der Vorzeitige Abgang vom Gymnasium.* Kultusministerium Baden-Württemberg. Reihe A Nr 6.

(28) Bristol University (1969) : *Faculty of Education Prospectus*, p. 17.

In periodic assessment it is, of course, possible to avoid specific tests and to assess the quality of pupils' work on laboratory note-books, extended essays or portfolios representing work done during the course, without controlling attendance at classes, but this clearly involves both a risk of cheating in the form of work which has not been done by the pupil at all and a very fine decision as to the extent that a pupil may legitimately be 'helped' in its preparation, for instance by an elder sister at the university. The best control here, which should certainly avoid deliberate cheating, is probably an oral examination on the work submitted, but it is in the nature of continuous assessment that in practice such oral examinations could only be carried out by the pupils' own teacher.

The reference to "factors of an affective nature" in the Bristol regulations leads us on to the third usage of the term 'continuous assessment', to indicate a complete assessment of the whole personality and record of the pupil. This, which would clearly provide the best criterion either for orientation or selection, we shall call "global assessment." It is, however, manifestly much more subjective and therefore unreliable than even cumulative assessment. Consequently it is at present used only by independent colleges whose assessment cannot be questioned on grounds of partiality, or, in official systems, for orientation rather than selection.

3.2. *Orientation*

In an ideal educational system, unaffected by either economic limitations, academic snobberies or the shortage and inexperience of teachers, all assessment would be for orientation rather than selection. For such a purpose a global assessment of the whole personality of the pupil would no doubt provide him with the best guidance in plotting his course through life. This ideal has long been recognised, particularly in France, where the term 'classe d'orientation' dates back to 1937.

In reality, however, educational systems are only able to approach towards this ideal where pupils are moving from a less to a more differentiated stage within the period of universal education. Even then the pure concept of orientation is compromised and selection enters in as soon as one of the different channels in the second stage is attracting, for whatever reason, more applicants from the first stage than it can absorb.

In the current European situation, therefore, we find the nearest approaches to a genuine system of orientation at the point of transfer from lower to upper secondary education rather than from upper secondary to higher.

The problems which such orientation poses have been very clearly set out in a recent report of the Institut Pédagogique National in Paris ([29]). Whether the actual advice on orientation is to be given by a school counsellor or by a commission, it has to be based on a dossier which records not only the social and family background of each pupil, but an assessment of his qualities by all those teachers who have been mainly responsible for teaching him. This means, as R. Gal has said, that the teacher must become a psychologist, and many teachers who are excellent teachers of Latin or Mathematics, and who can judge very well a pupil's performance in Latin or Mathematics, are ill-equipped for this wider function.

In the design of a 'dossier scolaire' there is a constant conflict between the points of view of the administrator who wants it to be a complete and tidy record, the teacher who wants it to be easy to fill in and to deal with, the counsellor or commission who, perhaps dealing with several hundred under pressure of time, want it to be short and contain only essential information, and the potential research worker who wants it to contain all conceivably relevant facts. The report quoted above gives two specimen dossiers one of twelve and one of twenty-two pages and appends some of the criticisms of those who used them. The first comment of a teacher, on even the shorter dossier, was that a packet of 35 or 40 dossiers for a single class made a heavy and cumbersome load to carry home. The criteria for a good dossier suggested by a 'user' were also significant :

- It should be short ;
- It should give a synthetic view but without neglecting important details ;
- It should stress the 'particularities' of the pupil.

It is important to remember that both teachers and 'assessors' are human beings with human limitations and that there is as much danger in demanding from them, or feeding to them, too much information as too little. It is presumably possible to envisage so structuring the information which is required that it could be coded and processed by a computer, but this would still require a highly

(29) Institut Pédagogique National (1969) : *L'Orientation Scolaire et la Recherche des Aptitudes.* Paris. Service d'Edition et de Vente des Productions de l'Education Nationale. Brochure N° 34.

skilled assessment by the teachers in the first place.

In so far as the teacher can be trained or train himself as a psychologist this type of global assessment seems possible, but it will undoubtedly be time-consuming, even if the professional psychologists come to his aid with batteries of tests. Moreover, it increases the individual teacher's responsibility, and as long as the advice based on his assessment has any mandatory force on the pupil there will be a need for some recourse to a second opinion. Otherwise the emotional pressure on teachers from pupils and parents is likely to become intolerable.

This recourse is provided in France by the fact that orientation to the Lycée, the most favoured channel, is the procedure only within the state system. Pupils from independent schools have to pass an entrance examination, and it is left open to pupils from the state system who have not been 'oriented' to the Lycée, and who disagree with their orientation, to take this examination also and, if successful, to enter. This device of providing a system of continuous assessment, with an examination open to those who disagree with the assessment, is one which might prove useful elsewhere.

The procedure followed in Sweden is very similar, except that standardised achievement tests play a large part in deciding the advice on orientation which the teachers give to the parents. These tests are, however, characterised by Henrysson as being 'monitorial' only in character and "it is considered to be of value for teachers in their marking to include qualifications in their students that cannot be evaluated by tests" ([30]).

The advice given by the teachers at the point of entry to the pre-gymnasium stream in the comprehensive school is not mandatory on the parents and whatever the views of the school authorities any parent may insist on his child entering this stream. Up to the age of 16 therefore we have here orientation in its pure form. At this point however selection enters in. The proportion of the age group in the pre-gymnasium stream over the last three years has been 45 % in 1967/8, 44 % in 1968/9 and 42 % in 1969/70 ([31]); yet by the terms of a Parliamentary decision the maximum proportion which can be admitted to the gymnasium is 30 %. Selection is therefore based on the applicant's school marks as 'monitored' by the standardised national tests ([32]). It seems possible that the gradually falling proportion in the pre-gymnasium stream represents a growing tendency on the part of parents to accept the teachers' advice at the earlier stage.

England and Wales are the only European countries which make use of a nationally organised examination of the conventional type, the General Certificate of Education at Ordinary Level or the Certificate of Secondary Education, for orientation at this point and some small elements of continuous assessment are now beginning to appear in these examinations also. If the recommendations of the Schools Council that these two examinations should be merged and should no longer play any part in University selection are implemented, it seems likely that the element of continuous assessment will increase.

The French experiments seem the most ambitious attempt yet made at orientation through global assessment in a national system, but, as Legrand points out in the introduction to the report quoted above, there are still very serious problems to be overcome, both in the training of teachers as assessors and in the conceptual analysis of the qualities to be assessed.

3.3. *The reliability of continuous assessment*

The advantages of continuous assessment, whether periodic, cumulative or global, in terms of validity, backwash and speed are apparent. It is not surprising therefore that much effort is now being given to improving its reliability. The most notable example of this is undoubtedly the new system which has been adopted in Sweden, both for orientation within secondary education and admission to higher education.

In Sweden the studentexamen, which previously closely resembled the German Abitur, has been replaced by continuous assessment of a cumulative type based on the school marks in a range of subjects. The procedure introduced to improve the reliability of this assessment and particularly the comparability between one school and the next is of great interest.

The first step was the establishment of standardised national achievement tests at the upper secondary level. This was done by asking gymna-

(30) Marklund, Henrysson and Paulin (1968) : *Kompetensutredningen* III. Stockholm. National Education Board. (English summary).

(31) *Statistika Centralbryan Reports* 1968 U2, 1969 and 1970 U5. Stockholm. National Education Board.

(32) Orring, J. (1969) : *School in Sweden*. Stockholm. National Education Board.

sium teachers to submit items for their respective subjects. These items were then reviewed and the most promising pre-tested in the most rigorous fashion, including, for instance, the pre-testing of language items in the countries where the language concerned was the mother tongue. The tests were then standardised over a large and representative sample of the Swedish school population. It is significant here that Henrysson reports that rural schools gave as good results as those in the town.

From these tests a national mean performance at each stage in the upper secondary course is calculated and it is assumed that results throughout the country, if related to the national mean, will correspond to a pattern of normal distribution. These tests are then sent to each school where they are administered and scored by teachers. Tests at the gymnasium level have now been devised by the central research unit for Swedish, foreign languages, mathematics, physics, chemistry and economics, i.e. the subjects normally presented in Europe for written matriculation examinations. Where the tests are wholly of the 'objective' type scoring is comparatively simple, but in languages, as in the USA it has been found desirable to require some piece of extended writing. In order to assist teachers in assessing this, corrected scripts from the standardising sample are sent to teachers who, if they are still in doubt after consulting them, may call in a second examiner.

The purpose of these standardised tests is, as we have seen above 'monitorial'. They do not determine each pupil's final assessment, but from them the teacher can judge how the general standard of his class and the intervals between pupils within it conform to the national norm. He is expected to make other assessments of his pupils' work over the year, but when at the end the final assessments are translated on to a five point scale the teacher takes into account the pattern established within his class by the national tests in determining the distribution of his grades.

Let us suppose, for instance, a class of 30 pupils whose raw scores in a certain subject range from 83 to 15 with a mean of 48 on a national test for which the national average is 53. Normal distribution of the grades would be as follows :

Grade	5	4	3	2	1
Number of Pupils	2	7	12	7	2

Looking at his raw scores the teacher sees that he has two pupils well ahead of the rest with scores of 83 and 82. Although his experience tells him that on the whole the class is rather below average and this is confirmed by their score on the national test, these two clearly deserve a mark of 5. Similarly he finds a group of seven between 77 and 62 who are clearly ahead of the next candidate at 59. He retains the normal distribution therefore for the mark of 4. Below that he has less than would normally be expected bunched about the national mean and so his grade distribution for this class comes out as below (normal distribution in brackets) :

Grade	5	4	3	2	1
Number of Pupils	2 (2)	7 (7)	9 (12)	9 (7)	3 (1)

Examples of possible distributions quoted by the National Board of Education for different types of class of 30 are given below ([33]) :

Mark

Nature of class	5	4	3	2	1
Average	2	7	12	7	1
Good	4	9	9	7	1
Poor	1	4	12	10	3
Even	1	7	14	7	1
Uneven	4	7	9	6	4

Considerable liberty is therefore left to the teacher to distribute the grades in accordance with his own assessment of the whole of his pupils' work but in doing so he has their performance in the national tests as a 'monitor'. Similarly the whole results from a particular school may be above or below the national norm, but it is the business of the inspector to ensure that if this occurs there are reasons which justify it.

This system, which is still in an experimental stage, has several important advantages. It introduces an objective control into continuous assessment and it ensures that the spread of grades is comparable as between different subjects. This is particularly important in Sweden where a pupil's final assessment is expressed as a sum of his grades and the difference between an average grade of

(33) National Education Board (1968) : *The New Gymnasium in Sweden*. Stockholm. p. 52.

4.3 and one of 4.5 may imply acceptance or rejection by the most exclusive faculty of the university.

Neither this nor the use of standardised tests in the comprehensive school, however, is proving wholly popular and there are still problems involved in the combination of objectivity with continuous assessment which have not been solved. It is easy to see for instance the resentment which might be caused under this system when identical individual performances on a standardised objective test are translated into different grades on the five point scale because of differing distribution patterns in the classes.

Moreover, the National Board of Education have found that, in the competitive situation which persists with regard to entry to the most favoured channels or faculties, continuous assessment has not substantially reduced the competitive pressure for which the studentexamen was blamed. Issue number 3/1970 of the Council of Europe *Newsletter* reports as follows :

> "The National Board of Education and the Ministry have declared their intention to reform the present grading system within the comprehensive school. The grading system has been the subject of considerable criticism, in some cases resulting even in the boycotting by pupils of standardised achievement tests.
>
> It is generally admitted that the present system of grading has several defects. There are many who have come to the conclusion that these are so serious that an entirely new approach will have to be adopted to find, in particular, more adequate selection methods for restricted intake lines. Above all, it is felt that the present grading system counteracts cooperation and collaboration between pupils, contrary to the goals of the comprehensive school.
>
> A decision is soon expected to be taken to greatly reduce grading within the comprehensive school. Experimental and investigational work is being done to obtain a basis for such reforms. Suggestions and ideas from both pupils and teachers being discussed at present in periodicals and newspapers, will also be taken into account" ([34]).

Experiments are also being carried out in England to determine and to improve the reliability of continuous assessment. In 1964 the Department of Education and Science gave a grant to Leicester University "to study ways of examining other than by conventional written papers." In introducing the record of their experiments which were carried out in the context of the certificate of secondary education Professor J. Eggleston writes : "provided the assessments arise directly from an adequate specification of the educational objectives, the continuous evaluation of attainment has important consequences. These include more emphasis on immediate and intermediate gains rather than on terminal or more remote outcomes ; higher levels of student motivation ; and encouragement to use a wider variety of teaching methods" ([35]).

The emphasis here laid on the specification of objectives was reflected in the experimental method. Taking the physics test as an example, teachers were first asked to specify the objectives of their course, as in the IEA experiment quoted earlier. These were then compared with the examiner's specification and an agreed list prepared consisting in this case of 'Inference', 'Organisation of Data', 'Observation', and 'Application of Facts and Principles to problem solving'. The procedure for assessment was then as follows :

- The teacher decides that pupils will exercise a particular ability drawn from the list during a specific lesson.
- The teacher then organises the lesson so as to draw a response requiring that ability, preferably in writing and 'under test conditions', during the lesson.
- The teacher grades these responses on a five point scale.

This was done on an average three times over an eight week period. Teachers agreed that although this kind of assessment was difficult it could be improved by practice and did not interfere with good teaching. The assessments were then checked against a 'moderating test' of the 'objective' multiple-choice type based on the same objectives. The teachers were warned of the 'halo' effect and asked as far as possible to isolate the specific ability being tested.

Three indications from this test may be of general interest :

— Although the average correlation with the objective test was reasonable (0.45), some teachers achieved markedly worse correlations than others and there was some indication that even a small amount of training would improve results.

(34) Documentation Centre for Education in Europe : *News-Letter* 3/70. Strasbourg. Council of Europe. p. 16.

(35) Eggleston, J. and Kerr, J. (1969) : *Studies in Assessment*. London. English Universities Press. p. 3.

— The intercorrelations between the grades on the tests of separate abilities were much higher on the teachers' assessment (between 0.44 and 0.54) than on the objective test (between 0.14 and 0.34), indicating that teachers conducting continuous assessment were considerably less capable of 'isolating' specific abilities than test constructors.

— The test constructors complained, as usual, that the teachers did not use the full range of the scale : to which the teachers replied that 'pupils in a streamed class might well represent a band of the spectrum of ability so narrow that differentiation even on a five point scale might be difficult'.

All these seem to point in the direction that continuous assessment by teachers requires special training and is more appropriate for certification or orientation than for selection.

In the test of practical work in Biology they found the reliability of cumulative assessment greater than that of periodic and of the same order as in team-marking of essay type questions. The introduction of a 'control' in the form of a number of questions in the written papers which referred to practical skills, a device now used by those GCE Boards which rely on teachers assessments for the measurement of practical work, was also investigated and regarded as providing reliability at an acceptable level for a test carrying not more than one third of the total weight of the assessment.

The general conclusion of the investigating team for Physics was : "When the extreme cases of disparity between grades were examined across all three assessment procedures (e.g. continuous assessment by teachers, objective test and conventional examinations) there was no evidence to make a case for one being a more accurate assessment of attainment in Physics than another. The decision to use one method of assessment rather than another will presumably depend on *a priori* assumptions about the known or presumed backwash effects of the procedures" ([36]). For the purposes of assessments made at a point of transition within a full secondary education, and concerned only with certification and orientation, this may well be true. It may be argued on the other hand that the cost of these elaborate teacher-controlled procedures in training, attendance at meetings and assessment is greater than the outcome justifies, and that the increased involvement and enthusiasm which they generate as an experiment would evaporate when they became a routine.

(36) Eggleston, J. and Kerr, J. : *Op. Cit.* p. 76.

3.4. *Analysis of objectives*

The experiments reported above laid great stress on the clear specification of objectives as an essential preliminary to assessment. Unless we are to be concerned solely with predictive validity, it seems clear that in order to make a valid assessment of what gains in knowledge, skills or attitudes pupils have made from following a course, we do need to know what knowledge, skills or attitudes that course is seeking to develop. Since the publication of B.S. Bloom's important work in 1956 this has often been known as a taxonomy of educational objectives ([37]). But the analysis of objectives can be a fruitful method of assessing the value of a course as well as the progress made by individual pupils. If a course at upper secondary level leading to a public assessment, whether continuous or by single occasion examinations, is intended to prepare pupils for the next stage in their lives, then we need to know not only whether the method of assessment is testing reliably the objectives of the course, but also whether those objectives are really relevant to the needs of the next stage.

That this is a useful exercise for those responsible for course and assessment planning seems beyond doubt, but it must be realised that it raises both the profoundest issues of educational philosophy and the most practical problems of administration and teacher training. Some of the teachers who took part in J. Eggleston's experiments were strongly of the opinion that a consideration of the methods of assessment was of great value in helping to clarify objectives.

On the other hand the historians were so convinced that the use of original sources was important as a teaching method that they finally abandoned any attempt to relate items in their tests to specific objectives and preferred to make the use of original sources the criterion of a good assessment procedure. The outcome, in their opinion, was that 'the children are better historians as a result'. But the question whether the objective of teaching history to fourteen and fifteen year olds of average or less than average ability is to make them 'better historians' raises profound issues of educational philosophy, to say nothing of semantics, and ought not, perhaps, to be left to the history teachers. Most teachers seemed to agree that objectives in what Bloom terms the 'affective domain' were of great importance, but also that it was difficult, if not impossible, to assess how far they were being achieved.

(37) Bloom B.S. *et al.* (1966) : *Taxonomy of Educational Objectives*. New York. David McKay.

This very understandable view, combined with the view that consideration of assessment procedures helps to clarify objectives, raises the doubt whether we are not faced with the risk which Desclos saw in objective testing — that in our attempt to break down the objectives of education into specific, operational and testable items of behaviour, we shall produce assessment procedures which are no more valid, because we have eliminated from our courses those elements, often the most valuable, which are difficult or impossible to assess. As one of the most distinguished empirical researchers has said, the relative emphasis on the knowledge function in comparison with the non-knowledge components in school courses is 'a question of educational philosophy and not very well suited for empirical research' ([38]). If this is so, it may well be important that the analysis of objectives should not be left too much in the hands of the specialist teachers or empiricists with a new found enthusiasm for taxonomies, and should quite clearly precede the design of the assessment procedure. The argument that assessment should always relate directly to the course which the teachers are teaching, and not to the course which the examiners think they ought to be teaching, should not arise in a system in which both teachers and examiners are working within a framework of agreed objectives.

In drawing up such a framework or taxonomy, few would seriously quarrel with the kind of hierarchy of learning at the basis of Bloom's work, at least in the cognitive area, but its application to course and assessment design is by no means simple. From a teaching point of view a Piagetian approach might lead us to advocate a much deeper analysis of the stages, including the tension between established schemata and incoming experience, through which a particular level of cognitive ability is best developed. This might lead to the specification of objectives in terms more of a 'spiral', to use J. Bruner's term, than of a series of levels, and so point in the direction more of continuous assessment than of sampling on single occasions. On the other hand, we know from experience, confirmed by Eggleston's experiments, that teachers find it extremely hard to differentiate their assessment into any but the simplest categories. 'Observation' may be, as Legrand has pointed out, a far from unitary activity, yet the teachers in Eggleston's experiment clearly found it quite difficult to assess it separately from even three other widely different 'objectives'.

(38) Dahllöf, U. (1963) : *The Contents of Education with regard to demands for different jobs and for further studies.* Stockholm. National Education Board.

Although much more fundamental work needs to be done in clarifying the objectives of educational courses and relating them to assessment, there are clearly certain easily identifiable skills, in language, for instance, or mathematics, where a more detailed definition of the level expected and the weight to be given to each area of performance might contribute to improving the validity of assessment.

In such a detailed taxonomy attention should surely be given to the second problem, the relevance of skills taught at one stage to the needs of the next. An exhaustive study of this problem was carried out by the Swedish National Board of Education in connection with the reform of the gymnasium and is summarised by Professor U. Dahllöf ([39]). The procedure was to break down the curriculum of the gymnasium into sixty-five separate areas of content and fifteen separate "general study-skills". University professors and industrialists were then asked to rate on a five point scale both the importance of each of these for further study in his own field or for employment, and the degree to which the present preparation of gymnasium students was adequate. The results from the universities emphasised the importance, in the area of content, of foreign languages (especially English and German), elementary mathematics and statistics, among which German and statistics were found to be inadequately treated. Among study-skills the greatest emphasis was placed on "rapid reading in order to identify the main points in the text, the making of notes, the collection of information from a library and its use in an essay or memorandum." These skills may seem rather obvious, but in how many upper secondary courses is their development a conscious objective or significantly tested in the assessment ?

4.0. SCHOLASTIC APTITUDE TESTS

Some of those who have been primarily concerned with the predictive validity of current assessment procedures rather than their validity as certificates of secondary education, have turned their attention to the possibility of using tests of aptitude for higher studies rather than of achievement at an earlier stage.

It might reasonably be assumed that, provided there is a good 'fit' between the content of upper secondary and university courses, success in the first would be a good predictor of success in the second and therefore a good criterion for selection. The whole process of selection at this point in the

(39) Dahllöf, U. (1963) : *Op. Cit.*

educational process has in fact been based on this assumption throughout Europe. Recent research in a number of countries has thrown considerable doubt upon it. Dr. Bagg, in England, working with students of chemical engineering and comparing GCE Advanced Level results with university first year and final examinations found a poor and decreasing correlation between them and concluded that "only 13.5 % of the variance of the Part II Finals marks can be accounted for by the three 'A' Level results taken together" (40). Other studies in England have shown a fair correlation between 'A' Level and first year examinations but little between 'A' Level and final university examinations.

Similar research in Germany appears to point in the same direction. A study reported by E. Weingardt of correlations between average Abitur grades and the 'Erster Lehrerprüfung' in four Pädagogische Hochschulen showed correlations of just over 0.40 in three cases (although only 0.29 in the fourth), but correlations with the 'Zwischenprüfung' in Science in five Universities ranged between 0.18 and 0.37 and the correlation between the Abitur grade in Chemistry and the final university examination in Chemistry in one study was 0.06 which is even lower than Bagg's finding for Chemistry in England (41). Orlik found only slightly better correlations between Abitur grades in individual subjects and university success in that subject with actually a negative correlation for medical students (42).

These results are confirmed by a much more extensive survey of 174 different studies of correlation between school or examination marks and 'success' in higher education in the Scandinavian countries carried out by Marklund, Henrysson and Paulin. In these studies involving approximately 30,000 pupils a mean correlation of 0.27 was found between matriculation marks and results in different kinds of further education using a wide variety of criteria. This accords well with other European findings. Findings in the USA may approximate to 0.50 but, as the Swedish report points out there is a far greater degree of similarity, both in structure and methods of assessment, between senior high school and university first degree work in the USA than in Europe. This report, which includes a valuable survey of the problems involved in defining 'success' in tertiary education, also records a mean correlation of 0.28 between the results of test batteries and results in further education, using the same criteria as for matriculation marks (43). In view of these almost exactly equal correlations and of the adverse backwash effects of matriculation examinations already referred to, it is not surprising that a number of European countries are now investigating urgently the employment of such tests for university selection.

In compiling such scholastic aptitude tests they are able to draw on the considerable experience of educators in the United States. The assumption that success in upper secondary courses, in so far as this was validly and reliably assessed by matriculation examinations, was a good predictor of success at the University depended on a large degree of University control of both courses and examinations in the upper secondary school. This, of course, has prevailed in Europe but not throughout North America. Consequently the need for tests, both of scholastic aptitude and achievement which were not dependent on a prescribed syllabus was felt all the sooner in the North American context.

The College Entrance Examination Board has been conducting Scholastic Aptitude Tests since 1926. The present tests are three hour objective tests designed to measure the development of mathematical and verbal skills. They are used by American Colleges as one element in a battery of assessments, on which they base their selection procedure. Because they have been in use for so long and have the financial backing of a continent-wide system they are based on cumulative experience and very substantial research. They are designed, however, for a population rather younger and also less homogeneous, both in social and academic terms, than that which is now seeking entry to higher education in Europe.

Typical questions from the American College Board Scholastic Aptitude Tests are given below :

1. "As warfare has come to engross an increasing proportion of the belligerant populations, so military ... has grown far beyond the problems of varying terrain.

(40) Bagg, D. (1968) : The Correlation of GCE 'A' Level Grades with University Examinations in Chemical Engineering. *British Journal of Educational Psychology*. Vol. 38. Part 2. June 1968.

(41) Weingardt, E. (1968) : Der Voraussagewert des Reifezeugnisses für wissenschaftliche Prüfungen. Roth, E. (ed.) : *Begabung und Lernen*. Stuttgart. Klett. pp. 433-447.

(42) (Reported by) Flitner, A. (1966) : Das Schulzeugnis im Lichte neuerer Untersuchungen. *Zeitschrift für Pädagogik*. Jg 12. Heft 6. December 1966.

(43) Marklund, Henrysson and Paulin (1968) : *Op. Cit.*

(A) custom
(B) life
(C) history
(D) strength
(E) geography."

This is a relatively difficult question. An analogy is drawn between the general expansion of warfare and whatever it is that is indicated by the missing word. A careful examination of the sentence should indicate that the missing word is somehow related to the problems of "varying terrain" and its effects on some aspect of military operations. Of the five choices, (A) and (C) are obviously incorrect ; military custom and military history have changed and grown and were never limited in their concern to problems of varying terrain. (B) and (D) seem plausible ; military life and strength do involve problems relating to terrain, but this is not their major concern. Of the five alternatives given, only (E) *geography* relates specifically to terrain, so this is the correct answer.

2. "If $x > 1$, which of the following increase(s) as x increases ?

I. $x - \frac{1}{x}$

II. $\frac{1}{(x^2 - x)}$

III. $4x^3 - 2x^2$

(A) I only, (B) II only, (C) III only, (D) I and III only, (E) I, II and III."

This question is slightly above average in difficulty and requires numerical judgement in a relatively new situation. Two principles must be understood and applied in this problem : (1) If the denominator of a fraction increases while the numerator remains constant, the entire fraction decreases (2) If x is greater than 1 and increases, then x^n increases more rapidly than x^{n-1} ; i.e., x^2 increases more rapidly than x and $4x^3$ increases more rapidly than $2x^2$. Thus one can show that expressions I and III increase as x increases, whereas II does not. Therefore the correct answer is (D) ([44]).

Another type of item used to test verbal skills is the provision of a relatively sophisticated piece of prose which then forms the basis of items designed to test comprehension. Scholastic aptitude tests are now being investigated in England for the Committee of Vice-Chancellors and Principals, in Germany by the Planungsgruppe Pädagogische Diagnostik sponsored by the Volkswagen Stiftung and in Sweden for the National Board of Education. The intention is clearly to use these as supplementary, rather than substitute, methods of assessment in order to reduce the dominance of traditional examinations or of periodic assessment. The promoters of the English project say that "it is hoped that if the test results prove their worth they will eventually relieve the pressure on the schools to concentrate on high achievement in A levels to the exclusion of other aspects of the curriculum" and the Swedish report referred to above : "Short-sighted striving for high marks may easily eclipse the more long term work of development of personality in the school. This experience is by no means new, but its significance has been greatly underestimated. If, in addition to school marks, other instruments are used in selection, the derogatory effects will be reduced."

Typical questions from the English *Experimental Test of Academic Aptitude* ([45]) designed for use at the European university entrance level are quoted below :

1. "*Directions.* In each of the following questions, a related pair of words, printed in capital letters, is followed by five pairs of words lettered A to E. Select the lettered pair which best expresses a relationship similar to that expressed in the pair printed in capitals :

REFORM	:	PROGRESS
A conscience	:	virtue
B change	:	results
C correction	:	improvement
D exercise	:	health
E legislation	:	welfare."

2. "*Directions.* Solve the following problems. You may use any blank spaces for rough work. Answer each question by marking the appropriate lettered space on your answer sheet :

The miles per gallon of petrol obtained by a certain car falls uniformly from 40 at 40 m.p.h. to only 20 at 80 m.p.h. How many miles per gallon will be obtained at 52 m.p.h. ?

A 33
B 34
C 35
D 36
E 37."

(44) College Entrance Examination Board (1965) : *Op. Cit.* pp. 13 and 26.

(45) Association of Commonwealth Universities (1970) : *Experimental Test of Academic Aptitude.*

Another type of question used in these tests is to provide a question and two statements, the candidate being asked, not to answer the question, but to state whether the data provided in either or both of the statements is sufficient to answer it.

The report of the German commission is already completed and pupils who took the first batch of English tests before entry to University will be completing their first degrees this year, so that a considerable amount of evidence about their predictive validity should be available very shortly. If it should prove that this is as high or higher than that of assessments, whether single or periodic, of achievement, the case for using them in selection will be strong since they have great advantages in cost and speed ; but it may well be that they test different qualities from those tested in achievement tests. Since achievement tests are likely to be retained, both for certification and for their incentive value, the decision to consider them as supplementary forms of assessment will probably be maintained. This is also the recommendation of the Swedish survey which suggests that there are more opportunities of improving the validity of school marks (based on both cumulative and periodic assessment) as predictors than that of tests, and that the general test of scholastic aptitude usually proves a better predictor than either the specialised test designed to measure individual skills or the attempts to test qualities in the 'affective domain'.

5.0. ORAL EXAMINATIONS

Oral examinations fall into two clearly defined types : tests of oral skill, that is the ability to understand and to speak a language, and tests of all the other knowledge, skills and attitudes discussed in previous sections by oral rather than written methods. It will be best to distinguish the two, although purely oral skill almost certainly affects the assessment in all other subjects, as we saw pure skill with the pen contributing to success in all written examinations.

Oral tests must form a part of any valid assessment of competence in foreign languages. If, however, they simply take the form of an exposé by the pupil or a conversation between examiner and examinee, they suffer from grave defects both in validity and reliability. Many of these, the reliance on the subjective judgement of a great number of examiners, the variations in psychological rapport between different candidates and examiners, the problem of the over-helpful or over-forbidding examiner, the difficulty of carrying on a conversation and an assessment simultaneously, are common to both types of oral examination. There are others, such as sampling adequately the range of phonemes to be distinguished, which are peculiar to language examining. Fortunately the rapid spread of electronic recordings, combined with the more detailed analysis of objectives, is producing new methods of oral language examining which should be much superior to the old.

Even the simplest analysis of objectives discloses that there are two separate skills to be tested here : aural comprehension and oral expression. Aural comprehension can be objectively tested by playing tapes, which include a wide distribution of the most commonly confused phonemes, to groups of candidates who are asked to record on answer sheets their responses to a series of questions which test what they have heard. This not only introduces objectivity into what has been the least objective of examinations, but greatly reduces the cost and increases the speed of the operation.

Tests of oral expression require a slightly more elaborate procedure : the candidate's response to a series of questions, posed to him by the examiner, can be recorded on tape, as can a dialogue carried on with his teacher on a subject chosen by the examiner, and the recordings assessed at leisure either by an examiner or by a jury. It has been objected that candidates will be rendered nervous by having to speak into a microphone, but this is, perhaps, to ignore the extent to which tape recorders are now used in language teaching and also the extent to which other candidates are rendered nervous by the attitude of a severe examiner. Experiments with this type of oral examination have been carried out by the International Baccalaureate Office as well as in national systems and though there are undoubtedly some candidates who feel happier with the traditional 'face-to-face' oral there are others who prefer the tape recorder.

Oral examinations of the second type, designed to assess pupils' work in subjects other than languages, present much greater problems. They have a far longer tradition than written examinations and in many national systems play a very important part. The various forms which they may take and the problems of reliability which they pose have been admirably analysed by Professor G. Panchaud (46) of the University of Lausanne, who draws attention to the fact that little or no

(46) Panchaud, G. *et al.* (1969) : La Valeur Objective des Examens. *Etudes Pédagogiques.* Lausanne. Editions Payot. pp. 55-72.

research has been done on the various extraneous factors which may affect oral assessment, whether carried out by the candidate's own teacher, in the presence of an assessor, or by a stranger. Nor do most systems seem to have examined very thoroughly the desired role of the examiner, with a view to laying down general instructions and so ensuring a greater harmony of approach between examiners. Experiments carried out in relation to the French baccalauréat, for instance, show that examiners vary greatly in the proportion of time during which they, rather than the candidate, are talking ([17]). Some seem to be giving a lesson, some to be 'drawing the candidate out' and others to be conducting an inquisition. One of the studies which have been carried out seemed to indicate that there is a considerably greater correlation between the grades awarded by different oral examiners when they are assessing a number of specific abilities and then averaging their grades than when they attempt to mark on 'global impression' ([18]). It may be, therefore, that the reliability of oral examinations could be improved by greater specificity about objectives as well as about procedures.

What then are the objectives of the oral, as opposed to the written examination, in such subjects as Literature, History or Physics? Their use as a control of independent work submitted by the pupil has already been referred to and can be carried out, as in the International Baccalaureate, by questions posed on tape, but this is only a special case of a general objective. Panchaud points out that a special feature of the oral examination is that it enables the examiner to make sure that the candidate does not misunderstand the question, if necessary by reformulating it, and even more important, that the candidate really understands his own answer, by posing further questions. This can make a most important contribution to the validity of examining, particularly, but not solely, where the oral is conducted by an examiner who has already read the written scripts. How often does an examiner find himself saying of a written script "This is excellent for a pupil of this age — but does he really understand what he has written, or is he merely repeating as jargon what he has half understood from his teacher?" The difficulties in introducing such a control are practical ones, of which the greatest is the conflict between speed, validity and reliability. Unless examining is distributed among a very large number of highly localised juries, the delay involved in completing the marking of the written scripts first and then sending the examiners who have marked them round the schools as oral examiners is likely to be very great.

Yet the distribution of examining among a great number of local juries, each working independently of the others, is a well-known source of unreliability. It is possible that a solution to this dilemma may be found in the 'moderation' of oral examinations by the submission of taped recordings of face-to-face examinations to a central commission. Some very interesting work on different methods of recording oral examinations has been done in the context of the French baccalauréat ([19]), but clearly we shall need much more experiment before we can hope to arrive at a satisfactory method of giving adequate reliability to oral examinations. In the course of these experiments the factor of cost is likely to play an important part. Educational researchers, like so many other researchers, are inclined to seek perfection without considering whether the outcome of their researches could be generalised and applied on a wide scale within the bounds of the resources available.

Oral examinations have other valuable functions beyond this testing of the pupil's real understanding. The quickness of mind which enables a pupil to grasp a new idea, to see the implications of a new piece of information, or to appreciate immediately a flaw in his own reasoning when it is put to him, is a quality invaluable both in higher studies and in future life. Its development should be an objective of upper secondary courses and can only be assessed orally. I have myself found, when examining for the International Baccalaureate, that one of the most important objectives of literature courses, the development of a genuine delight in good literature, likely to last beyond the period of formal education, can be better assessed in an oral than in any written examination. It is to these qualities which Panchaud refers, in recommending oral examinations as a way of assessing "la vivacité d'esprit du candidat, son habileté à se tirer d'embarras, sa façon de s'exprimer, la solidité de ses connaissances, son émotivité, etc." ([50]). Yet, as the same time, the 'global' nature of such assessments increases still further the risk of unreliability.

(47) Piéron, Reuchlin and Bacher (1962) : Une recherche expérimentale de docimologie sur les examens de physique au niveau du baccalauréat de mathématiques. *Biotypologie*. March/June 1962.

(48) Timble, O. (1954) : The oral examination : its validity and reliability. *School and Society*. New York. Vol. 39. pp. 550-552.

(49) Panchaud, G. (1969) : *Op. Cit.*

(50) Panchaud, G. (1969) : *Op. Cit.* p. 60.

It seems possible that the rapid development of the video-tape recorder may help towards the achievement of greater reliability, by providing us with better means both of training oral examiners in this type of assessment and of moderating the assessments made. There seems some justification for the expenditure of some resources in this direction, since unless reliability can be improved there is a danger that this type of examination will be abandoned and a process of assessment which is, in the opinion of many contemporary students, too impersonal already depersonalised still further.

One further type of assessment to which video-tape or film might contribute is the assessment of practical laboratory work in the sciences. We have already seen that in some examining systems continuous assessment by teachers is being adopted in place of the single practical test and that this probably has superior validity ; but some kind of control of reliability by means of a more objective test would be desirable. If the skills which it is desired to test in science practicals are analysed, it becomes apparent that while some such as manual dexterity, can be best tested through continuous assessment, others, such as observation, the formulation of hypotheses and the design of further experiments, could be tested by the use of a film or video-tape which represented a typical laboratory situation. Candidates could be shown the film, perhaps more than once, and perhaps stopping at certain points, and asked to record their observations, interpretations or suggested further experiments, and to identify certain pieces of equipment, on prepared answer sheets, as for the aural test in languages.

The final advantage of oral examinations which must be recorded is their backwash effect. They encourage a 'dialogue' form of teaching more appropriate to modern youth than the magisterial ; they develop powers of oral communication, increasingly important in the age of the telephone and dictaphone ; and in so far as they employ audio-visual aids in assessment they encourage their use in teaching.

6.0. CONCLUSION

Assessment procedures can be used for certification, selection or orientation. The borderline between these three purposes is sometimes rather blurred. Many secondary school terminal examinations, for instance, were originally certification procedures. As such they worked reasonably well, so long as full secondary education was confined to an elite and there was genuinely no numerus clausus at entry to universities. Today, however, with government grants dependent on the grades achieved, with an official or unofficial numerus clausus in many faculties, and with extreme competition for entry to the most favoured faculties of the most favoured universities, they have become in effect selection procedures. This is true even for countries like the USA, Japan and Sweden which have approached most nearly to open entry to tertiary education. On the other hand procedures which were designed ostensibly for orientation such as the 'eleven plus' in England or the new procedure in the Swedish comprehensive school, described in section 3.3 above, also become selection procedures when the proportion of candidates anxious to follow one of the 'channels' available exceeds the capacity of that channel to absorb them.

This conflict of purposes is responsible for much of the confusion about techniques. For certification a combination of periodic and cumulative assessment is probably the most valid measure available to us and it may be significant that this is widely acceptable even for certification of professional competence provided no selection or numerus clausus is involved. Its reliability however demands a high degree of training in techniques of assessment and a high degree of professional integrity from the teachers, who are the only assessors in a position to carry it out. The first of these, as we have seen in section 3.2, can be very expensive in teachers' time and in the opportunities for teaching, rather than assessment, which must be foregone. The Northern Universities Board in England which has been experimenting with this type of examination in English for the last four years found, when it sought to extend the experiment in 1969 that of 256 schools approached, only 41 finally joined in the experiment. Their report states : "The reasons put forward for not taking part are of some interest. Not a few teachers feel that it is their duty to teach and the Board's to examine" (51). The second demand raises more delicate issues and the extent to which it can be met will depend on the differing social conventions and social pressures in different countries. Here it is perhaps enough to say that few examiners or inspectors with international experience would simply subscribe to the somewhat idealistic optimism of A. Agazzi's view that "there must be confidence in the teacher's honesty, sense of responsibility and sense of vocation" (52).

(51) Northern Universities Joint Matriculation Board (1970) : *Sixty-Sixth Annual Report.* Manchester. JMB p. 9.

(52) Agazzi, A. (1967) : *Op. Cit.* p. 56.

For orientation there seems every reason to suppose that global assessment is the ideal procedure. But global assessment is even more expensive and time consuming than periodic or cumulative assessment and involves so many factors that its reliability is almost impossible to measure and quite impossible to demonstrate. How, indeed, could we demonstrate what would have happened to a student, who has been oriented into one channel, if he had in fact been oriented into another, any more than one can demonstrate that a woman would have been happier had she chosen one suitor rather than another. Confidence in a global process of orientation depends on our confidence in a whole battery of assessments, many of them subjective. This confidence may well be justified, but it will not be achieved if many students are oriented, against their will and that of their parents, into channels which carry less social prestige and inferior life chances.

Thus orientation procedures also are affected as soon as the element of selection enters into them.

Reviewing the state of research and innovation in Europe, it would seem that the most promising line of development is the 'examen bilan'. What is needed is a process of assessment which is as valid as possible, in the sense that it really assesses the whole endowment and personality of the pupil in relation to the next stage of his life, but which is at the same time sufficiently reliable to assure pupils, parents, teachers and receiving institutions that justice is being done. Yet such a process must not by its backwash effect distort good teaching, nor be too slow nor absorb too much of our scarce educational resources. Would not the best way to work towards such a process be to analyse in significant, not over-sophisticated terms, the qualities we want to measure and to adopt for each the appropriate measuring technique, objective tests for some part of the course, cumulative project assessment including self-assessment for another ? And in arriving at this balance perhaps we shall decide that the reliability of all our methods is so questionable that the method with the best 'backwash effect' is the one to be preferred. The dilemma is well posed by J. Valentine of the educational testing service in Princeton : "A 'good' examination from the educational impact point of view, that provides an effective model of desired student behaviour, suffers as a measuring device because of the limited sample of behaviour it produces and the unreliability of marking. A 'good' examination from a strictly measurement point of view, on the other hand, that generates an adequately large and representative sample of behaviour, is likely to resort to efficient but essentially indirect and artificial measurement devices, such as multiple-choice questions, which have limited value as classroom exercices" ([53]).

The xperience of four years in the International Baccalaureate Office has shown how much Europe has to contribute to this programme of improving assessment, but often also how ignorant we are of each other's systems and how much we should gain by coordinating our research. Is there, for instance, any cross-national comparative study of the costs of assessment, in time spent by pupils and teachers both on taking examinations and practising to take them ? Is there any comparative study of the improvements made in the reliability of written examinations in different countries since the report of the Carnegie Commission ? If so I have been unable to find them, and Professor Panchaud's study of oral examinations quoted above indicates our equally thinly covered field for them. Considering the vital part which the baccalauréat or its equivalent plays in the lives of so many young people today, there is surely a case here for a concerted European programme of research studies.

---

(53) Valentine, J. (1969) : The Unbearable Burden of External Examinations in England and the United States. *Comparative Education*. Vol. 5 No. 2.

SECONDARY SCHOOL-LEAVING EXAMINATIONS

by E. EGGER

The report on "Secondary school-leaving examinations" [Doc. CCC/EGT (71) 6], drawn up by Professor E. Egger, University of Geneva, comprises two parts. In the first, Professor Egger discusses the results of an enquiry conducted by the Committee for General and Technical Education in Council of Europe member States and outlines the present position of various examination systems as well as the observed trends. The second part consists of individual contributions by specialists, who make suggestions relating to the various problems inherent in examinations. Three of these papers are reprinted below.

Examination research: Results thus far and outlook for the future

by M. REUCHLIN, Paris

Interest in the problems raised by examinations has never been so keen; it is shown in a large number of publications, many of which contain expressions of opinion and reflection or describe planned or attempted reforms. Accounts of events observed or systematic experiments using material gathered in such a way that facts can be confirmed or hypotheses objectively tested are less frequent; and it is these observations and experiments which constitute the field of examination research docimology (1).

The earliest work in this field was almost certainly that done in 1922 by H. Pieron, M. Pieron and H. Laugier, on the French primary school-leaving certificate, comparing exam results with those produced by the same pupils in a series of psychological tests. The object then was to find out whether school examinations could be used as a criterion of the value of psychological tests; results were not conclusive. They are mentioned here, however, because they indicate a trend which must be noted: the first research proposed to establish experimentally which of two testing processes was most satisfactory. One fact became so instantly and disturbingly clear, however, that it may well have driven the new field of research off its original path: the fact that traditional examination methods were monumentally inadequate. Almost immediately criticism of those methods became the chief activity in the field, seeking principally and almost exclusively to accumulate objective, verifiable data confirming the faults of traditional methods of evaluating school and university achievement. This attempt was highly successful, although its success gives no cause for rejoicing. Examples of such work were the following: *Etudes docimologiques,* published by H. Laugier, H. Pieron, M. Pieron, E. Toulouse and D. Weinberg (in the series *Travail humain*); the international enquiry financed by the Carnegie Corporation on *Les conceptions, les méthodes, la technique et la portée pédagogique et sociale des examens et concours*; a volume published by the English Carnegie Commission in 1936 (*An examination of examinations,* twice reprinted since, including a 1941 edition by Macmillan, entitled *The marking of English*); and a publication by the French committee, the same year, called *La correction des épreuves écrites dans les examens.* In 1956, the French advisory council for scientific research and technological progress, at the suggestion of H. Laugier, included the problem of examinations on its list of subjects requiring urgent national investigation and financed a new course of research to be carried out by the national

(1) Cf. H. Pieron, *Examens et docimologie,* Paris, Presses Universitaires de France, 1963.

Institut d'étude du travail et d'orientation professionnelle, the results of which have been published in several periodicals (*Bulletin de l'institut national d'étude du travail et d'orientation professionnelle, Travail humain, Biotypologie*). This work, directed by H. Pieron, M. Reuchlin and F. Bacher, remains largely critical, but also attempts to make a positive contribution to the investigation of examinations, thus returning to the original and, it may be, too quickly abandoned, intention of the first studies in the field.

Current investigations are continuing in this direction. The teams carrying them out have been predominantly trained in psychology, and employ all the techniques of measurement and control developed by that science since the beginning of the century. The adoption of these methods of objective evaluation necessitates an explanation of the value-scales employed, and consequently of the aims of education as well.

A situation report on docimology

These very cursory historical notes explain why an experiment-based critique of traditional examination and marking methods forms such a large proportion of the work done in the field. The results being without exception convergent, and having been abundantly confirmed, there is no longer any research of this type being carried out. In so far as they give teachers (especially those who agree to participate in them) information concerning the unreliability of their usual methods of assessment, however, such experiments are still valuable.

With regard to the French baccalauréat in particular, statistical analysis of series of marks actually given in the examination has been relevant.

For example, a comparison was made between the average marks given in the same examination subject by 17 boards in the philosophy exam and 13 in the mathematics exam (July 1955 examinations), the candidates in each set being allocated to the boards at random (in alphabetical order). Variations in averages were found to be considerable : from 5.81 to 9.06 for the written paper in maths, from 8.2 to 9.5 for the written paper in philosophy, from 8.3 to 13 for the physics oral (mathematics section), from 9 to 14.4 for the natural sciences oral (philosophy section), etc. These fluctuations are greater than one would expect for repeated estimates of an average based on a series of random samples from a single population group.

These differences in averages in the marking scales used by different examiners lead naturally to similar variations in percentages of candidates passed by the different boards. In the example given above, pass percentages ranged from 48 % to 61 % in philosophy, and from 31 % to 53 % in mathematics, depending on the board.

The consistency in marks given to the same candidates in different subjects has also been studied, and has been found to be very low indeed, even as between pairs of subjects such as physics and mathematics, or the written and oral parts of an examination in the same subject.

To establish these conclusions with greater accuracy, experiments have been specially conducted using not the marks actually awarded to examination candidates, but figures specially obtained experimentally.

In the Carnegie project, sets of 100 scripts were selected from the examination (baccalauréat) office files, five copies of each set were made, and issued, for marking to five different teachers, all experienced examiners. Averages varied as widely in this instance (where the material was identical in all 5 cases) as in the analysis of actual examinations : from 6.32 to 10 in French composition ; from 7.01 to 9.16 in maths ; from 7.65 to 11.23 in philosophy ; from 7.11 to 9.48 in physics, etc. Consistency among examiners (which is affected only by the rank assigned to scripts, not by the marking scale) varied considerably from subject to subject and also between pairs of examiners.

Also as part of the Carnegie project, 3 French compositions were marked by 76 different readers, with results ranging, respectively, from 1 to 13, 3 to 16 and 4 to 14.

More recently, experiments have been carried out involving the multiple evaluation of identical data in physics orals (2nd part of the baccalauréat in mathematics). Twenty oral examinations were recorded on tape and listened to by 16 lycée teachers with considerable experience of examining orally for that particular examination, each teacher marking each unit separately. Averages ranged from 13.4 to 8.03.

These difficulties in assessment are not, of course, restricted to the baccalauréat. Other experiments at other levels (from the primary school-leaving certificate to university first-degree examinations) reach similar conclusions. Nor are they peculiar to French examinations, which have been used as examples here. The general tenor of the English Carnegie Commission report is identical.

Such critical material, which exists in abundance, was the first contribution of examination research. The uses it has been put to are mistaken and in a sense deceptive in my opinion : it has been taken as evidence that examinations create a false problem, or adduced in favour of solutions to the problem which, although offering the appeal of simplicity, are probably false solutions.

If critical research into examination evaluation shows that the traditional type is a poor solution to the problem of examining, the problem nevertheless remains ; it remains to be solved, and the imperfections of the present systems are no justification for the total abolition of any form of examination (2).

The advances in science and technology which have revolutionised our world have given greater importance than ever before to the qualifications hierarchy. Qualifications have lost most of the specificity they may have possessed in a craft-based production system, and now serve mainly to distinguish between individuals in terms of their general level of education. Workers can move relatively easily from one job to another at the same level of qualification ; but it is extremely difficult today for them to make any vertical progress up the qualifications ladder in the exercise of their trade or occupation. The level of general education reached before leaving school (or at the price of great effort while working) is the deciding factor.

The school and university machinery which provides this general education can no longer concern itself exclusively with the transmission of disinterested culture, as it could at a time when protracted study was possible only for a privileged minority not subject to economic pressures. The vast majority of today's pupils and students come seeking access to the most highly qualified occupation possible. The mere length of time spent at school is not in itself any guarantee of the education actually received by anyone. If the university did not trouble to make individual assessments of such education, the "consumers" themselves would assuredly do so ; and in accordance with scales of values which, however distasteful they may be to the universities, would nonetheless tellingly affect the direction of instruction.

It is hard to imagine that the problem so imperfectly dealt with by examinations would not arise naturally in the course of an education. Long years of study must inevitably carry some people to the highest levels of qualification ; and our observation tells us that all individuals are not equally capable of reaching those heights, or equally prepared to submit to the long hours of effort and way of life required of those who do want to reach them and remain there.

Long study also means diversified study. Some subjects are spontaneously chosen by a number of students out of all proportion to the employment available at the end of their study. In technologically advanced societies, there are other reasons why some form of check upon individual capacities becomes inevitable. Instruments of considerable potency may be put into the hands of individuals in a wide range of activities ; because of their cost to the public and their potential danger, society lays down very strict conditions for establishing the abilities of those who use them. How many citizens would consent to the measures proposed by M. Lobrot, writing on the subject of examinations : "Examinations must be replaced by something else. But what ? The answer is : by nothing. It should be enough for a man to walk into an industry and announce, 'I have learned chemistry' (perhaps signing a document to that effect), his employers would judge him by his work, and would soon see whether he has told the truth. It should be enough for a man to hang up a sign and say, 'I'm a doctor', and his patients would judge him by his cures."

If we will admit the reality of the problem which traditional examinations are struggling to solve without much success, we must also admit the reality of certain related problems which will presumably be with us as long as the first one is : whatever process is used to guide pupils in their studies, and assign them to a particular occupational level, that process will affect the imparting and receiving of education and the emotions and stability of students.

Critical research on examinations, then, cannot make the problems disappear by proving that traditional examinations are a poor solution to them. Nor can it give unreserved support to more recent alternatives whose only value may be in their apparent simplicity.

Suggestions for measuring scales and types of test

Some of these would-be solutions are technical : certain reformers hold that it is desperately important to abandon the 1 to 20 scale in favour of the 1 to 5 scale. A scale with fewer grades would lessen the likelihood of two correctors

(2) Cf. on this point the controversy between M. Lobrot and M. Reuchlin in *Pourquoi des examens ?* — Paris, Société des éditions rationalistes, 1968.

placing the same essay in different categories, it is true ; but by using a scale with a single grade, the risk would be done away with altogether ! It is equally clear that each single error would be magnified in proportion as the number of grades diminished. The net result of this reform, thus, would be to convert a total sum of error, which would itself be in no way reduced, into larger units. It is possible that a five-grade scale would be better suited than a 20-grade scale to the sensitivity and accuracy of that measuring instrument which is the corrector : using a five-grade scale would at least avoid the absurdity of using tenths of a millimeter to talk about something which has been measured with a dressmaker's tape. But this is only a hypothesis, which could and should be tested before being adopted. Until doubt on this point has been dispelled, the dangers of imposing a broad-band scale (increasing the weight of each single error) are greater than those of putting up with a narrow-band scale for the time being.

Proposals to replace numerical evaluations by verbal ones form a second set of non-solutions. According to them the five grades mentioned above would not be designated by the numbers 1 to 5, but by words, such as Poor, Fair, Average, Good, Very Good. Some teachers, the literature-minded in particular, hope great things from this type of reform. Partisans of this view apparently see a significance in words which could provide an absolute scale of reference, by offering a means of expression in language understood by all. Experience (3) unfortunately shows that language is understood differently by different assessors : evaluated verbally, the same set of homework papers produces even greater disagreement among correctors than when those same correctors use numbers.

But some say that words can be used otherwise than as a direct translation of numerical marks. They can be arranged into sentences expressing the overall impression made on a teacher by each of his pupils. This proposal is often associated with the idea that a teacher, having had regular contact with a pupil throughout a year (chiefly in secondary education), is well-placed to make a solidly-based general assessment of him, without recourse to any formal means of testing what he has learned.

A discussion of this proposition, which involves more fundamental considerations than the previous ones, would demand more time than can be given to it here (4). I would only offer two or three reminders. In the first place, the material conditions, in which teaching takes place in many countries (the teacher-pupil ratio in particular) are such that a teacher cannot always know every pupil personally ; as things are now, he may even be completely ignorant of the home, family and interests of a child who has attended his classes for a year. Where closer relations have developed between teacher and pupil, each may have reached conclusions regarding the other which are felt to be self-evident and beyond question, but this self-evidence and certainty are, of course, totally subjective, based wholly on one system of personal relations which would be different with another teacher, and in some cases permanently influenced in nature and manner by one small detail.

The comparison of impressions by several teachers meeting in "council" is a palliative. Its value is limited : the pupil may have developed a temporary and superficial attitude towards the entire teaching staff ; also, the group comparison itself cannot escape the laws of group dynamics, which confer very different degrees of persuasiveness upon the testimony of different persons, according to the structure of relations between those persons. I would add that individual judgments of this type inevitably reflect the average level of the whole class, and levels vary far more than one might suppose. They also reflect the values of each teacher, the relative importance he attaches to each of the elements considered in the overall evaluation.

Here again, objective research has shown that different teachers have very different scales of values. It would be extremely optimistic to suppose that all these sources of potential error would ultimately cancel each other out in a group assessment, provided only that it covered everything. There is every reason to fear, on the contrary, that errors are not independent, and in particular, that the emotional tone of relations between one teacher and his pupil will pull several other sources of potential neuralisation into the same orbit. As a result, an overall assessment, even covering a relatively long period of time, will not solve the problem of examinations. It has been established (5)

(3) M. Demangeon, S. Larcebeau, *Une expérience de corrections multiples,* B.I.N.O.P., 1958, special issue, pp. 131-156.

(4) Cf. M. Reuchlin, F. Bacher, *L'appréciation des élèves par leurs professeurs,* Revue française de pédagogie, 1968, No. 2, pp. 19-25.

(5) F. Bacher, M. Reuchlin, *Le cycle d'observation. Enquête sur l'ensemble des élèves d'un département,* B.I.N.O.P., 1965, 21, No. 3, pp. 149-236.

that at the end of five years of primary school, a teacher's general estimate of his pupils' probable performance in lower secondary education was less accurate than a standard one-hour achievement test, when the two predictions, made at the same time, were compared with actual results at the end of the seventh year of school. This does not mean that the teacher's observations do not contain a mine of potentially precious information ; the problem is how to render that information useful. It will not be solved by recourse to intuition, generalisations, and verbalism.

At an even higher level of generalisation, we find another pseudo-solution. This consists in affirming that examinations create problems only if viewed as an "eliminating selection" but not if they are seen as an "advancement selection" ; or that replacing selection by orientation would solve them. Clearly, this is more rhetoric, which may be justified by particular social circumstances but does nothing to alter the terms of the problem. It is essential that those who do not wish or are unable to embark upon studies requiring vertical or steep ascension should be able to find a horizontal or more gently graded alternative at every level of education. It is essential that their choice cease to be between success or nothing, and become a choice between different successes. They will be qualitatively different, of course, but it cannot be seriously argued that they are not primarily hierarchical. The development of this hierarchy is not a political issue ; it is inextricably linked to the decisive role being played by science and technology, in every system. This is the problem with which the present examination system is failing to contend, this is the problem which the non-solutions I have mentioned either displace or deny.

Examination research cannot pretend to offer any satisfactory, full solution. To the extent that it adopts a positive, constructive attitude towards its traditional critical function — an attitude too often lacking in the past — it can, however, offer some partial suggestions at the technical level, and also attempt to state the wider problem more clearly.

As far as techniques are concerned, simple statistical operations may be of some value. For example, a group of examiners making a random selection from a large set of papers should find their marks distributed around equal averages. Where there is a large gap between the averages of two correctors, both may justifiably be asked to alter all the marks they have given (by adding to or subtracting from each the same number of points), in order to lessen the disparity. This process has been recommended to baccalauréat boards of the Académie of Paris.

In many fields standardised questionnaires may be used to check factual knowledge. They contain a large number of questions and cover an entire syllabus. Answers may be open or multiple-choice. A good deal of the opposition to this form of check is based solely on ignorance or prejudice.

An interesting variant involves building up a public "question bank" during the course and with the aid of teachers, from which test questions are subsequently drawn. Another worth-while experiment would be to give standard tests to extremely large groups of students. By consulting a published account of the results (percentage of correct replies to each question), teachers would be able to place their own classes in relation to the large group and make the necessary deductions about their own teaching and assessment scales.

Another and very interesting possibility is the "open book" test, in which the candidate is given all the material necessary for making a synthesis, following a line of reasoning, etc. The preparation of such tests (choice of subjects and material) and their assessment, however, are very difficult. Multiple, independent correcting is probably the only (and expensive) means of achieving an acceptable level of objectivity here.

Introduction to docimology

In addition to these suggested measuring scales and types of test, it is to be hoped, in broader terms, that all teachers might be given an introduction to docimology marking in the course of their training. This should, without fail, cover five aspects :

— The purposes served by pupil assessment ;
— Elementary statistics ;
— Study of published research ;
— Experience of multiple correcting ;
— Study of new types of checks.

Such an introduction would not provide teachers with ready-made solutions, for these are still to be found. But it would make them conscious of the problem, and enable them to take a more active part in the search for solutions, both in practice (the work of the English examinations boards might serve as an example) and in research.

At a still higher level of generalisation, work on examination evaluation leads in turn to an effort at clarification (6). A better solution to the problem which traditional examinations solve so poorly cannot be found without first defining the exact function academic assessment is supposed to perform, e.g. does it relate mainly to teaching, or should it simply describe attainment, or is its purpose principally to estimate suitability for further study of some particular type ?

This effort to clarify the functions of assessment and examination leads in turn, however, to an effort to clarify the aims of education — since the object of any such operation, after all, is to find out whether or not they have been achieved. It is certainly not for teachers alone, and still less for the examination research experts, to define those aims, nor is that the intention of recent research, in particular, that of B. S. Bloom ; but it is part of their task to state the question in terms that will lead towards a useful answer, and this is at least one of the directions which modern work on examinations should pursue.

(6) M. Reuchlin, *La docimologie, effort d'explication.* Les Amis de Sèvres, 1968, No. 2, pp. 33-40.

Objective testing and educational assessment

by W. D. HALLS, Oxford

The doubts raised on all sides about the educational value of conventional examinations have increased in recent years. In particular their validity as a means of accurate grading and differentiation has been called into question. Too much, it is alleged, is left to chance. Most "essay-type" examinations preclude a systematic sampling of the knowledge, skills and behavioural attitudes of learning. The setting of only a few questions to answer means that large areas of what has been learnt remain untested. If these questions contain alternatives the unfairness to the candidate becomes even more manifest : only rarely can such alternative questions be of the same order of difficulty and complexity. Above all, it is alleged, subjectivity in making is great, and may lead to erratic results (1). On the purely material plane, moreover, there are great difficulties surrounding conventional examinations : the security of the question papers is only one aspect of this ; the phenomenal growth in numbers of candidates, particularly in the vital examinations of secondary education, is another. Thus everywhere national education systems are seeking out new methods of evaluating their pupils. In Sweden, experiments are being made in continuous assessment, using objective tests as one element in this process. Elsewhere, more attention is being paid to the school record and to the interview, but there are insuperable difficulties in standardising these two procedures. It is for these reasons, therefore, that objective testing is put forward as yet another alternative. The purpose of this paper is to discuss its validity, as a method of evaluation.

Objective testing may be defined as a systematic method for evaluating by sampling procedures an individual's psychological behaviour, mainly in relation to ability, aptitude or achievement. The first example of such a test was the "intelligence test". Although its historical origins can be traced back to the work of Galton, Pearson and Cattell in the late nineteenth century, the credit for devising general tests of scholastic ability must go to A. Binet, who, with Simon, first devised in Paris a scale of intelligence from which was derived the concept of the Intelligence Quotient (IQ). From this European work in individual testing the Americans elaborated group testing of large numbers. The first achievement tests were devised by E. L. Thorndike, working at Teachers' College, Columbia, in New York in the early 1900s. Testing as a method of evaluation first supplemented and then largely supplanted conventional examinations, so much so that today in North America most children by the age of 12 have undergone at least half a dozen intelligence or achievement tests of one kind or another. The extent of the "testing industry" may be gauged from the fact that in 1964, 148 million test booklets and accompanying answer sheets were sold. Despite their European ancestry, however, tests as a means of assessment, have not been widely adopted

(1) Hartog and Rhodes : *An Examination of Examinations,* 1935 : also the *Report* of the *Commission Française pour l'Enquête Carnegie sur les examens et concours en France.*

this side of the Atlantic. (United Kingdom, where there were few linguistic difficulties, is a possible exception ; certainly in England so-called "intelligence tests" were generally in use as part of the notorious 11 + examination for selection to secondary education.) One American writer ascribes this reluctance on the part of Europeans to adopt testing as a routine procedure as due to "basic ideological and cultural differences of opinion about the nature of human abilities" and to "technical and social problems which make the large-scale use of objective tests either difficult or impractical" (2). That this attitude in Europe is changing there can be no doubt. More and more, at the upper levels of secondary education, and particularly in relation to access to higher education, testing is being considerd. It is this "classification and promotion" aspect of testing, rather than its more frequent use as a "diagnostic for counselling and treatment", which is undoubtedly at present of the greatest interest.

Objective test

What, precisely, is an "objective test" ? According to Ebel (3), a test contains "a small but statistically significant number of short-answer questions — 'items' designed to test the most important areas of knowledge, skills and behavioural attitudes". The adjective "objective" is mainly applied to contrast with the subjectivity inherent in marking the convention examination. Certainly the scoring of a test is usually so simple that no bias can enter into the marking and no technical expertise is required. Where, however, subjectivity remains a danger is of course in the devising of the test, which is a lengthy, highly-skilled and extremely costly process, and in the pre-determining of what is the most appropriate answer from a number of possibilities. Tests, in fact, contain a number of "items" which may be of different kinds. The simplest and crudest sort is the one which poses a short question and supplies two possible answers, from which a choice has to be made between the "true" and the "false" one. A refinement of this — and the most widely-used type of item — is the "multiple-choice" question where several possible answers are supplied — perhaps as many as five — all of which are responses of differing degrees of plausibility, but only one of which is absolutely correct. To answer this kind of question requires both knowledge and judgment. Another kind of item may be described as "classificatory" : the candidate has to assign each object in a list to its appropriate class ; the "objects" may consist, for example, of names, descriptions, pictures and statements. A further kind of item is termed the "matching" one : two lists of statements or symbols have to be exactly matched, detail by detail, against each other, using the principle of "closest association". To avoid the effect of guessing, a mathematical correction can be applied to the final score, if the above types of items are used. A final possibility, although this does not fall strictly within the purview of "objective testing", is to present a number of short-answer questions — usually described as being "open-ended" — to which the pupil may give a "free" answer, as distinct from the other items, where his choice is limited in advance. This latter kind of test obviously requires the corrector to exercise judgment as well.

Some of the advantages of objective testing have already been mentioned. The ease with which large numbers of candidates can be dealt with and the tests scored are not inconsiderable merits. It is also clear that, with a larger number of items, or questions, a much more systematic "sampling" of the content of a syllabus can take place. With the help of a taxonomy, "weighting" can be applied to various parts of the test ; a selector for candidates for university arts faculty may, for example, be more interested in how well potential entrants answer questions demanding verbal rather than mathematical ability, and can "weight" his marking accordingly. Moreover, it is also claimed that a test has great reliability ; if well-devised, it will be consistent in its measurement of what it is intended to measure. Thus the assessment made of pupils in one year can be compared with assessment of different pupils made in previous years, by standardising scores. An even more tangible advantage, for example, in selection for higher education, is the claim that tests have high predictive validity (although in fairness it must be added that most authorities agree that a candidate's school record is as good a predictor. The drawback in using records, however, is that although as a whole they are usable, the individual cases may be assessed on different yardsticks : so many variables, from teacher quality to standard of instruction, lie totally outside the control of the selector). All in all, therefore, the advantages of using tests for assessment are very considerable.

(2) D. A. Goslin : *The search for ability : standardised testing in social perspective*, New York, 1966. This is the best work on the social impact of testing.

(3) R. L. Ebel : *Measuring educational achievement*, Englewood Cliffs, New Jersey, 1965. This is one of the best standard works on the techniques of educational measurement, and is easily understandable by the non-specialist.

Interpretation of the scores

Are they, however, overwhelming? Vital to the understanding of objective testing is how the scores should be interpreted. Any test, however carefully devised, does not yield a score that is anything more than an approximation : it means that the estimate of the candidate's ability will fall within a certain range, of which the score he obtains is the mean. The extent of this range represents the "standard error of measurement". The large claim is made that this standard error is less than in conventional examinations. But, as far as is known, no European examining body, using conventional methods, has ever published what the standard error of its examinations is. Until this is made public, the only verdict possible here on the testers' claim must be one of "not proven".

A more substantial objection concerns the development of new theories on the nature of intelligence. These have called into question the permissibility of objective testing. We now know that intelligence is to a large extent environmentally conditioned. The tests are no better than conventional examinations in eliminating an element — possibly the preponderant one — which gives a socio-cultural advantage to children from educogenic families. An even more serious criticism is that objective tests — and here they may even be more biased than conventional examinations — fail to measure creative capacity. Too often "intelligence" has been conceived in terms of what has been called "convergent" thinking — the ability to see relationships that, once perceived, appear obvious. But intelligence, in its new definition, must also embrace "divergent" thinking — the ability to perceive unusual relationships, which entails the use of creative thought. Tests have as yet not been devised which could measure this capacity.

By their *modus operandi* existing tests require the answer that a majority would give ; moreover, in short "responses" it is impossible to justify any answer that is unusual, however ultimately well-founded that answer may be. Yet it is this ability to discover the unusual, rather than the usual relationships, to see connections that have not been discerned before, which is a quality that must be highly prized in modern society, where innovative capacity is rare. In a conventional examination, where creativity and imagination may be given free rein, although its spark cannot be accurately measured, a qualitative assessment of it is possible. The more rigorous selection is, as for example, for entrance to higher education, the more this creative aspect of intelligence is of value. In tests, in any question that calls for relating one phenomenon to another, the tester, because he cannot interrogate the candidate to find out whether an unusual answer is well-founded, must automatically penalise those who do not arrive at the expected answer. Such an objection to present forms of testing seems grave, and not easily overcome.

Linked with this, is the allegation that tests cannot measure powers of synthesis and analysis, or the capacity to follow through a chain of thought to its logical conclusion, whether in discursive prose or in the symbolic language of mathematics. Testing is regarded as a passive process, in which even the ability to express oneself well in the mother tongue — a prerequisite for success in any field of human activity — is not required nor fostered.

Further objections concern the undoubted fact that tests are, despite thorough checking of items, far from perfect (4). Items lend themselves to ambiguities and obscurities. But perhaps this may be described as a mere fault of design, which can be remedied with greater care in drawing up the test, and in pre-testing it before it is actually used.

A last objection concerns the abuses to which testing may give rise. Chief among these is the "backlash" effect upon teaching. There is little or no evidence that well-designed tests such as those of the US College Entrance Examinations Board (CEEB) yield vastly different scores when coaching has taken place. A gain of a few points may be registered, but since we have been warned that test scores should be interpreted as demonstrating that a candidate's ability lies within a certain range, and cannot be determined as an exact mathematical quantity, this is no insuperable drawback. Yet there have grown up in the larger US cities coaching "schools" to prepare pupils for tests. The College Entrance Examinations Board has expressed its concern at this detrimental development, "because we see the educational process unwillingly corrupted in some schools to

(4) Mention must be made here of an attack on testing by B. Hoffman : *The tyranny of testing*, New York, 1962. In 1968, when the present writer attended the annual meeting in Chicago of the College Entrance Examinations Board, which uses the services of the Educational Testing Service of Princeton, and is the largest testing organisation in the world, an « Anti-Test Protest » was under way. To the present writer it seems that not the principle of testing is called into question by the movement, but the abuses to which indiscriminate testing may lead.

gain ends which we believe to be not only unworthy, but ironically, unattainable". Nevertheless, such commercial institutions continue to exist, and are even emulated by public schools who see the salvation of their pupils encompassed by submitting them to endless testing.

Scholastic aptitude tests

Objective tests have been used to test achievement, scholastic aptitude, and personality and character traits. Unfortunately, the tests of personality and character that have been devised up to now have been the least accurate, yet it is precisely these which at the upper secondary level would be of the greatest utility. Of particular interest, however, in view of the difficulties of prediction of success in higher education, are scholastic aptitude tests, a refinement developed from the old-style "intelligence" tests. In the US the College Entrance Examinations Board has developed a Preliminary Scholastic Aptitude Test, taken in the eleventh grade, and a Scholastic Aptitude Test proper taken a year later. Their purpose is to serve as a guide to universities and colleges of the future academic success of their would-be entrants.

In England these tests have aroused so much interest that the Committee of Vice-Chancellors, in its search for new methods of university selection has conducted pilot experiments with English adaptations of the tests on a wide scale to determine their feasibility as a selection instrument. No definite conclusions have yet been reached. The aptitude tests have no passing mark as such, but simply yield a score (which is to be interpreted as accurate within a given range) which is passed on to the appropriate college or university that a candidate seeks to enter. The score is standardised on a national scale, and it is up to the receiving institution to decide whether it is sufficient to justify admitting the candidate to the particular courses it offers. It must be emphasised that the tests are aimed at supplementing the evidence provided by the schools, not at replacing it. Used by the Ivy League universities as Harvard, there is no doubt that their value is much appreciated.

The tests have as their object the measurement of general ability. They are not yardsticks of attainment, but seek to measure the basic learning skills required for university success. They are of two kinds, mathematical (SAT-M) and verbal (SAT-V). Whereas the SAT-M is alleged to measure ability "to reason with numbers", the SAT-V avowedly measures the ability "to read with skill and to understand and use words correctly". As well as the experiment mentioned above, a trial of the tests, without any alternation in their formulation, was carried out on pupils in England. The general conclusion reached was that "United States tests can work well in other countries but that certain items in any test might have to be changed to take into account cultural differences". If this is so as between two countries where linguistic and cultural differences are comparatively small, it is even more so where such differences are large.

An even more challenging experiment is being mounted in Canada. A Canadian version of the SAT is being worked out in English and French as a criterion for university entrance. Such a test, in a bicultural situation, must be of equal fairness in both languages, particularly in provinces such as Quebec and Ontario, each of which has respectively a strong English-speaking and French-speaking minority. (Incidentally, the Quebec provincial Department of Education has pioneered the modern use of objective tests in French, and there are indications that French-speaking nations everywhere are interested in its work.)

In order to compile a SAT-V test a very complicated procedure is followed by the CEEB. Each year some 2,000 items are drafted by a team which includes a psychologist, English specialists and a number of "laymen". After pre-testing, only 90 items are retained in the final test. The first part of the test can be broken down into items that ask for sentence composition, antonyms, and analogies, and relate to categories such as the esthetic and philosophical, the world of practical affairs, science, human relationships and general matters. The second division of the test is of items that ask questions on passages of reading comprehension relating to the sciences, the humanities and social sciences. Such passages may consist of straight narrative or discursive prose. The candidate is asked questions on the ideas, the inferences to be drawn, the logic of the argument, and its style. Similar care goes into the compilation of the SAT-M test, where the main branches of mathematics are used — but no previous specialist knowledge is assumed — and types of thinking connected with computation and numerical judgment, and relational thinking, are evaluated. Each year a certain number of "dummy" items (which do not count towards the candidate's final score) are slipped into the SAT tests ; this constitutes an experimental section for pre-testing items. After the SAT tests have been worked a most detailed analysis of each item is made to see whether it

discriminated well between the brighter and the weaker candidates, and was otherwise viable.

The difficulties of "transfer" of such tests from one language to another — of transliteration or even translation there can be no question — seem almost insuperable. Nevertheless, the prospect of such a test being used as providing additional evidence uniculturally for selection to higher education is an attractive one. So also is the prospect of devising a satisfactory multicultural one on a European scale as a new instrument to be used in solving the problem of equivalences, or at least the "acceptability" of students from foreign countries. A first step might be the production of suitable tests to serve French and German-speaking areas of Europe. It is also likely that it would be feasible to produce a test which might serve both the Italian and the Spanish-speaking peoples. Those in English already exist, and could serve as exemplars, if not models.

Achievement tests

If one now turns to consider the use of achievement tests in the various disciplines it can be seen that their utility will vary considerably according to the subject. In the mother tongue, where "free composition" and literary appreciation are required, they would seem to be of only marginal value. In modern languages, where the aim is to test grammar and comprehensive and the correct use of language, they would seem to serve a more useful purpose. In aural-oral tests, easy standardisation is possible. But as with the mother tongue, neither creative writing or literay appreciation would seem capable of adequate testing in this way. (We omit translation altogether.) In history and geography much would depend on the nature of the test. If a series of questions, for example, were asked regarding an historical document, or even a picture, much could be learned about a candidate ; in geography the best American tests now consist, for example, in presenting a map, from which deductions have to be made. In the sciences, where much factual knowledge is still required, the test offers a rapid way of systematically checking on the candidate's memorisation of facts, and is also useful for verifying his ability to solve small problems. The same holds true for mathematics.

What has to be avoided at all costs is that tests are used exclusively, and solely as a means of ensuring that a candidate possesses the requisite knowledge. This was the trap that the Americans fell into initially. As soon, however, as it was emphasised that a curriculum consisted of more than its content, and postulated general cognitive aims of learning such as flexibility, judgment, intuitive ability and other intellectual qualities, as well as aims intrinsic to a subject, the character of achievement tests changed rapidly and radically. But to test such qualities systematically by the use of items that are difficult to devise and cannot be used too frequently (although they can obviously be used more than once, for no candidate can keep the "question paper" when he leaves the examination room) is obviously a difficult process.

Whole books have been written on the social impact of testing. Its pedagogical repercussions have already been touched upon. There is no doubt that testing can build up neuroses among pupils and parents — not to mention teachers. What is needed is a clear idea of the limitations of the testing process. The widespread use of tests as a means of making decisions of vital importance to individuals, whether in school, in business or industry, has caused much heartsearching in the USA. In school, apart from their use in selection for higher education, tests have been used as part of the counselling and guidance process, for differentiating pupils according to their ability, and for the weeding-out of gifted or retarded children to be placed in special classes and schools. It must be emphasised, however, that in the context of schooling, only rarely is testing used as the sole criterion : before decisions are made other evidence, such as teachers' evaluations or school records are almost invariably used. Testing is therefore no panacea for all the educational problems connected with evaluation that face us in Europe. There would seem to be a strong case for a series of controlled multinational and multidisciplinary experiments in evaluation in which the advantages of testing are weighted against those other modes of evaluation in current use, from conventional examinations to class work and the use of school records. Correlations of predictive validity, for example, might be established through using each of the evaluation processes mentioned on all candidates for entrance to higher education and then comparing their efficiency. In any case, if the return across the Atlantic, to where it originated, of objective testing has served any purpose, it has been to re-awaken interest in producing a more systematic — dare one use the word "scientific" ? — process of evaluation than the subjective methods, too often based upon prejudice and dogmatism, which have hitherto held sway in Europe.

From "point-in-time" examination to general assessment

by J. CAPELLE, Bergerac, Dordogne

In October 1966, the Council of Europe held a course on examinations and access to higher education, in Brussels ; in a report read on that occasion, I proposed certain definitions of examination types which I have again adopted in the following pages.

The examination was defined as an operation designed to assess a candidate in relation to a definite goal.

With regard to the structure of the examination itself, a distinction was made between the "immediate" type which I have since called a "point-in-time" exam (examen ponctuel), using a term which underlines its distinctive feature, and the "long-term" or general assessment based on a large number of sources of information.

The hazards of the "point-in-time" examination

The particular feature of this examination is that it is performed in a period of time which is extremely short in relation to the period of preparation for it and to the period during which the pupil may be observed and assessed by his teachers. In essence, it selects, more or less at random, one point on the graph of the pupil's performance in the subject throughout the whole of an academic year or course of study.

It may be of several types :

- the candidate may be given a series of psychotechnical tests ;
- he may be "interviewed" on more or less narrowly defined subjects by a small group of "judges" ;
- he may be asked to write papers referring to a pre-determined syllabus in a series of academic subjects, and judged according to the quality of his performance (written or oral).

In schools, the third type is by far the most common and what I have to say hereafter refers chiefly to it.

The point-in-time exam is unreliable because it judges a candidate solely on his reply at one particular moment not of his own choosing to a proposition selected from a number of possible propositions, to which he is required to address himself. The quality of his reaction cannot be the same at every moment and for every proposition.

Any numerical assessment of his performance accordingly involves a margin of uncertainty, bearing in mind his possible response to the questions which might have been put to him instead of the one that was. Given that success depends upon his obtaining a certain minimum mark, the element of chance will mean something very different for the candidate well above the minimum and for the one who comes close to it.

In addition to this uncertainty, which we may call intrinsic, there is another, extrinsic element of chance, resulting from the diversity of assessments which different correctors may assign to the same performance, or which the same corrector may assign to the same performance at different times.

If the evaluation of a candidate's capabilities in a particular subject is subject to both these elements of uncertainty, their evaluation on the basis of work produced at one moment of time represents an extrapolation which adds yet a third element — and a sizeable one it is, too — to the fog of uncertainty surrounding the traditional examination system.

Must point-in-time examinations be abolished ?

The love of gambling and fighting, that is, the love of risk, is inbuilt in us ; it is one of our most effectives motors. Therefore, it would be absurd to lose the stimulating effect of a point-in-time examination. What needs to be carefully watched, and possibly changed is the aim of such exams : the stake must not be so high that the candidate will be thrown off his stride by undue nervous tension.

Stimulus-examinations can be used with profit, when given moderately often and as a competitive game, like a contest between athletes. But the pupil who does not do well must not be permanently harmed by his failure on one occasion ; like the athletes, he must be able to tell himself that he will do better "next time", for there has to be a "next time".

When the point-in-time examination is also a turning-point examination, there may be no "next"

time, and this is what causes the stress and disruption which are prejudicial, not favourable, to continued study.

This being so, is it necessary to abolish this form of competition when its results have a decisive effect on a young person's career?

I do not think it is, provided that throwing open a prized career to competition of this sort does not ruin all a student's chances for fulfilling his ambition and desires in his chosen field of study.

Competitive examination as a means of access to some careers is not an evil if it is not the only access to them, and if there are a large enough number of similar occupations which can be obtained through other and, in so far as possible, less hazardous routes. The situation of a participant in a competitive examination must be a little like that of the worker who plays the pools: if one of his attempts is unsuccessful he may be disappointed, but he will not be cast into despair.

A second point must be made: competitive point-in-time examinations become virtually indispensable, or in any event more acceptable, when their object is to detect exceptional ability; those possessing it seldom do poorly in this type of examination, and a challenge merely brings out their virtuosity. The "concours général" in French lycées has no adverse effect on those who fail, but those who succeed are never undistinguished, and the prospect of being allowed to sit for it provides an excellent stimulus for pupils in those schools.

The international mathematics Olympics in Moscow may be seen in the same light. Offering a competition of this type in certain subjects to the best pupils in European secondary schools certainly could not lessen the effectiveness of secondary education: they would be striving for a success which, like the pine branch for which Olympic athletes competed, would be symbolic and no less ephemeral, but it would have a definite stimulating effect and would provide a valuable opportunity for singling out the most gifted.

Continuous assessment

Evaluating the capabilities of the young, principally at the time when they must make a choice which will decide the success of their future studies and career, is a goal as difficult of achievement as it is desirable.

The potential of an individual is to some degree contingent, no doubt; it will evolve with his personality, his physical development and the maturity he acquires with experience. It can rise and fall in response to a change of circumstance (discovery of a motivation, determination to surmount an obstacle, assumption of unforeseen responsibilities).

There are so many imponderables among the elements which propel an individual towards a career and which, more often than not, deposit him in a vocation for which he never suspected he was prepared.

But setting all these difficulties aside, along with the knowledge that at any given moment the components of an individual personality are not all intrinsic, we must still be able to define those components far more objectively than is possible with the patently fortuitous deductions that can be obtained from a student's performance on a single examination.

It is in this context that the long-term assessment becomes relevant.

Unfortunately, the more we seek to define this form of assessment, the more we realise how little we know about the art of evaluating the individual. For too long, university circles may well have regarded educational research as being of little importance, even incapable of reaching beyond conventional verbalism; and new horizons may well have been slow to appear for that reason; but research on examinations — which, so to speak, perform the same function for education as inspection in the factory does for production — is undeniably even more inconsistent in doctrine, and has even less to show for itself in actual results. In order to define as concretely as possible the concepts that go to make up the idea of long-term assessment, we shall look at the specific problem of assessing a pupil at the end of the upper secondary course (usually lasting three years), with a view to establishing a summary of his attainments and a portrait of his potential qualities.

These two aims call for two types of evidence, one relating to what the pupil has already achieved and the other to what he may be able to achieve hereafter. The former is analytic, the latter synthetic.

Analytic evidence

The object here is to produce a condensed and if possible standardised account of the pupil's attainments in each subject throughout the course (e.g. in the French system, during the three years of

the maturité). This account should not, of course, be confined to a list of marks or place obtained on the traditional composition. The profession of results obtained over a period of time on other forms of exercises or tests should also be noted.

One possibility would be to give tests throughout the course to all pupils in schools of the same level in one region, or all over the country. The use of these tests, which would act as stimulants and also make comparison more easy, should never give rise to the tension and agitation caused by the institution in February 1960 of a part-session of the baccalauréat exam.

After the Easter holidays, in each of the three years of the course, pupils would write an essay on a subject drawn by lot (and for which the necessary documentary material would be supplied, so as to dispel all needless anxiety). Continuing with this hypothesis, the subject — maths, French civilisation, physical sciences, etc. — would not be disclosed until the exam. In this way, pupils could both work and sleep undisturbed during the preceding days. The essay would be corrected carefully, following instructions given by a regional or national examinations board.

The boards themselves should work in close association with the educational research institutes whose creation within universities was urged by the Caen Conference in November 1966.

Such tests, being "point-in-time", should not, of course, be decisive ; the results would be recorded in the analytic summary alongside other results, as an additional source of information. Their importance could be considerable, however, and not only because they would facilitate comparison and increase comparability between school marking systems.

Whatever the fate of this proposal, the object for each teacher remains to mark and classify performances on a relatively large variety of exercices and over a relatively long period of time, in order to build up a picture of the pupil's attainments which shows the work he has done and his efficiency.

Synthetic evidence

Going beyond this account of attainments, the object now is to compose a portrait of the pupil himself, which should be viewed as a synthetic assessment of his potential performance in different branches of study and response to the responsibilities he may have to assume as a student.

From a look at subject-teachers' end-of-term or end-of-course reports in, say, the school record, we can plainly see the hazards of such an undertaking. The comments are vague and cannot be compared : they run along the lines of "poor", "good worker", "achievement poor", "does his best", "intelligent", "can do better", "able but erratic", etc.

The truth is that we have not defined what qualities we want to assess, we have not determined their places on the spectrum of a pupil's possible responses in each subject, and we have not agreed upon terms for expressing that place. The portrait I am thinking of could be compared to the diagnosis which a group of medical experts would produce, each for his own specialisation, to describe a person's state of health. Each part of the diagnosis would be expressed in standardised terminology, and the degree of quality or deficiency expressed by this terminology could be converted numerically. If the same person is examined by a different group of doctors representing the same specialisations, their diagnosis will be very much like the first, in both the terms employed and their numerical translation.

In this respect, the field of examination research is far behind the medical profession !

Nevertheless, let us try to suggest how the portrait we want might one day be composed, however indistinct its outline now. We will not attempt to have it show every aspect of a personality, which is always complex.

It could express two types of features :

- polarised features, or those revealed in the pupil's behaviour in his relations with each subject taken separately ;
- general features, or those which are constant in the pupil's personality regardless of the object of his activity.

These features are composed of qualities which need to be defined precisely enough to distinguish each from the others, and evaluated by means of a convention which must be as simple as possible and hence numerical.

Even without any decimals, the 20-point scale, however accepted it may be in France, implies a precision of evaluation well beyond that with which human judgment can express itself objectively, in an area so full of contingencies and so difficult to defend against subjective interpretation.

It would seem that a quality could be sufficiently subtly graduated on a five-degree scale, expressed by the numbers 1 to 5 (for insufficient, adequate, fair, good, excellent).

Polarised features

For each subject, the qualities would be assessed by the teacher in charge. Greater comparability of modes of assessment and greater reliability would be achieved by having teachers of the same or related subjects work together. A teacher should be able to follow his pupils' progress throughout a course by consulting the rest of his department — in contrast to the present professorial compartmentalisation where every teacher scrupulously "respects" the isolation of every other, even when they are all teaching the same subject.

The next question is, what qualities do we mean to assess at the level of individual subjects ?

It is convenient, and indeed desirable, to consider the same qualities for all subjects, specifying where necessary how one of them — imagination, for example — is to be interpreted in a particular instance.

In the following scheme, which may be oversimplified and excessively arbitrary, I have suggested six qualities.

— Three are constant :

- aptitude for analysis,
- aptitude for synthesis and composition,
- imagination and creativity.

— Three are dynamic :

- rapidity of assimilation,
- curiosity,
- initiative.

On this basis the table of polarised features could be presented in highly condensed form, as follows :

Analysis	*Synthesis*	*Imagination*	*Subject*	*Rapidity*	*Curiosity*	*Initiative*
4	5	3	French	2	3	2
3	3	2	History	3	4	3
			Mathematics			
			- - - - - - -			
			- - - - - - -			

General features

The assessment of these qualities requires us to go even more deeply into the individual's personality.

Here we are no longer concerned with qualities as they may relate to one or another subject of study : we want a diagnosis of the candidate's mental attitude, intellectual worth, morality, and accessibility.

To some extent a portrait thus composed would be a synthesis of the polarised features listed above, but it would go beyond that to reveal the basic character of the individual.

Two problems immediately arise, one of objectivity and the other of ethics.

As regards the former, it would be necessary to define those features which can be deemed characteristic, and which would indicate the type and amount of responsibility an individual might assume ; and a means of evaluating them would have to be found. The evaluation would be performed by the staff as a body, including in particular those teachers who have had occasion to observe the pupils outside the classroom (physical education instructors, study supervisors). Tests might also be used, and their results compared with those of more academic assessment processes, etc.

The ethical problem arises when we cease to judge the work produced by a pupil and begin to assess the value and deficiencies of his personality.

Proceeding in the order outlined, with all the caution and reservations required in the interests of objectivity, we should arrive at a true portrait of the individual's basic personality structure, and thus obtain the best possible basis on which to compare him with others and place him in relation

to the requirements of a particular type of study or career. We should also, however, be trespassing upon his conscience and privacy.

It is possible to state on a school certificate that a candidate received a particular mark in mathematics without inflicting pain ; but it is a far more delicate matter to record some personality deficiency there.

Teachers in a position to compose such a portrait, disclosing every feature of an individual's personality (assuming the thing to be possible), should be pledged to secrecy on the same terms and for the same reasons as medical practitioners.

The continuous assessment, which achieves the fullest possible knowledge of an individual, also encroaches upon the realm of the confidential. Its full contents, therefore, can never be made public ; but the person concerned should himself be informed of them, and should be able, if he so desires, to communicate them to whatever authority takes the decision on applications he may submit.

To conclude : I do not ask for the abolition of "point-in-time" examinations, whose value as stimulants and detectors of exceptional ability cannot be questioned. But they must not be used for other purposes : that is, in conditions which leave the pupil gasping and his work in tatters.

The long-term assessment would appear to be the best basis for those decisions which have to be taken at the end of a course of study concerning future study or the choice of a career. It alone can attest to the continuity of work done, and leave the candidate with the comforting thought of an evaluation which is free from the vagaries of chance and human moods.

But our schools and universities have not been adequately prepared for the responsibilities it entails. Sustained effort is required in two directions : research on examinations, and acceptance by teachers and parents alike of a new principle, i.e. that the best guarantee of objectivity is the teacher who knows the pupil, not the one who does not.

Publications

The two series of educational works "Education in Europe" and the "Companion Volumes", published in English and French by the Council of Europe, record the results of the studies of experts and intergovernmental surveys carried out within the framework of the programme of the CCC. We here present the latest publications in both of the series, obtainable from the Council of Europe Sales Agents, as well as some other books published with the support of the Council for Cultural Co-operation of the Council of Europe.

Series "Education in Europe"

THE TEACHING OF MATHEMATICS AT UNIVERSITY LEVEL

by F. FIALA

Published by George G. Harrap & Co. Ltd., London, 1970, 163 pages, £ 1.50.

This book is a comparative study of mathematics at the university level in various western European universities.

A questionnaire was sent to 150 higher education institutions in 16 countries, and some 50 usable replies were received. It is mainly on this information that the book has been based. The first draft, prepared by the late Professor F. Fiala, has been modified to take into consideration the comments and additional information provided by a group of mathematicians at a meeting held in Strasbourg in February 1969.

The book concentrates on mathematics at the undergraduate level, and on those students for whom mathematics is a principal subject. Information is given concerning university admission requirements, length of study, the various degrees or diplomas offered, course content, general organisation of examinations, teaching conditions. Within these various sections brief details are given concerning the situation in the various countries. Teaching methods applicable to mathematics in higher education institutions are touched on indirectly.

The aim of the book is to contribute "to a search for criteria enabling some sort of equivalence to be established between the materials studied, in the hope of ultimately being able to attain a legal recognition of equivalences between the academic qualifications". In so doing, attention is drawn to the great divergence in university studies not only at the international level but also, in certain cases, at the national level.

Companion Volumes

HOW TO VISIT A MUSEUM

by Pierre REBETEZ

Strasbourg, 1970, 136 pages. Distribution free of charge.

The main aim of the study is to encourage schools and museums to unite their efforts to further the use of the latter for teaching purposes and to promote the full development of creative facilities. In emphasising its educational function, the book seeks to show the different capacities of the museum : its vitality, its possibilities of contact with the public, its organisation.

The relationship museum-school is examined in three chapters : the museum, its aims and its means ; the school curriculum and the museum's activities ; the museum as a school.

The author is more specifically concerned with the 13-18 age group. He also stresses the importance of better collaboration between authorities, teachers and curators, so that museums may be used more effectively by schools.

Other publications

PAEDAGOGICA EUROPAEA :
THE CHANGING SCHOOL CURRICULUM IN EUROPE. - VOL. VI.

L. C.G. Malmberg, N. V. Uitgever and Georg Westermann Verlag, 1971, 268 p.

The theme of the latest issue of the European Yearbook of Education Research is "The Changing School Curriculum in Europe", a theme which dominates the European educational scene. Eminent authors from France, the Federal Republic of Germany, Sweden, Switzerland and the United Kingdom have contributed seventeen studies which cover the major aspects in the field of curriculum research and development and which show the convergence of developments and trends. The editor, Professor S. J. Eggleston summarises this convergence as follows : "The moves to 'curriculum autonomy' have introduced a number of further variables previously unknown and certainly unanticipated by the educators who initiated 'curriculum development'. Most notably these are the decision making powers that have been claimed and won by the clients of the system — the parents, students and pupils."

The reader will find this volume to be a most useful, and indeed necessary, handbook on the changing nature and determination of the school curriculum in western Europe. He may regret, however, that the British element dominates this issue and that the French and Scandinavian language areas are underrepresented. There is no Italian contribution. It is hoped that the next two volumes which are to deal with the diversification of post-secondary education and guidance and assessment respectively, will remedy this imbalance.

Paedagogica Europaea has now found its place and role in the international book market as a European forum for the discussion of educational issues. The possibility of it being published biannually, and in a cheaper edition, would most certainly be welcomed. This would enable *Paedagogica Europaea* to follow more closely developments in the rapidly expanding field of research and innovation in education in Europe, and also to reach a wider public.

Editor : The Director of Education and of Cultural and Scientific Affairs
Strasbourg