League with accelerators of greater than 30 Gev and beams traveling clockwise, an AAA league with energies between 3 and 30 Bev, etc., etc. Further possible extensions of this system, such as "trading" of team members, relegation of members to minor leagues, and so forth, are left to the imagination of the reader.

We remain, Sir, etc., etc.

Ralph Emerson Weston
R. Emerson Weston
R. E. Weston
Ralph E. Weston
and several others
Library Lane, Brookhaven, L. I., N. Y.

References

1. Brown and 16 others, Phys. Rev. Letters **8**, 255 (1962).
2. CERN, Ecole Polytechnique, and Saclay, ibid. **8**, 257 (1962).
3. F. Reif, Science **134**, 1957 (1961).
4. S. A. Goudsmit, Phys. Rev. Letters **8**, 229 (1962).

Information Retrieved

In an article beginning on page 22 of the February 1962 issue, it is stated that existing information retrieval tools are inadequate for compilations of data. It is suggested—perhaps rhetorically—that readers should ask their library to find three quoted references. I took the authors at their word and put the enquiry to the Harwell Library in the following terms:

"Could you please try and find references to compilations of information on the following subjects:

(a) Thermal neutron capture gamma rays
(b) Range and energy loss of charged particles in aluminum
(c) Radial Coulomb integrals

A single reference on each subject would suffice."

The library staff were not told of the nature of the enquiry, and using perfectly conventional library methods (i.e., consultation of the catalogues and abstract journals, reference to experts on the site, and use of the searcher's memory) the following references were produced:

(a) Atlas of γ-Ray Spectra From Radiative Capture of Thermal Neutrons, by L. V. Groshev, V. N. Lutsenko, etc. Translated by J. B. Sykes. Pergamon Press, London, 1959, pp. 198.
(b) Nuclear Data Tables. US Atomic Energy Commission, 1960. Part 3, pp. 1–19.
(c) Radial Coulomb Integrals. Iron Series Hartree-Fock Calculations. MIT Tech. Report No. 12, 1959.

You will notice that the references produced are in all three cases to more recent articles than those quoted in your article. Had the searcher looked for additional references, there would have been no difficulty in tracing those quoted as they all appear in *Nuclear Science Abstracts* or in the Harwell Library Subject Catalogue. Both these tools were used in the search, but the searcher worked back from the most recent reference.

R. M. Fishenden
AERE, Harwell, England

Testing Again

Dr. Frank J. Fornoff,[1] of the Educational Testing Service, has taken on the task of defending ETS in the matter of certain multiple-choice science questions used by me in a challenge to that organization.[2] But, as will be shown, he has not taken the preliminary step of carefully reading what I have written.

His principal defense of the "potassium" question consists of an application of the well-known "Daddy knows best" routine, which can work wonders when one is dealing with children: he says that the question "was reviewed by a group of competent college and school chemistry teachers, and, in the context for which it was intended (high-school chemistry), they were satisfied it had merit". If the question is defective, this type of defense carries implications that Dr. Fornoff may well find surprising.

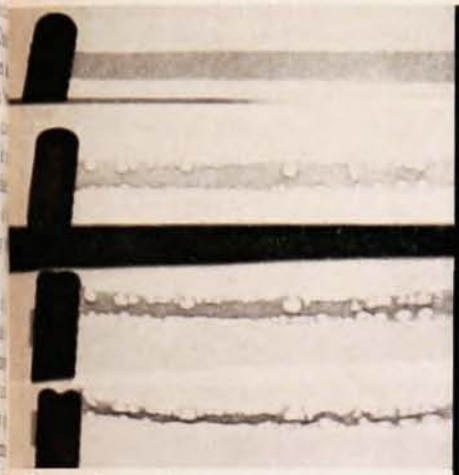
Let us look at the other aspects of his defense of the "potassium" question. His inability to appreciate the strength of the hierarchical relationship of the answers does not necessarily do him credit. And his illustration of the way the committee system for test review operates shows that he failed to understand my argument, for he is under the misapprehension that Mr. Hart[3] and I are in disagreement, though the fact is that I had been well aware of Mr. Hart's cogent point and had specifically stated, in *Testing*, that it was *for the sake of argument* that I was assuming that the potassium and lithium were supposed to be in a gaseous state—an act that Mr. Hart characterizes as "too generous" on my part.

Again, it is strange that Dr. Fornoff should tell of the multiple-choice testers' "belief"—I would myself have said "hope"—that "the rare high-school student who has pushed his knowledge of the field far beyond his textbook will be sophisticated enough to recognize the context associated with the question and will answer it correctly." Can it be that Dr. Fornoff does not realize, for example, that he is here making a damaging admission of the fact that the question penalizes the "rare" student? Though the "rare" student has negligible effect on the testers' statistics, he is not therefore unimportant, and one should not dismiss him thus easily. Nor should one impose on him the burden of estimating what degree of "sophistication" to apply to a particular multiple-choice question, especially since the appropriate degree of "sophistication" varies from question to question. Indeed, one would do well not to condone tests that penalize him for failing to subordinate truth to "sophistication".

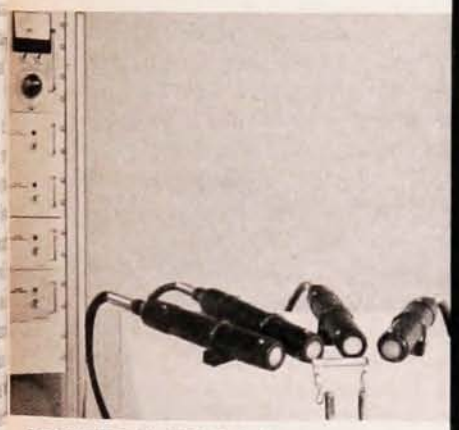
Dr. Fornoff's defense of the "potassium" question is notable for what he does not say. He makes no reference at all to my charges, in *Testing*, that ETS' official defense of this question, in *Explanation*, was invalid, contained elementary blunders in science, used an argument that boomeranged, and revealed that ETS did not understand what its own question was about. These are strong charges. Did Dr. Fornoff read *Testing* in such haste that they escaped his notice? If so, was

You can do more with a Flash X-ray tube that has:

1. small size

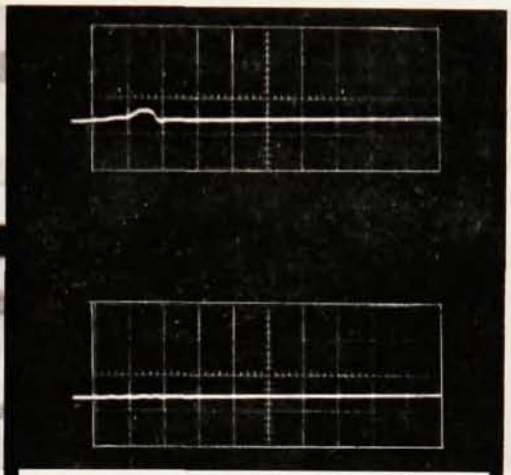


Cineradiographs:
of exploding copper foil show magnetic saw and pinch effects, avoid "blinding" visual flash.

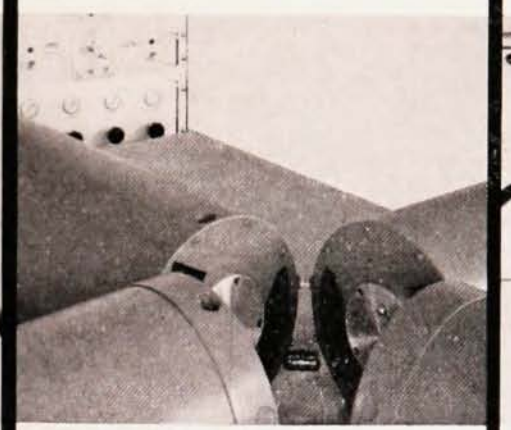


FEXITRON 730-4-C/232
Several tubes may be located in a small volume and flashed in sequence to avoid film motion; this system has four each tubes, pulsers and delay generators for interpulse periods from 1 to 1000 μsec; voltage 100 kv, dose rate of 10^7 rads/sec, choice of pulse lengths from 30 to 100 nanoseconds; Price $15,170.

2. high dose rate

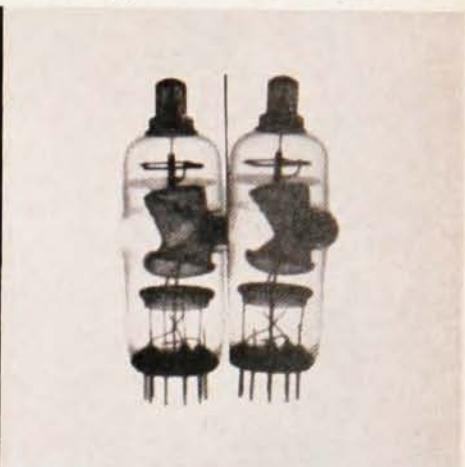


Radiation Effects:
Current of several ma induced in dielectric (capacitor) at 10^8 rads/sec; lower trace shows low noise level.

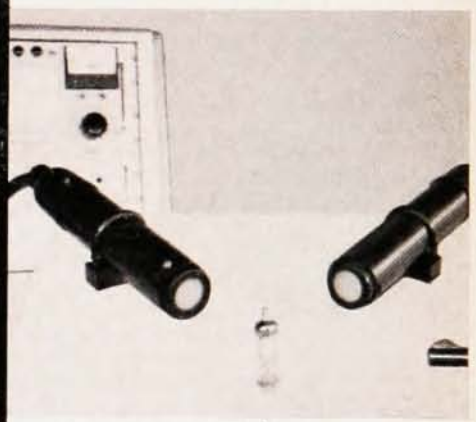


FEXITRON 730-4-233
Multiple x-ray sources are fired simultaneously to maximize dose rate and uniformity of illumination, or in time sequence (1-1000 μsec) to provide a pulse train; 300 kv, cold cathode tubes are operated inside grounded, metallic pulser can to give high x-ray signal-to-noise; dose rate 10^8 rads/sec (single tube) and pulse length choice 30 to 100 nanoseconds; price of four channel system as shown, $29,550.

3. high resolution



3-D Radiographs:
bullet stopped in flight penetrating vacuum tube; will resolve 7 mil particles at 16,000 fps.



FEXITRON 730-C/232
Source sizes from 1 to 4 mm provide high resolution; two x-ray tubes fired simultaneously with a common film provide 3-D views; this system available with the same range of pulse lengths, and the same voltage, dose rate and interpulse periods as the 730-4-C/232; all systems shown are complete with 30 kv dc supply, cabinet and nitrogen regulator (pulsers are pressurized to isolate performance from atmospheric effects); Price $8,715.

Field Emission Corporation

Telephone 472-5101

McMinnville, Oregon

FOR ADDITIONAL INFORMATION: Write or call Field Emission Corporation, Dept. T-1, McMinnville, Oregon ■ Telephone: 472-5101, or our field office: Riviera Central Bldg., Dept. T-1, 205 Avenue "I", Redondo Beach, California ■ Telephone: 375-5510

there no one in ETS to bring them to his attention and to suggest that he refute them?

The "prism" question asks why "a ray of white light is broken up into a spectrum by a prism of colorless glass." Dr. Fornoff and I agree that answer *B*, "white light is a composite of many frequencies," gives a necessary but not a sufficient reason. He asserts, though, that answer *D*,[4] "the amount of refraction differs for light of different wavelengths," is "more nearly, though not perfectly, both a necessary and a sufficient condition."

To justify this assertion, he says that *B* is "not a sufficient condition. A plane mirror may receive the [polychromatic] light and yet produce no spectrum." Suppose, though, that *B*, instead of *D*, had been the wanted answer. We can well imagine him then pointing out that if white light were monochromatic it would not be "broken up into a spectrum" by the prism or by anything else even though *D* were valid. Neither *B* nor *D* is at all a sufficient condition, but, as I explained in *Testing*, *B* is a deeper answer than the wanted answer, *D*. And I repeat here what I said there: that in the case of the good student this question "does not measure understanding of science so much as understanding of the workings of the mind of the test maker".

Believing that he has made his point Dr. Fornoff goes on to remark, not without a show of virtue, that "to eliminate judgments such as that required in this question and that asked for in the potassium electron question, tends to reduce a test to questions which can be answered by rote memory and so lead to a test favoring students with photographic memories". It does indeed. And his remark emphasizes the fact that test experts who seek to make multiple-choice questions that cannot be answered by rote memory resort too often to ambiguity and worse.

In defending the "gasoline" question, Dr. Fornoff again falls back on the dangerous "context" argument, stating, as his principal defense, that "there is a reasonable context associated with the question and that this context is normally discerned by chemistry students who understand chemical change". This, however, is merely a reiteration of part of ETS' official defense of this question, in *Explanation*, and if Dr. Fornoff wishes to know my response he has but to refer to *Testing*.

On February 14, 1956, some time before I reluctantly decided to make a public issue of testing, I pointed out, in the course of correspondence with ETS, how $E = mc^2$ affected the "gasoline" question. Referring to this communication, Dr. Fornoff says that "the ETS staff at that time recognized that some people might miss the context in which the question was written" and therefore changed the wording of answer *E*. In mentioning this episode, which ETS had not done in its official defense, Dr. Fornoff can hardly be regarded as having bolstered the cogency of his "context" defense of the "gasoline" question.

Even if we did accept the "context" defense, the fact

would remain that ETS displayed professional ineptitude in framing the "gasoline" question, and that Dr. Fornoff condones this ineptitude by defending the question. For it clearly favors the candidate who, because he is totally ignorant of the significance of $E = mc^2$, faces no problem regarding the "context". Indeed, for such a candidate mere superficial rote learning suffices —a fact that Dr. Fornoff unwittingly emphasizes by his quotation from *Chambers' Encyclopaedia*. What is surprising, and disquieting, about Dr. Fornoff's reiteration of ETS' official line of defense is that it gives no indication that he took cognizance of my analysis of this defense in *Testing*, where I wrote: "Note how damaging are the implications if we do assume that ETS was fully aware of the meaning of $E = mc^2$ and deliberately included answer E nevertheless. For we must then ask: what was its motive in doing so? To make a question with no correct answers? Let us hope not. Then what? To penalize the superior student? One doubts that ETS would say so; yet the question is surely easier for the student who does not understand $E = mc^2$ than for the student who does. Is the latter student supposed to compensate for the deficiencies of the test maker by reading possibly hazardous amendments into the question as worded? That way lies chaos—not 'objectivity'. If the superior student does decide to pick answer E, does he not do so with contempt for the test maker, and with cynical disregard of scientific facts? Should he be rewarded for his willingness thus to place expediency above scientific integrity? If tests are training students to respond in this way, are they not having a deleterious effect on education?"

Dr. Fornoff makes no mention of the conclusions regarding the above three questions reached by the Western Pennsylvania Section of the American Association of Physics Teachers.[5] He probably realizes that they offer no aid or comfort to ETS. If he, or ETS, believes otherwise, and will specify precisely why, I shall be happy to discuss the matter in detail.

In *Harper's*, I pointed out that if the testers "defend a bad question by their 'statistics show . . .' maneuvre, they risk the implication that their use of statistics is improper or that their statistics are untrustworthy," and that "if they defend it by pointing to the high caliber of their staff experts and consultants they may well start people wondering whether the caliber is high enough." These warnings, as well as the related points raised by Mr. Hart, seem to have been ignored by Dr. Fornoff. He brings in statistics whose limitations are manifest, and he boasts implicitly of the caliber of ETS' consultants by boasting of the caliber of the institutions of learning with which some of them are associated. My object in exhibiting defective questions and challenging the testers to defend them specifically was to make a sharply focused prima-facie case for the setting up of a distinguished committee of inquiry to look into the whole matter of testing. Dr. Fornoff has resorted to the testers' routine defenses, ignoring the fact that the challenge strategy was spe-

cially designed to turn these defensive weapons into boomerangs. He has not only ignored the essential points in my criticism, in *Testing*, of ETS' official defense, in *Explanation*, but has answered my challenge mainly by rote, and I am loath to think that his article constitutes ETS' "best answer".

Banesh Hoffmann
Queens College, N. Y.

References

1. Frank J. Fornoff, *Physics Today*, April 1962, p. 36.
2. See Banesh Hoffmann, *Harper's Magazine*, March 1961, p. 37; "Explanation of Multiple-Choice Tests," (April 1961) obtainable on request, and without charge, from Educational Testing Service, Princeton, N. J.; and Banesh Hoffmann, *Physics Today*, October 1961, p. 38. For convenience, these are referred to, respectively, as *Harper's*, *Explanation*, and *Testing*.
3. See Robert Hart, *Physics Today*, February 1962, p. 62.
4. In this connection, see footnote 1 in *Testing*.
5. Robert B. Gray, *Physics Today*, February 1962, p. 62.

And Again . . .

As a former educator (field of physics) but now a member of industry's technical management fraternity, I was taken with the splendid and illuminating articles dealing with "Objective Tests" which Dr. Hoffmann (*Physics Today*, October 1961, pp. 38–42), and Dr. Fornoff (*Physics Today*, April 1962, pp. 36–44) have written.

Yet, perhaps I am somewhat "old-fashioned" in my concepts of education, and as such have leaned in the direction of those who prefer to think of comprehensive final tests, College Board tests, Regents' Examinations (New York State), and other similar types of tests as a means of estimating what a student has learned. I recognize and have carefully excluded those types of tests used for diagnosis which may have as one of their prime objectives an estimate of what a student doesn't know. These tests, too, are valuable for their purpose.

It seems to me, sitting on the sidelines, that both of your presentations only allude to the aforementioned issue, and tend to dissect the validity of particular scientific questions and subsequently their reliability as a forecasting tool. From my own point of view, I am inclined to think that the major issue has been somewhat overshadowed. I am left with a number of questions, the answers to which are not definitive in the articles. What is the prime purpose of the tests? Are written tests (whether essay or objective type) the best means of achieving the purpose? How accurately do the means of evaluating (written tests or other techniques) serve the purpose? How closely do the means typify or simulate "real life" situations or evaluations?

Frankly, I do not purport to know all of the answers to such questions. Nonetheless, it has become evident to those of us in industry that one of the better methods of evaluating happens to be a "guided" discussion with an applicant. By subtle direction, conversations