THE

INTERNATIONAL ENCYCLOPEDIA OF EDUCATION

Research and Studies

Volume 2

C

Editors-in-Chief

TORSTEN HUSEN

University of Stockholm, Sweden

T. NEVILLE POSTLETHWAITE

University of Hamburg, FRG

PERGAMON PRESS

OXFORD · NEW YORK · TORONTO · SYDNEY · PARIS · FRANKFURT

caused such widespread concern that in the Federal Republic of Germany, the Philippines, and Singapore, for example, video arcades have been banned.

However, the success of the video game offers a challenge to the educator. The computer is the basis of the modern-day video game, but it is apparent that much of the educational software initially developed for computer-based learning does not have the same addictive qualities. To understand better the pleasure that students derive from computer-based games and to build this in to future educational programs is an important short-term goal for educational psychologists.

It is difficult, now, to gauge the full significance of "the computer revolution." Being in the middle of it, researchers lack the perspective to adequately weigh such assessments, but it is already clear that computers have brought about fundamental changes in the way various classical problems in education are approached.

See also: Computer Technology and Telecommunications; Computer-managed Learning; Computer-assisted Learning

Bibliography

Bork A M 1981 *Learning with Computers*. Digital Press, Bedford, Massachusetts

Ellis A B 1974 *The Use and Misuse of Computers in Education*. McGraw-Hill, New York

Metropolis N C, Howlett J, Rota G-C (eds.) 1980 *A History of Computing in the Twentieth Century*. Academic Press, New York

Minsky M L 1967 *Computation, Finite and Infinite Machines*. Prentice-Hall, Englewood Cliffs, New Jersey

Papert S 1980 *Mindstorms: Children, Computers and Powerful Ideas*. Basic Books, New York

Seidel R J, Rubin M (eds.) 1977 *Computers and Communications: Implications for Education*. Academic Press, New York

Smith S G, Sherwood B A 1976 Educational uses of the PLATO computer system. *Science* 192: 344–52

Statistical Analysis Systems 1979 *Introductory Guide*. Statistical Analysis Systems Institute, Cary, North Carolina

Turing A M 1937 On computable numbers, with an application to the Entscheidungsproblem. *Proc. London Mathematical Society, Series* 2. 42: 230–48

Weizenbaum J 1976 *Computer Power and Human Reason: From Judgement to Calculation*. Freeman, New York

B. H. Choppin; D. McArthur

Confidence Marking

According to Ebel (1965), "the term confidence weighting refers to a special mode of responding to test . . . items, and a special mode of scoring those responses . . . . The examinee is asked to indicate not only what he believes to be the correct answer to a question, but also how certain he is of the correctness of his answer. When his answers are scored he receives more credit for a correct answer given confidently than for one given diffidently. But the penalty for an incorrect answer given confidently is heavy enough to discourage unwarranted pretense of confidence."

This article considers how the credits and penalties are assessed.

1. The Underlying Models

The choice of a level of confidence (from amongst those available) must be considered in the perspective of decision theory. Technical problems such as validity, reliability, and acuity of confidence answers must be based on a sound model of mental activity, on carefully written instructions, and on the selection of appropriate tariffs. The tariffs are the points attributed: for a correct response (T_c), for an incorrect one (T_i), or for an omission (T_o). A set of three tariffs (T_c, T_i, and T_o) is called a *t*-scale. The best known *t*-scale is the *St*-scale (simple *t*-scale) where $T_c = +1$, $T_i = 0$, and $T_o = 0$. The expected score on a given question (ESQ) is computed according to the following formula:

$$\text{ESQ} = (p \cdot T_c) + (q \cdot T_i) + (r \cdot T_o)$$

where p = probability of a correct answer, q = probability of an incorrect answer, and r = probability of an omission.

1.1 Classical Models and Corrections for Guessing

In the usual "correction for guessing" scoring formula, T_i is equal to $-(1/k-1)$. This tariff characterizes the *Gt*-scale (where *G* stands for Guessing), contrived in such a way that ESQ = 0 when the probability for a successful guess is $1/k$, and is based on the model "Students who do not know the correct answer choose randomly among the k alternatives." Procedures such as eliminating the incorrect alternatives instead of choosing the correct one (Coombs et al. 1956) are based on a second model of test taking behavior: "Students who do not know the correct answer first eliminate the alternatives they know to be wrong. Then they choose randomly amongst those that remain." This gives rise to a generalized *Gt*-scale that allows for the existence of partial knowledge.

It is well-known that the *Gt*-scale is unfair. In some cases, it undercorrects and in others it overcorrects (see *Guessing, Correction for*). Other statistical "corrections for guessing" have been proposed (Chernoff 1962), but in spite of their sophistication, these approaches are not sound for two reasons. First, the percentage of students that have chosen a given alternative (i.e., the alternative's popularity) is not a good index of the (subjective) individual attractiveness of the alternative. Second, item difficulty must be combined with students' ability to explain the probability of a correct answer (a basic assumption of the Rasch model) (see *Rasch Measurement Models*).

1.2 Subjective Probabilities

The apparent difficulty of a question is a subjective experience since it depends upon the student's own ability that varies from individual to individual. The way to cope with this is to take into account the personal attractiveness of the alternatives (i.e., the individual's subjective estimate of the probability that each alternative is correct). Accordingly, the tariffs depend on the subjective probability *and* on the correctness of the response. This suggests a matrix of tariffs. In educational tariff matrices the T_c values are positive, the T_i values are negative, and the T_o value is equal to zero.

This third model is distinguishable from the earlier ones because it does not refer to the dichotomy "when the student knows . . . when the student does not know." This model supposes a continuity of cognitive states, from perfect knowledge to null knowledge. Between these two extremes exist many intermediate states of knowledge, of partial information.

Model 3 states that when a student is faced with the various alternatives, he or she attributes to each of them a probability of being the correct one. If a student is requested to give only one answer, he or she will choose the alternative with the greatest probability.

More and more researchers and teachers share De Finetti's opinion (1965): "Partial information exists; to detect it is interesting, necessary and feasible. Instruction in using these methods . . . has, moreover, a high educational value. Such methods, including the way of *scoring*, and not only the response systems, must be appropriately chosen by the experimenter and clearly explained to the subjects who must understand the nature of the game they are playing. If this is done, questions about guessing disappear."

2. The Instructions

In order to obtain "admissible measures of subjective probability" (Shuford et al. 1966), that is, valid and reliable measures, instructions must be carefully designed. It is not sufficient to ask the students to emphasize those answers of which they are confident, or to ask them to choose among the degrees of an ordinal scale (such as "1 = not sure, 2 = fairly sure, 3 = very sure," etc.). Such definitions are vague and will be interpreted in different ways by different students, although despite these shortcomings, they have been frequently used in combination with various tariff matrices.

It is better to express the level of confidence in probabilistic terms. A student can hardly distinguish 0.3 from 0.35, so it would appear useless to request answers of such an acuity. For this reason, a few areas are delimited on the probability scale, so that the student has to choose a zone of probabilities and not pinpoint a single probability. Table 1 gives one set of instructions as used in Belgium (Leclercq 1982), with a nondecision tariffs matrix (A) and a relevant tariffs matrix (B).

Table 1
Tariff matrices for model 3

If you attribute to your answer a chance of success from	then choose confidence degree number	Tariffs *A*		Tariffs *B*	
		T_c	T_i	T_c	T_i
0 to 25%	0	0	0	0	0
25 to 50%	1	+1	−1	+3	−1
50 to 75%	2	+2	−2	+4	−2
75 to 100%	3	+3	−3	+5	−5

3. Psychometric Considerations

The score on a test computed with a *Ct*-scale (Confidence *t*-scale) is a payoff, a reinforcement, but not a measure. It is an arbitrary weighted mixture of two different measures: the measure of the student's knowledge in the content area and the measure of the student's ability in self-estimation. Various indices of this last ability have been suggested (Lichtenstein et al. 1977, Leclercq 1982): indices of individual coherence, realism, calibration efficiency, acuity, and stability in the use of confidence degrees.

De Finetti (1965) stresses that "It is only subjective probability that can give an objective meaning to every response and scoring method."

Nevertheless, more valid or more reliable information about a student's ability in a given domain can be obtained from his or her confidence responses only if he or she is a good estimator of his or her own capacity and if he or she tells the truth.

Tariff matrices are computed according to a decision theory in order to encourage students to express their convictions honestly and without faking.

4. Implementation Considerations

Students usually use poor strategies (e.g., follow optimistic or pessimistic rules) instead of simply telling the truth. They need to experience the consequences of their choices (i.e., the tariffs must be applied to them) in two or three tests before they will fully appreciate that the matrix is "fair" for the honest student.

The whole scoring system is more challenging to the student. The best performance must not only be correct, but moreover be strongly backed by confidence. The procedure should include a "correct-for-severity" coefficient to adjust the final C scores to the usual ones (as computed with the St-scale).

5. Educational Importance

Sound estimates of the chances of success in an undertaking are of importance for physicians, nurses, and surgeons in a hospital, for teachers in a school, for workers in a factory, and for drivers on the road. If someone is confident of his or her correctness, but is wrong half of the time, he or she is a perpetual source of annoyance and danger to himself or herself and to others. It is thus important that adults are able to estimate realistically the chances of success of any project they undertake.

6. Research Interest

In many cases, overt behavior is more related to the individual's beliefs than to objective measures of knowledge.

Research on subjective probabilities in education raises theoretical and methodological problems similar to those that arose in psychology when psychophysics developed. Strict definitions had to be found for "stimulus" and for "response." In the same way, concepts like "knowledge," "doubt," "uncertainty," "confidence," "information," begin to acquire a new operational meaning in the framework of subjective probability.

Subjective probability enables people to assess partial knowledge with valid, reliable, sensitive, and convenient methods and techniques. Specific principles must be respected, and clear distinctions must be made between measures, payoffs, and degrees. When this is done, promising perspectives appear concerning the study of information processing by humans, a central concern for educators.

Bibliography

Chernoff H 1962 The scoring of multiple choice questionnaires. *Ann. Math. Stat.* 33: 375–93

Coombs C H, Milholland J E, Warner F B 1956 The assessment of partial knowledge. *Educ. Psychol. Meas.* 16: 13–37

De Finetti B 1965 Methods for discriminating levels of partial knowledge concerning a test item. *Br. J. Math. Stat. Psychol.* 18: 87–123

Ebel R L 1965 Confidence-weighting and test reliability. *J. Educ. Meas.* 2: 49–57

Leclercq D A 1982 Confidence marking: Its use in testing. *Eval. Educ.* 6(3)

Lichtenstein S, Fischhoff B, Phillips L 1977 Calibration of probabilities: The state of the art. In: Jungermann A H, De Zeeuw G (eds.) 1977 *Decisionmaking and Change in Human Affairs*. Proc. of 5th Research Conf. on Subjective Probability, Utility, and Decision Making, Darmstadt, 1–4 Sept, 1975. Reidel, Dordrecht

Shuford E H, Albert A, Massengill H E 1966 Admissible probability measurement procedures. *Psychometrika* 31: 125–45

D. A. Leclercq

Confucian Theory of Human Development

Over the centuries in Chinese culture four books have represented the core of the Confucian philosophical tradition. The first of these is the *Analects of Confucius*, a collection of brief proverbial sayings attributed to the teacher and social philosopher after whom the tradition was named, Confucius or Kung Fu-tzu (551-479 BC). Second is *The Book of Mencius*, composed by the sage Mencius or Meng-tzu (circa 372–289 BC), a student of Confucius' grandson. The last two books, *The Great Learning* and *The Doctrine of the Mean*, were written somewhat later but are still considered to be faithful representations of Confucius' beliefs.

In Chinese society prior to the twentieth century, these four classics served as the material to be mastered by candidates for the civil-service examinations that determined who would be awarded positions in the imperial government. Today the four continue to be viewed by followers of Confucian philosophy as the most authentic description of the proper conduct of life in the well-ordered political state. The four can be used as well as the source of a Confucian theory of proper human development.

The version of Confucian theory offered here has been formulated by the writer's addressing a series of questions about human development to the contents of the four books, then organizing the derived answers under four topics: (a) personality structure, (b) the goal of development, (c) learning theory, and (d) implications for child rearing and education.

1. Personality Structure

From a Confucian perspective, human nature is divided into two general types, the physiological and the virtuous. The physiological represents lower level characteristics and is disposed toward certain aspects of the world of behavior—the mouth disposed to tastes, the eye to colors, the ear to sounds, the nose to odors, and the four limbs to ease or rest. The higher level of human nature is that of virtue, which makes the difference between humans and brutes. Unlike animals, humans innately possess a heart that is sensitive to the suffering of others. This sensitive heart is a composite of the four hearts containing virtues—hearts of compassion, of shame, of courtesy and modesty, and of a sense of right and wrong.

In the heart of compassion is the seed of the virtue of humanity (jen). The heart of shame is the basis for righteousness (i). The heart of courtesy and modesty is the source of the observance of the Confucian