will often take years to complete, that dependent measures must be chosen because they are meaningful, and that the purpose of research is to obtain knowledge, the case for psychotherapy will remain unproven.

REFERENCE

Smith, M. L., & Glass, G. V. Meta-analysis of psychotherapy outcome studies. *American Psychologist,* 1977, *32,* 752–760.

PHILIP S. GALLO, JR.
San Diego State University

An Exercise in Mega-Silliness

The article by Smith and Glass (September 1977) begins promisingly by referring to my "tendentious diatribes" (p. 752) on the outcome problem in psychotherapy, inviting the reader to study two papers of mine, one of which I have no recollection of writing, and which the reader will look in vain for in the journal appearing in the list of references—indeed, the year and volume number given in the reference do not agree! The authors go on to the "astute dismantling of the Eysenck myth" (p. 752) by Bergin, not mentioning that the Bergin myth has in turn been astutely dismantled by Rachman (1971). No discussion of the issue can be regarded as meaningful which accepts the quite erroneous and—indeed, in places—absurd arguments of Bergin and pays no attention to the serious criticisms brought forward by Rachman. Indeed, the latter is not even mentioned in the bibliography, although his book *The Effects of Psychotherapy* is a classic in the literature. Smith and Glass have a somewhat arbitrary method of reference selection that does not augur well for their major opus.

This major opus disregards completely some of the major findings of Rachman's book, for example, that there are large differences in recovery (spontaneous remission) between different types of patients; the analysis takes no account of this effect. The analysis accepts without a word of warning subjective reports of therapists as a major source of information on outcome, although it is well known that such assessments are extremely unreliable even when made by well-qualified psychiatrists who have no personal involvement in the patient's recovery (Block, Bond, Qualls, Yalom, & Zimmerman, 1977); the distortion which is likely to arise when a therapist assesses his own patients' progress can well be imagined. It is noted that "subjectivity of the outcome measure" has much the highest correlation with effect size; this alone would invalidate all the complex statistics offered. Smith and Glass also do not mention the problem of selection, so well discussed by Rachman; patients for psychoanalysis are much more highly selected (for high intelligence, emotional resources, ego strength, etc.) than are patients for behavior therapy, and hence much more likely to improve spontaneously.

The most surprising feature of Smith and Glass's (1977) exercise in mega-silliness is their advocacy of low standards of judgment. More, they advocate and practice the abandonment of critical judgments of any kind. A mass of reports—good, bad, and indifferent—are fed into the computer in the hope that people will cease caring about the quality of the material on which the conclusions are based. If their abandonment of scholarship were to be taken seriously, a daunting but improbable likelihood, it would mark the beginning of a passage into the dark age of scientific psychology.

The notion that one can distill scientific knowledge from a compilation of studies mostly of poor design, relying on subjective, unvalidated, and certainly unreliable clinical judgments, and dissimilar with respect to nearly all the vital parameters, dies hard. This article, it is to be hoped, is the final death rattle of such hopes. "Garbage in—garbage out" is a well-known axiom of computer specialists; it applies here with equal force. There is only one sentence in the article with which one can wholeheartedly agree: "Extracting knowledge from accumulated studies is a complex and important methodological problem which deserves further attention" (p. 760). Until it has received such further attention, it would be highly dangerous to take seriously the "results" reported by Smith and Glass. Only better-designed experiments than those in the literature can bring us a better understanding on the points raised; in particular, placebo groups must be included in all designs which aim to study therapy-specific effects, and several therapists must be included for each method in order to obtain evidence on the therapist variance. I would suggest that there is no single study in existence which does not show serious weaknesses, and until these are overcome I must regretfully restate my conclusion of 1952, namely that there still is no acceptable evidence for the efficacy of psychotherapy.

REFERENCES

Block, S., Bond, G., Qualls, B., Yalom, I., & Zimmerman, E. Outcome in psychotherapy evaluated by independent judges. *British Journal of Psychiatry,* 1977, *131,* 410–414.

Rachman, S. *The effects of psychotherapy.* London: Pergamon Press, 1971.

Smith, M. L., & Glass, G. V. Meta-analysis of psychotherapy outcome studies. *American Psychologist,* 1977, *32,* 752–760.

H. J. EYSENCK
Institute of Psychiatry
London, England

Reply to Eysenck

We have numbered our responses to correspond to the successive paragraphs in Eysenck's (this issue) rejoinder:

1. Of our two citations of Eysenck's work, the 1952 reference is correct. The 1965 reference should have been to the *International Journal of Psychiatry,* not the *Journal of Psychology.* We read Rachman (1971) when it appeared and again when we began the project on which we recently reported. On the sub-

stantive issues involved, we continue to stand with Bergin (1971).

2. "Spontaneous remission" was not ignored. We chose to consider only studies that used control groups. The purpose of using a control group is to render consideration of "spontaneous remission" immaterial.

Our concern for the "subjectivity" of outcome measurement was exercised in the most impartial and sensible way we could imagine. "Subjectivity"—we called it "reactivity" —was quantified and included in regression analyses of outcomes. The effects of "reactivity" can be assessed and controlled by use of the regression equations in Table 5 in Smith and Glass (1977). Likewise, other aspects of "the problem of selection" (e.g., intelligence and diagnosis of clients) can be assessed and corrected statistically by reference to the regression equations in our Table 5. Again, the putative more likely "spontaneous remission" of "patients for psychoanalysis" is irrelevant; we considered only studies that used control groups. (If, incidentally, Eysenck's speculation were true, it would work *against* psychodynamic therapies' showing large effects; aspirin would show no superiority to a placebo among patients, all of whom recovered from headaches for extraneous reasons.)

Not only did we control ex post facto with regression analysis against the points Eysenck nonetheless raised, but the results in our Figure 4 are based only on those studies in which Eysenck's concerns are controlled experimentally. We regard these data as particularly persuasive.

3. Eysenck's points in his third paragraph are largely ad hominem and rhetorical; and one might not readily suspect that an important methodological issue lies buried here; but it does. One of us has addressed the point recently (Glass, in-press-a, in-press-b). We will attempt to put forward the issue briefly, aware that there is too little space here to elaborate upon it satisfactorily. Designing future experiments and appraising the results of completed experiments are quite different endeavors. There are many prudent safeguards that should be incorporated into the next experiment to be performed; but once completed, it is an empirical question whether the presence of that safeguard is systematically related to the findings of many experiments. In our analysis of psychotherapy outcome studies, such features as use of randomization versus matching and double versus single versus no-blinding had virtually no correlation with study findings. The mass of "good, bad, and indifferent" reports show almost exactly the same results. Connoisseurs' distinctions about which studies are "best" and which ought to be discarded would lead, in this instance, to a profligate dismissal of hundreds of findings. We and our students have studied the covariance of many properties of designs and analyses with experimental findings (Glass, Smith, & Miller, in press; Hartley, 1977; Hearold, 1978; Miller, 1978).

It is one thing to say that questions of "good design" are insignificant in evaluating the impact of this particular collection of 400 experiments, and another matter to say that the next experiment I propose to conduct needn't have a "good design." The former may very well be true, the latter virtually never. The two circumstances are quite analogous to two uses of a test and the related issue of test reliability. There is no contradiction in maintaining that a test must have high reliability if one wishes to make correct decisions about individuals, but far lower reliability is tolerable if only certain group characteristics are of interest (since measurement errors will average out in groups). One wants one's next study to be well designed because one wants to believe its findings and plan subsequent studies on the basis of them. But once the study is completed and its findings join those of dozens of other studies on the same topic, the sophistication and validity of the study design become an entirely different matter, and conceivably a minor one.

4. We disagree with Eysenck's opinion of the psychotherapy outcome literature, namely, that "there is no single study in existence which does not show serious weaknesses." The effects of psychotherapy could hardly have been so treacherous and elusive as to have tricked hundreds of investigators over the past 40 years (disguised as spontaneous remission in Jones's study, masquerading as the placebo effect in Brown's study, hiding cleverly behind a bad statistical analysis in White's study, etc.). As arduous as were the reading and reanalysis of the nearly 400 studies we included, we were often pleased to find excellent experiments, well executed and well reported. Probably, the best of them was the comparison of psychodynamic psychotherapy and behavioral therapy reported recently by Sloane, Staples, Cristol, Yorkston, and Whipple (1975). It satisfied many more criteria than Eysenck specified in his fourth paragraph. We were gratified to see that Sloane et al.'s results nearly exactly matched our compilation of effects across the 400 studies: essentially no difference between behavioral and nonbehavioral therapies and an average effect of about two thirds of a standard deviation above control groups for each.

REFERENCES

Bergin, A. E. The evaluation of therapeutic outcomes. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change.* New York: Wiley, 1971.

Glass, G. V. Integrating findings: The meta-analysis of research. *Review of Research in Education,* in press. (a)

Glass, G. V. Reply to Mansfield and Busse. *Educational Researcher,* in press. (b)

Glass, G. V, Smith, M. L., & Miller, T. I. *The benefits of psychotherapy.* Baltimore, Md.: Johns Hopkins University Press, in press.

Hartley, S. S. *Meta-analysis of the effects of individually paced instruction in mathematics.* Unpublished doctoral dissertation, University of Colorado, 1977.

Hearold, S. L. *Meta-analysis of the effects of television on social behavior.* Unpublished doctoral dissertation, University of Colorado, 1978.

Miller, T. I. *The effects of drug therapy in psychological disorders: A meta-analysis.* Unpublished doctoral dissertation, University of Colorado, 1978.

Rachman, S. *The effects of psychotherapy.* London: Pergamon Press, 1971.

Sloane, R. B., Staples, R. F., Cristol, A. H., Yorkston, N. J., & Whipple, K. *Psychotherapy versus behavior therapy.* Cambridge, Mass.: Harvard University Press, 1975.

Smith, M. L., & Glass, G. V. Meta-analysis of psychotherapy outcome studies. *American Psychologist,* 1977, *32,* 752–760.

GENE V GLASS
MARY LEE SMITH
Laboratory of Educational Research
University of Colorado

A Guide for Reviewers: Editorial Hardball in the '70s

The policies and practices of journal reviewers have undergone a marked shift recently. In the past, reviewers responded primarily to the judged importance, theory relevance, and originality of the idea under study. Such judgments are inherently subjective, however. In recognition of the scientific nature of the psychological enterprise, the new trend is toward simply listing the number of methodological flaws and conceptual alternatives that can be imputed to a given study. A recommendation as to whether to accept the article is then based simply on the sum of the flaws and alternatives. Some reviewers are not fully aware of this shift in practices, and the following rules are offered for those who wish to understand this new trend and join ranks with the forefront.

1. The best way to handle the question of the importance of the idea under investigation is to operate under the assumption that there is a perfect inverse correlation between its importance and the number of methodological flaws and alternative interpretations of the study.

2. The reviewer may similarly finesse the question of novelty of the idea under investigation. Be vigilant, however, for novel methods of investigation. Novel methods are suspect because untried.

3. Add points for "real world" investigations because these are still in vogue. Subtract points for ecological validity in the laboratory setting, however, because this is not so much in vogue and, other things being equal, more ecological validity means messier stimulus materials. Messy stimulus materials mean many artifactual explanations.

4. Bear in mind that *every* study has artifactual possibilities and that your prowess as a reviewer consists in adducing as many of them as possible. Keep the list below handy, since most of the examples can be applied to any study.

a. *Experimenter demand characteristics.* The nice thing about this one is that you don't have to think about exactly *how* demand characteristics could have produced the results. Simply assert it. Editors know that all studies are prone to them and God knows the authors realize they are guilty. They know they didn't take all those Rosenthal precautions.

b. *Social desirability.* This is another one that doesn't actually have to be thought through. The lovely thing is that obviously the subjects wouldn't have said or done what they said or did if they didn't think it was desirable. There's really no way out of this one.

c. *Motivational relevance.* If the study engages the subjects' interest and deals with matters that are important to them, then assert that the findings were obtained only because of the motivations or defenses that were aroused by the procedures.

d. *Limitation to trivial situations.* If the study does *not* engage the subjects' interest, then so much the better. It may be claimed that the phenomenon under study would not hold up under any but the barren laboratory situations studied.

e. *Subjects' comprehension of instructions.* If the instructions to subjects were long and complicated, assert that the subjects probably didn't understand them. The same criticism may be applied if the instructions were brief.

f. *Confounded manipulations.* Except for the very smallest and most circumscribed of manipulations, almost all independent variable manipulations involve changing more than one stimulus component. This provides a rich source of artifactual explanations. It is important to note that the artifactual explanations need not be at all plausible. The important thing is that they exist, and the author should be forced to rule them out methodologically. This may often mean that the manipulation will become so small and inconsequential that the infant phenomenon is thrown out with the alternative bathwater. But that's the author's problem, not yours.

g. *Verbal versus behavioral measures.* Criticize any article employing only verbal measures on the grounds that verbal measures are only verbal measures. Criticize any article having only behavioral measures on the grounds that the behavior might have been produced by any of a number of a cognitive states that were not tapped verbally. Criticize any article in which verbal measures follow the behavioral measures on the grounds that the verbal results were obtained only because the subject was trying to be consistent with his or her previous behavior. Criticize any article in which verbal measures precede the behavioral measures on the grounds that the behavioral measures were obtained only because subjects were trying to be consistent with their verbal reports. If the order of verbal and behavioral measures is counterbalanced, employ both of the latter criticisms.

5. Occasionally it may be difficult to find trivializing explanations, or the author may have spoiled things by arguing cogently against them. The enterprising reviewer is not daunted by this state of affairs. It is nearly always possible to find some interesting, nontrivial, alternative explanations for the results. In