cross-link. (Of coure, many other factors may also affect the development of specific cross-links.) Thus, the determination of the relative abundance of the reducible cross-links, while potentially interesting, may not provide direct insight into various biologic phenomena which involve collagen. For example, although the abundance of various carbonyl compounds changes as a function of wound healing, age, vitamin D deficiency, animal species, source of tissue, and a hereditary disorder of connective tissue (*28*), it is not clear what factors are responsible for these differences. While it is tempting to speculate that a "Rosetta stone of aging" may originate in the covalent cross-links of collagen, it is too early to come to any reasonable conclusions on this subject. However, it does seem probable that once more structural information is available, the mechanisms of the regulatory processes can be probed in detail.


## Summary

The formation of collagen cross-links is attributable to the presence of two aldehyde-containing amino acids which react with other amino acids in collagen to generate difunctional, trifunctional, and tetrafunctional cross-links. A necessary prerequisite for the development of these cross-links is that the collagen molecules be assembled in the naturally occurring fibrous polymer. Once this condition is met, cross-linking occurs in a spontaneous, progressive fashion. The chemical structures of the cross-links dictate that very precise intermolecular alignments must occur in the collagen polymer. This seems to be a function of each specific collagen because the relative abundance of the different cross-links varies markedly, depending upon the tissue of origin of the collagen.


### References and Notes


1. C. Franzblau, in *Comprehensive Biochemistry*, M. Florkin and E. H. Stotz, Eds. (Elsevier, Amsterdam, 1971), vol 26, part C, pp. 659–712; S. M. Partridge, *Gerontologia* **15**, 85 (1969).
2. F. O. Schmitt, in *Biophysical Science—A Study Program*, J. L. Oncley, Ed. (Wiley, New York, 1959), pp. 349–358; J. Gross, in *Connective Tissue, Thrombosis, and Atherosclerosis*, I. H. Page, Ed. (Academic Press, New York, 1959), pp. 77–92.
3. J. Gross, *Science* **143**, 960 (1964).
4. M. L. Tanzer, *Int. Rev. Connect. Tissue Res.* **3**, 91 (1965).
5. A. S. Narayanan, R. C. Siegel, G. R. Martin, *Biochem. Biophys. Res. Commun.* **46**, 745 (1972).
6. M. L. Tanzer, *J. Biol. Chem.* **243**, 4045 (1968).
7. A. Deshmukh, K. Deshmukh, M. E. Nimni, *Biochemistry* **10**, 2337 (1971); C. Franzblau, A. H. Kang, B. Faris, *Biochem. Biophys. Res. Commun.* **40**, 437 (1970).
8. A. J. Bailey and C. M. Peach, *Biochem. Biophys. Res. Commun.* **33**, 812 (1968).
9. M. L. Tanzer and G. Mechanic, *ibid.* **32**, 885 (1968); G. Mechanic, P. M. Gallop, M. L. Tanzer, *ibid.* **45**, 644 (1971); G. Mechanic and M. L. Tanzer, *ibid.* **41**, 1597 (1970).
10. R. C. Siegel and G. R. Martin, *J. Biol. Chem.* **245**, 1653 (1970).
11. P. Bornstein, A. H. Kang, K. A. Piez, *Proc. Nat. Acad. Sci. U.S.A.* **55**, 417 (1966); K. Deshmukh and M. E. Nimni, *Biochemistry* **10**, 1640 (1971).
12. P. Bornstein and K. A. Piez, *Biochemistry* **5**, 3460 (1966); A. H. Kang, B. Faris, C. Franzblau, *Biochem. Biophys. Res. Commun.* **36**, 345 (1969); M. Rajkind, A. M. Gutierrez, M. Zeichner, R. W. Lent, *ibid.* p. 350.
13. M. L. Tanzer, T. Housley, L. Berube, R. Fairweather, C. Franzblau, P. M. Gallop, *J. Biol. Chem.* **248**, 393 (1973); M. L. Tanzer, R. B. Fairweather, P. M. Gallop, *Biochim. Biophys. Acta*, in press.
14. M. L. Tanzer and G. Mechanic, *Biochem. Biophys. Res. Commun.* **39**, 182 (1970); A. H. Kang, B. Faris, C. Franzblau, *ibid.* p. 175.
15. M. L. Tanzer, R. Fairweather, P. M. Gallop, *Arch. Biochem. Biophys.* **151**, 137 (1972); S. P. Robins and A. J. Bailey, *Biochem. Biophys. Res. Commun.* **48**, 76 (1972).
16. M. L. Tanzer and R. Fairweather, unpublished data.
17. W. Traub and K. A. Piez, *Advan. Protein Chem.* **25**, 243 (1971).
18. A. H. Kang, *Biochemistry* **11**, 1828 (1972).
19. E. J. Miller, *Biochem. Biophys. Res. Commun.* **45**, 444 (1971).
20. O. DeLuque, G. Mechanic, M. L. Tanzer, *Biochemistry* **9**, 4987 (1970).
21. Y. Kubokl, M. L. Tanzer, G. Mechanic, *ibid.*, in press.
22. M. L. Tanzer and P. M. Gallop, unpublished data.
23. K. Deshmukh and M. E. Nimni, *Biochem. Biophys. Res. Commun.* **46**, 175 (1972).
24. R. Fairweather, personal communication.
25. B. Pereyra, O. O. Blumenfeld, M. A. Paz, P. M. Gallop, *Fed. Proc.* **31**, 480 (1972).
26. A. J. Bailey, *Biochim. Biophys. Acta* **221**, 652 (1970); D. R. Eyre and M. J. Glimcher, *ibid.* **243**, 525 (1971); A. J. Bailey, *Fed. Eur. Biochem. Soc. Lett.* **18**, 154 (1971).
27. M. L. Tanzer and N. A. Kefalides, *Biochem. Biophys. Res. Commun.*, in press.
28. A. J. Bailey and M. S. Shimokomaki, *Fed. Eur. Biochem. Soc. Lett.* **16**, 86 (1971); A. J. Bailey and S. P. Robins, *ibid.* **21**, 330 (1972); L. Forrest, A. Shuttleworth, D. S. Jackson, G. Mechanic, *Biochem. Biophys. Res. Commun.* **46**, 1776 (1972); G. Mechanic, *ibid.* **47**, 267 (1972); G. L. Mechanic, S. U. Toverud, W. K. Ramp, *ibid.* p. 760.


29. Supported by research grants from NSF (GB-31077), NIH (2R01 AM 12683-04), American Heart Association (71-704), and Life Insurance Medical Research Fund (G-69024).


# Evaluation of Instruction


Peter K. Gessner


An essential role of teaching institutions is to teach—that is, to transmit knowledge, skills, and attitudes (*1*). The degree of success different institutions achieve in this respect may be viewed as dependent on the intellectual caliber of both its incoming students and its faculty and on the teaching effectiveness of that faculty. To ensure high intellectual caliber among their incoming students, many institutions place heavy emphasis in their admission policies on the applicant's score on national normative examinations such as the Scholastic Aptitude Test, the Graduate Record Examination, the Law School Admission Test, the Medical College Admission Test, and the Dental Aptitude Test. Likewise, to ensure teaching effectiveness among their faculty, teaching institutions may be reasonably expected to foster it by their promotion policies. Such institutional efforts, however, are likely to be hindered by the lack of agreement regarding the criteria upon which to base methods for measuring teaching effectiveness quantitatively.

The high correlation found between academic rank and publication output (*2*) supports the widespread impression that publication output has been the major determinant in promotion decisions (*3*). As the premise that good researchers are ipso facto good teachers is being challenged (*2*, *4*, *5*), increasing attention has been given to more direct methods of measuring teaching effectiveness (*2*, *6–9*). Among those proposed are (i) student ratings of instruction and (ii) class performance in examinations. Although both methods have certain disadvantages, the expressed desire of students to participate in the evaluation of courses and the view that students, as customers of the educational service, are in the best position to evaluate its worth (*7–9*) have resulted in increasing use of student ratings (*2*). This can be considered a reasonable devel-


The author is associate professor of pharmacology, State University of New York at Buffalo 14214.

opment if the two methods mentioned above measure different, but related, aspects of an individual's teaching effectiveness. If this be so, one would expect to find a substantial positive correlation between teaching effectiveness as measured by student ratings and teaching effectiveness as measured by class performance on examinations. However, a recent report (*7*) suggests that, instead, a negative correlation exists and concludes that students rate most highly those instructors from whom they learn the least. The momentous implications of such conclusions for the nation's educational system suggest that both a searching and critical analysis of these conclusions and the publication of other relevant data be undertaken as matters of high priority.

The data on which the present study is based were collected under conditions that were more controlled than is usually the case in studies of this type. Thus, by virtue of their participation in a single course, all the faculty of a department were rated by the same group of students. Moreover, the measure of class performance used primarily in this study (to measure relative class performance in the different subject areas of the course) was one computed from how well, relative to the national sample, the class answered questions on a national normative examination, an examination about the content of which none of the faculty had any foreknowledge. For comparative purposes, class performance in the same subject areas on departmental examinations was also computed.

## Design of the Study

The class consisted of 119 sophomore medical students taking a one-semester basic science course. The instructors were the department's ten faculty members. Each assumed sole responsibility for instruction in one or more of the 23 subject areas of the course. Part of the responsibility for instruction in a given subject area was the preparation of a corresponding section for a course manual. This manual, numbering some 560 pages of lecture outlines, notes, tables, diagrams, and so forth, was made available to the students at a modest fee and was purchased by all of them.

Students were graded pass or incomplete (failure) by the department on the basis of their cumulative raw grades on three departmental examinations. These examinations were prepared by a departmental committee from questions submitted by individual faculty members in the subject areas for which they were responsible. Special emphasis was placed on the elimination of ambiguous or unfair questions. One examination was administered at the end of approximately each third of the course. In an effort to further ensure the fairness of the examination (and, in part, to make it more of a teaching device), students were encouraged to submit written challenges of individual examination questions. All such challenges were reviewed by the departmental committee. If the challenges were considered valid, alternative answers were accepted or the question was removed from scoring. Details of the grading procedures are presented because of the possibility that the procedures could affect the ratings of instructors (*7, 10*).

Student ratings of teaching effectiveness were obtained by means of a questionnaire that the 78 students attending the last lecture of the course were asked to complete anonymously. Students were asked to rate each of the subject areas with regard to (i) content and organization and (ii) presentation, on a three-point scale "good," "satisfactory," and "unsatisfactory."

The ratings were converted to a numerical scale by (i) assigning the values $+1$, 0, $-1$ to the three attributes, (ii) using for each attribute a weighting factor reflecting the information content of the response, and (iii) calculation of a weighted mean rating for each subject area (*11*). The second of these steps was introduced because the frequency with which the three attributes were employed by the students in their responses was not uniform. On the basis of information theory, the information content of a response is proportional to $-\log P$, where $P$ is the population value for the probability of occurrence of the response (*12*). Thus, the smaller the probability of the response occurring, the greater the information content of the response if it does occur. The weighting factor, $w$, for each attribute was calculated from $w = -\log_2 f$ where $f$, the total frequency with which that attribute was elicited as a response (*13*), was used as a sample estimate of $P$ (*14*).

Class performance relative to a national norm was evaluated by an analysis of how the 141 questions on Part I of the National Medical Board Examination were answered by the 116 students (97.5 percent of the class) who elected to take this examination 5 weeks after the end of the course. Each of the questions was inspected and classified as to whether or not it pertained to material covered in one of the subject areas (*15*) by two members of the faculty, who referred, when necessary, to the course manual to ensure correct assignment. The National Medical Board provides the percentage of both the department's students and the nationwide sample who answered each question correctly. The latter can be considered as a normative measure of the inherent difficulty of the questions. Accordingly, the difference between the two percentages, averaged for all of the questions pertaining to a given subject area, was used as a measure of class performance in that subject area.

With respect to the departmental examinations, since no normative information was available, the percentage of students answering a question correctly was taken as the class performance for that question; class performance for a given subject area was defined as the mean class performance for the questions pertaining to that subject area.

## Results and Discussion

The correlation coefficients for the two measures of teaching effectiveness—namely, class performance in 20 subject areas on a national examination and the student ratings of the content and organization and of the presentation of course instruction in these areas—were .77 and .69, respectively. In both instances the results are statistically highly significant ($P < .001$). The 95 percent confidence limits for the true value of these correlations in the population are given by $.52 \leq \rho \leq .90$ and $.43 \leq \rho \leq .86$, respectively. Figure 1 shows the scatter diagram for the first of these two measures.

One variable that could have spuriously raised the above correlation would be the relative emphasis accorded to a given subject area in the course and the national examination. Thus, the greater the relative amount of time devoted to a subject area in the course, the greater the opportunity to cover the subject in depth; and this might lead to both better class performance and higher student ratings. Accordingly, a normalized measure of

the relative emphasis accorded a given subject area in the course and the national examination was calculated by subtracting the percentage of examination questions devoted to that subject area from the percentage of lectures devoted to it in the course and then dividing the resulting difference by the mean of the two percentages. The correlation between this variable and class performance was found to be .43, indicating a positive trend, although statistically not a significant one ($P = .06$). Moreover, the partial correlation coefficient ($r_{12.3}$) for class performance on the national examination and student ratings of content and organization and of presentation in the various subject areas, with relative emphasis held constant, were .74 and .62, respectively. These values are also highly significant ($P < .001$) and are not very different from the values of .77 and .69 presented above (*16, 17*).

Another variable that could have spuriously raised the correlations between student ratings and class performance on the national examination would be the time elapsed between when a given subject area was covered in the course and the date of the national examination. The greater the amount of time that had passed, the greater the opportunity for students to have forgotten the material. The correlation between the number of days elapsed and class performance on the national examination ($r = -.39$), however, was not a significant one ($P = .15$). Moreover, the partial correlation coefficients between student ratings and class performance on the national examination with the time elapsed held constant ($r_{12.3} = .77$ for content and organization, .69 for presentation) were identical to the correlation coefficients obtained without the time elapsed being held constant. Thus, it is apparent that differences in the time elapsed between a subject area's being covered in the course and the date of the national examination had no effect on the correlations reported.

Results of student and class performance on departmental and national examinations present somewhat of a paradox. The correlation between how well individual students scored in the departmental and national examinations was quite high ($r = .68$; $P < .001$), which suggests that the scores of individual students were a good measure of their individual proficiency relative to national norms. On the other hand, the correlation between class performance in the 20 subject

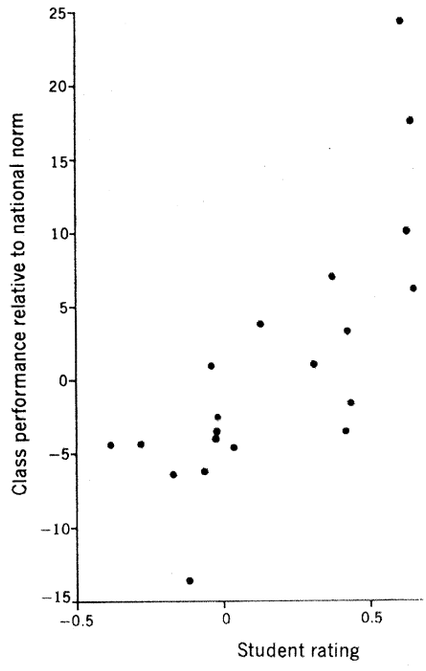


Fig. 1. Scatter diagram of class performance on a national normative examination in 20 subject areas of a course and of the student ratings of the content and organization of instruction in these subject areas.

areas on the departmental and national examinations was rather low ($r = .18$). This latter finding is not so startling if one remembers that the way class performance on the departmental and national examinations was computed differed, there being no way to allow for variations in difficulty among questions in the departmental examination.

The correlation between class performance on the departmental examinations and student ratings was quite low ($r = .11$ for content and organization; $r = .17$ for presentation). Although the reason for this low correlation is also probably related to the way class performance was computed, this result does point to the problems inherent in using class performance on internal examinations as a measure of the teaching effectiveness of the faculty.

The validity of all of these correlations is enhanced by a number of factors. The faculty were all rated by the same group of students; they were not involved in the grading of the students except through the medium of departmental examinations. These examinations were made as emotionally neutral as possible by being composed exclusively of multiple choice questions and by being subject to student challenge. The students were graded pass or incomplete, with a historically derived expectation of a low incomplete rate (5 out of 124 students received an incomplete grade, a rate not significantly different from that of the previous year). Finally, the students rated the instruction before the last departmental examination. Thus, their final grade could not have affected their ratings. Moreover, their rating of instruction in subject areas covered by the first two examinations, the results of which they knew, and by the last examination, which had not yet been given, did not differ significantly (mean ratings ± standard error were .16 ± .08 and .21 ± .20, respectively). This and the low correlation between student ratings and class performance on departmental examinations suggests that student ratings of instruction were not affected by the nature or perceived difficulty of the questions on departmental examinations. In addition, it might be noted that, since class performance on the national examinations was computed from the performance of the class relative to the performance of the nationwide sample, it would not have been affected by the inherent difficulty of the subject matter in the different subject areas (*18*).

Insofar as class performance on the national examination measured learning, it reflected material retained for a period of 5 weeks after the end of the course. The questions on the national examination were of the multiple choice type, and the grading was therefore wholly objective. As pointed out earlier, the instructors had no foreknowledge of the questions on the national examination.

In considering the marked discrepancy between the data reported here and the work of Rodin and Rodin (*7*), it should be noted that, although the Rodins claim that their examination procedure reflected how much the students learned from the instructors they were rating—that is, the teaching assistants—the fact is that the teaching role of the teaching assistants appears to have been an ancillary one. Thus, 60 percent of the course was spent in lectures given by a professor; one must assume most of the material to be learned was transmitted at these lectures. It was also the professor who constructed the problems and, one infers, the homework. The teaching assistants' role was to answer questions about the professor's lectures and homework, to administer the problems, and to go over them. Insofar as the students' ability to solve the problems tested teaching effectiveness, it presumably tested in large part the teaching effectiveness of the professor. In-

sofar as the ability of students in the various recitation sessions to solve the problems differed, such differences might be more reasonably taken as indicative of how "in tune" with the professor the various teaching assistants were; that is, it might be more a reflection of their teamwork abilities rather than of their teaching effectiveness. Details that could shed further light on this would be of interest. How, for instance, did the ratings of the teaching assistant and the professor by students in the various recitation sessions compare? Rodin and Rodin speculate, "Perhaps students do not wish so much to maximize the amount learned as to reach an equitable compromise between the effort involved in learning and the perceived importance of what is being learned" (*7*, p. 1166). One might equally well speculate, on the other hand, that those students who found the professor's teaching approach least effective tended to rate highest the teaching assistants who departed most from the professor's approach; such students could be expected not to do as well on an evaluative device (the problems) set up by the professor.

Some question also arises regarding the evaluative device used by the Rodins to determine the amount the students learned. From the details furnished by the Rodins, it appears that the students were presented with one or more problems a week. If they failed to solve these, they could attempt variants of them, presumably on subsequent occasions, without penalty for up to six times per problem. It would seem likely that such an evaluative device would differentiate between students able to solve the first problem presented to them and those who, although they may have been able to solve the fifth problem variant presented to them, failed to solve the first four. The use of an evaluative device of this nature in a study such as the Rodins' might lead to results reflecting unduly the correlation between the scores and ratings of the students for whom the learning process had been slow.

## Measures of Teaching Effectiveness

What conclusions can be drawn regarding the appropriateness and validity of class performance on examinations and of student ratings as measures of teaching effectiveness?

Examinations are sometimes thought of as objective measures of student learning, and thereby of teaching effectiveness, while student ratings tend to be regarded as subjective (*7*). This view is open to challenge. Thus, although there is no question that examinations, particularly multiple choice ones, can be scored objectively, there is significant debate (*19*) regarding what they measure (namely, recall or problem-solving); moreover, they frequently reflect the subjective view of those who devise them as to what it was that the students were taught (*20*). Accordingly, if no significant correlation is found between student ratings and class performance on examinations, then there would appear to be no a priori basis for singling out, as Rodin and Rodin (*7*) have done, one of the variables as reflecting teaching effectiveness more accurately. On the other hand, when a high correlation is found between these two variables, then they would appear to validate each other, each reflecting a related aspect of teaching effectiveness.

The high correlation between student ratings and class performance on national normative examinations suggests that such examinations themselves could be used as a measure of teaching effectiveness. The routine use of national normative examinations for this purpose, however, would have certain predictable consequences and should not be undertaken without cognizance of what these would be. For instance, because an obvious way to improve class performance would be to ensure that the material presented in class corresponded to the material tested for in the examination, there would be a tendency to lock curricula and the material presented by individual instructors to that customarily covered by such examinations (*21*). This, in turn, could tend to make instructors more concerned with the content of such normative national examinations and more determined to have a voice in deciding what that content should be. Whether or not the overall effect would be a desirable one remains a question.

The high correlation found in this study between student ratings and class performance on a national normative examination would appear also to validate student ratings as a measure of teaching effectiveness. The work of Rous *et al.* (*9*) suggests that giving faculty members access to student ratings of themselves can improve the ratings they receive on subsequent occasions. If such ratings measure teaching effectiveness, a contention supported by the results of the present study, then there appear to be some grounds for the argument that the use of student ratings can lead to enhancement of the faculty's teaching effectiveness.

## Conclusions

It appears quite clear that student ratings of instruction and class performance on national normative examinations are positively related: the higher the student ratings of the instruction they receive, the higher the class score relative to a nationwide norm. On the other hand, no significant correlation is found between student ratings and class performance on institutional examinations. This suggests that both student ratings and class performance on national normative examinations are valid measures of teaching effectiveness.

### References and Notes


1. *Education at Berkeley—Report of the Select Committee on Education* (Academic Senate, Univ. of California, Berkeley, March 1966); E. J. McGrath, *J. Higher Educ.* **33**, 148 (1962).
2. J. R. Hayes, *Science* **172**, 227 (1971).
3. R. H. Knapp, in *The American College*, N. Sanford, Ed. (Wiley, New York, 1962), pp. 290–311; T. Caplow and R. J. McGee, *The Academic Marketplace* (Basic Books, New York, 1958).
4. P. H. Abelson, *Science* **157**, 759 (1967).
5. W. V. Voeks, *J. Higher Educ.* **33**, 212 (1962).
6. D. S. Gromisch, J. C. Bamford, S. N. Rous, S. Sall, S. Rubin, *J. Med. Educ.* **47**, 281 (1972).
7. M. Rodin and B. Rodin, *Science* **177**, 1164 (1972).
8. S. N. Rous, J. C. Bamford, D. Gromisch, S. Rubin, S. Sall, *J. Surg. Res.* **11**, 311 (1971).
9. S. N. Rous, J. C. Bamford, D. Gromisch, H. Rich, S. Rubin, S. Sall, *ibid.* **13**, 262 (1972).
10. H. H. Remmers, *Sch. Soc.* **28**, 759 (1928); *J. Educ. Res.* **21**, 314 (1930).
11. Although assignment to "satisfactory" of a value exactly halfway between "good" and "unsatisfactory" (that is, halfway between $+1.0$ and $-1.0$) was arbitrary, it may be defensible on the basis that the resulting correlation coefficients are relatively insensitive to the exact value assigned to this attribute. Thus, assignment to "satisfactory" of values ranging from $+1.0$ to $-0.2$ resulted in a shift in the correlation coefficient between class performance and student ratings of content and organization of less than 2.7 percent.
12. P. Elias, in *Biophysical Science—A Study Program*, J. L. Oncley, Ed. (Wiley, New York, 1959), pp. 221–226.
13. That is, the total number of times that attribute was elicited as a response divided by the total of 1778 responses.
14. W. R. Garner, *Uncertainty and Structure as Psychological Concepts* (Wiley, New York, 1962).
15. For 7.6 percent of the questions, the knowledge tested for derived from more than one subject area. In these instances, the question was assigned on a fractional basis. For 8.3 percent of the questions, the subject matter was not covered in the course; these questions were excluded from further analysis, as were three subject areas in the course about which no questions were asked.
16. Had weighting factors proportional to the information content of the responses not been used, the correlation coefficients reported above

would have been .75, .68, .69, and .61, respectively. The $P$ value for the first three of the correlation cofficients would have been $P < .001$, and $P < .005$ for the fourth.

17. It might be argued that, in calculating the correlation coefficients, the subject areas should be weighted proportionally to the number of questions relating to that area in the national examination. Since the correlation being investigated is that between independent variables, the appropriateness of such weighting is questionable. Moreover, if applied, it would not alter the conclusions of this study, although it would lower somewhat the values of the various correlation coefficients reported (to .58, .50, .59, and .43, respectively). The $P$ values for these four correlation coefficients would be $P < .01$, $P < .05$, $P < .01$, and $P = .065$, respectively.
18. This contention is supported by the fact that calculation of the correlation coefficients between a measure of the difficulty of the various subject areas (namely the mean performance of the nationwide sample in these areas) and (i) class performance on the national examination, (ii) class performance or departmental examinations, (iii) student ratings of content and organization, and (iv) student ratings of presentation, gives values for $r$ of 0, .18, .14, and −.01, respectively.
19. C. McGuire, *J. Med. Educ.* **38**, 556 (1963); in *1963 Yearbook: National Council on Measurement in Education* (Lansing, Michigan, 1963), pp. 7–16.
20. B. Hoffman, *The Tyranny of Testing* (Collier, New York, 1964); R. Cox, *Nature* **237**, 489 (1972).
21. Although efforts might be made, as indeed they were in this study [see (*15*)], to exclude from computation questions on material not covered in the course, such efforts would likely not be foolproof.
22. The author is grateful to D. S. Riggs for the classification of the National Medical Board Examination questions by subject area and to him and N. Solkoff, T. Gessner, F. J. Bruce, and C. H. Ehret for helpful comments on the manuscript. He also wishes to thank R. Spangler and W. J. Walbesser for suggestions regarding the mathematical treatment of the data.

**NEWS AND COMMENT**

# Arms Control Agency: Fred Iklé, New Captain of a Disabled Ship

Congress established the U.S. Arms Control and Disarmament Agency (ACDA) in 1961 as a modest step toward redressing the balance in a government then still feverishly engaged in building up the nation's strategic weaponry. The director of ACDA was assigned, by law, to serve as the President's principal arms control adviser and to assume, under the Secretary of State's direction, "primary responsibility within the government" for arms control matters.

No miracles were expected of ACDA, and none were performed. But this small agency, with a staff of never more than 270 people and an annual budget of never higher than $10 million, has proved its value by playing a key role in bringing about agreements such as the Non-Proliferation Treaty of 1968 and the ABM treaty of 1972. Today, however, ACDA seems to be on the Nixon Administration's list of agencies marked, if not for extinction, for obscurity. A number of members of the Senate Foreign Relations Committee will give voice to their growing concern over this situation at the public hearings, now set for 9 May, on the confirmation of ACDA's newly designated director, Fred Charles Iklé.

Iklé's confirmation itself does not appear in any danger, for Iklé is a political scientist with a respectable if not luminescent record of scholarship into questions of modern weaponry and international negotiations. Indeed, the recent White House announcement of Iklé's nomination was received with relief by those who had feared that the nominee might be someone wholly unqualified, such as one particular Republican senator from the West who was defeated for reelection last November. Iklé, once a professor at Massachusetts Institute of Technology, comes to ACDA from the Rand Corporation where he has been head of the social science department. If some professionals in the field of arms control are uneasy at certain of Iklé's ideas, they at least recognize him as one of their own kind and as someone with whom they can communicate.

The real significance of the upcoming hearing is that it will give senators their first opportunity to question an administration spokesman closely, and publicly, as to ACDA's future. If Iklé is unable to provide satisfactory answers, the committee or its arms control subcommittee can proceed from there, scheduling other Administration officials to testify about ACDA and possibly considering legislation intended to enhance the agency's status and influence.

ACDA seems to be undergoing a transition from an agency entrusted with important arms control negotiations to one discharging a modest advisory role, yet apparently without being allowed to keep the tools necessary to perform even that latter role adequately. First, note how ACDA has been stripped of a major part of its role in negotiations.

A few months ago the agency was denied the leadership in SALT II negotiations when the White House named a career diplomat, Ambassador-at-Large U. Alexis Johnson, to head this second round of strategic arms talks with the Russians. In fact, Gerard C. Smith, who in January resigned as ACDA director, last May was cut out of the final negotiations for SALT I—which he had led for some 2 years—and was not invited to be present in Moscow with President Nixon and Henry A. Kissinger when the SALT agreements were signed.

ACDA will provide some staff support for SALT II, but whether this will be done largely through Johnson's negotiating team or through Kissinger's National Security Council staff is not yet clear. What is clear is that, in its new advisory and staff support role, ACDA's influence on policy will probably be weak by comparison with what it would have been if the agency were still actually leading negotiations. (The ACDA official currently assigned to the SALT negotiating team is Sidney N. Graybeal, the agency's deputy assistant director for science and technology.) ACDA remains in charge of U.S. participation in the multilateral arms control negotiations going on at the United Nations Conference of the Committee on Disarmament (CCD) in Geneva, but whether this will continue to be so if these negotiations should suddenly begin moving toward important agreements appears very much a question.

Just the fact that Iklé is an academician without practical diplomatic or high-level governmental experience itself suggests that he was named to head a think tank of sorts and not an agency with the "primary responsibility" for arms control. His qualifications are in marked contrast to those of his two predecessors. William C. Foster, director of ACDA from 1961 to 1969, served as director of the Economic Cooperation Administration and as deputy secretary of defense during the