Accuracy and Expert Testimony

Steven K. Hoge and Thomas Grisso

In an article in the journal Science, Faust and Ziskin have asserted that the testimony of psychologists and psychiatrists cannot “meet legal standards for expertise” for purposes of trial evidence.¹ This assertion, given visibility and credence by a journal of the stature of Science, understandably has raised concern among forensic mental health professionals.² Attorneys have begun to use Faust and Ziskin’s arguments to attack mental health professionals’ testimony. Moreover, many attorneys and judges who must evaluate these assertions may not have an adequate frame of reference for evaluating Faust and Ziskin’s arguments.

These concerns are augmented by the nature and quality of Faust and Ziskin’s accusations which, by our analysis, are largely unwarranted and misleading. The mental health professions should not seek to minimize the existence of limitations and valid complaints that may be made against psychological and psychiatric evidence. Faust and Ziskin’s article, however, makes no contribution to our understanding of those issues. Moreover, the charges leveled by Faust and Ziskin are serious ones, and the status of psychiatrists and psychologists may suffer if they are left unanswered. In this article, we critique Faust and Ziskin’s explication of the law governing expert testimony and their review of the clinical literature. We reach very different conclusions regarding the value of mental health professionals’ testimony.

Legal Standards for the Admissibility of Expert Testimony

The central premise of Faust and Ziskin’s arguments can be found in their interpretation of Frye v. United States.³ This landmark case provided a standard by which courts could judge the admissibility of novel scientific evidence. However, the Frye test is not the sole test used by courts, and a broader consideration of evidentiary rules is necessary for a full understanding of the proper role of expert testimony.

In order to be properly introduced—regardless of the evidentiary framework applied—evidence must meet a test of relevance. The Federal Rules of Evidence,⁴ which embody the common law, succinctly define relevant evidence as:

evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence. [Rule 401]⁴

Relevance, then, has two components: the evidence must address a fact that is the subject of legal dispute, and it must be of probative value with regard to that

¹ Dr. Hoge is affiliated with the University of Virginia, Schools of Law and Medicine, Box 100, Blue Ridge Hospital, Charlottesville, VA 22901. Dr. Grisso is affiliated with the University of Massachusetts Medical Center.
fact—it must make an incremental contribution to the fact finder’s determination of the truth or falsehood of the fact. Testimony judged to be not relevant will be excluded. Under some circumstances, relevant testimony may be excluded if the likelihood that it will be prejudicial outweighs its probative value, or if it is repetitive, or confusing, or misleading to the jury; if the premises on which the evidence rests cannot be tested, it may also be excluded. [Fed. R. Evid. 403 and 801]4

Regarding expert testimony, the Federal Rules of Evidence state:

If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise. [Fed. R. Evid. 702]4

Under these rules, which have been adopted in many jurisdictions, expert testimony is subjected to the same relevancy test that other evidence must meet. Challenges regarding the certainty, reliability, or precision of opinion testimony would be left for the consideration of the fact finder and influence the weight given to the expert’s opinion, not the admissibility of that opinion.

In many jurisdictions, an additional test—the Frye test—is applied to novel scientific evidence. Faust and Ziskin set the stage for their condemnation of mental health professionals’ testimony by explaining that the Frye test requires “two essentials.” First, the professional’s opinion can be admitted, they explain, only if the professional can state the opinion with “reasonable medical certainty.” They interpret this to mean “pretty likely accurate.” Second, they explain that Frye requires that the professional’s expert testimony must help the court “reach a more valid conclusion.” Then they pose two questions to guide their review: “(1) Can expert witnesses in psychology and psychiatry answer forensic questions with reasonable accuracy? (2) Can experts help the judge and jury reach more accurate conclusions than would otherwise be possible?”

Faust and Ziskin's explanation of Frye, therefore, suggests that courts require and are interested primarily in a high level of scientific precision and accuracy in expert testimony. This is the test against which they proceed to condemn the use of psychologists and psychiatrists as expert witnesses in legal forums. Their interpretation of Frye, however, is misleading.

Defendant Frye, who was being tried for murder, was not allowed to introduce expert testimony about “the systolic blood pressure deception test,” which apparently supported his contention of innocence. He appealed, claiming that the evidence “did not lie within the range of common experience or common knowledge,” which had been the test of admissibility until that time. The court in Frye then produced the following test for deciding the admissibility of expert testimony:

Just when a scientific principle or discovery crosses the line between the experimental and demonstrable stages is difficult to define. Somewhere in the twilight zone the evidential force of the principle must be recognized, and while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the
Accuracy and Expert Testimony

thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs. (Emphasis added.)

We think that the systolic blood pressure deception test has not yet gained such standing and scientific recognition among physiological and psychological authorities. . . .

_Frye_, therefore, requires that the mental health professional’s testimony must not only be relevant, but it also must be buttressed by “general acceptance” within psychology or psychiatry (or subfields therein). The _Frye_ test is meant to be applied to novel scientific theories and methods; it is a conservative rule intended to insulate legal decision making from being influenced by unproven innovations. Courts may acknowledge general acceptance through the testimony of experts, scientific and legal literature, or judicial precedent. . . . Note, however, that _Frye_ does not require that the scientific field generally accept the expert’s conclusions, deductions, or inferences. That which must be accepted by the field is “the thing from which the deduction is made.” This “thing,” for example, might be a generally accepted theory of psychopathology, method of psychological measurement, or systematic procedure for assessing symptoms of mental illness. But _Frye_ does not mention accuracy, validity, or even “general acceptance” of the opinion or conclusion that the expert reaches on the basis of these theories and methods. It accepts individual, potentially idiosyncratic, conclusions by the expert who is applying generally accepted theory or method to an area of investigation or to an individual case.

A brief illustration might be helpful. Imagine that a clinical psychologist is asked to testify concerning whether a criminal defendant is mentally retarded. In addition to collecting information about adaptive functioning, the psychologist administers the Wechsler Adult Intelligence Scale-Revised (WAIS-R) in order to obtain an index of general intelligence. It is well known among psychologists, however, that an intelligence test score has a range of expectable error (the standard error of measurement), so that the “true score” of a person scoring somewhat below 70 may actually be above that cut-off score for classification. Moreover, an examinee might score beyond the range of error itself due to peculiar circumstances on that particular day in the examinee’s life. An examinee in some circumstances might obtain a low score by feigning ignorance for reasons of personal gain associated with a finding of mental retardation. Finally, intelligence test scores often do not accurately correlate with success in everyday functioning in various occupations or educational settings. In summary, intelligence test results are not always accurate and precise, and sometimes they are not valid predictors of everyday functioning.

Nevertheless, the body of research with the WAIS-R has clearly demonstrated its general reliability as well as its construct validity. It meets the standards of the field of science and professional practice to which it belongs, so much so that any clinical psychologist’s assessment for the question of mental retardation would be considered inadequate without such a measure. This is not because intelligence scores actually
Hoge and Grisso

define mental retardation or because they are highly predictive of adaptive functioning, neither of which can be claimed by the WAIS-R. It is simply because such measures play a vital role in forming opinions about mental retardation, when combined with data from various other sources.

This is what Frye requires. It does not require that clinicians, or their methods of assessment, must be correct. It requires that the various methods and procedures that have contributed to clinicians’ opinions must meet professional and scientific standards.

In contrast, Faust and Ziskin interpret Frye to ask the question, "Can expert witnesses in psychology and psychiatry answer forensic questions with reasonable accuracy?" They imply that what the courts need to support the admissibility of testimony is a high degree of certainty by the science or profession that the expert's opinion is correct.

What Frye actually requires, however, is for courts to look to the professions to define the conceptual grounding for their experts' opinions. The courts intercede only to limit testimony that falls outside those boundaries. Opinions are excluded when they are based on theories and methods that are so untested by their scientific and professional fields that they are not generally accepted by those fields. The "reasonable medical certainty" to which Faust and Ziskin refer has been consistently interpreted as referring to the degree of confidence the clinician places in his or her opinion, not to the opinion's accuracy in an empirical sense.

This way of dealing with issues of admissibility of expert testimony occasionally has been responsible for legal decisions that may seem anomalous from the scientific perspective. For example the American Psychiatric Association offered evidence indicating that mental health professionals' opinions about future dangerous behavior often will be inaccurate. Nevertheless, in Barefoot v. Estelle, the U.S. Supreme Court found that there were no constitutional barriers to the use of mental health experts' opinions regarding defendants' future dangerous behavior in capital sentencing hearings. Similarly, legal reform of civil commitment standards in the last two decades has led to standards that call for judgments about dangerousness, which require mental health professionals to make predictions of future violence in the absence of conclusive research.

Discrepancies between what we are able to offer accurately and what the law will accept as expert testimony are found in medical as well as mental health testimony. Appellate courts have upheld decisions based on causally impossible medical testimony—for example, that a cancer was caused by blunt trauma—citing the certainty of opinion expressed by the expert. Some courts have expressed the matter very directly:

In a courtroom, the test for allowing a plaintiff to recover in a tort suit of this type is not scientific certainty but legal sufficiency...[T]he fact...that science would require more evidence before conclusively considering the causation question resolved is irrelevant.

Does this mean that the law is not concerned about the validity of our
Accuracy and Expert Testimony

opinions? No, the law hopes for validity in expert testimony, just as it hopes for truthfulness in the testimony of other witnesses. It knows no way to judge validity, however, but by assuring that the procedures of justice allow expert testimony to be challenged at trial, and by excluding opinions based on procedures and methods that are not recognized by the profession and science of the expert. It is the collective responsibility of mental health professionals, not the law, to limit our testimony to those matters about which we have reasonably accurate information and can offer reasonably valid conclusions. Indeed, the most common problem with the general acceptance standard is not, as Faust and Ziskin believe, that mental health professionals exploit it to foist unreliable opinions on the courts. Rather it is that the judicial system seems to have an unquenchable thirst for the input of mental health professionals, often insisting that they go beyond the limit of general acceptance.

The courts' desire to hear the opinions of professionals is especially apparent in testimony on forensic assessments, when professionals are asked to answer the ultimate legal question (e.g., whether the defendant is criminally responsible.) Ultimate legal questions never correspond to clinical diagnoses or any specific psychological functions. The answers to these legal questions represent society's contextual appraisals in specific instances; these judgments are bounded by notions of morality, justice, and social policy. Many commentators in the field of forensic mental health agree that clinicians providing assistance to the legal system should stop short of answering the ultimate legal question. Alternatively, some mental health professionals feel that addressing ultimate legal questions can be helpful to the trier of fact when the basis of the opinion and reasoning process has been fully explained. Yet many courts insist on experts answering ultimate questions, even when they are not provided the opportunity to explain their reasoning fully. Such testimony obscures the special expertise of the mental health professional and the distinctions between clinical and moral dimensions.

If organized psychology or psychiatry believes that substantial error or ethical conflict is inherent in offering certain types of testimony, then it should urge its professionals to so limit their testimony (e.g., to offer testimony about the degree of risk of future violence, rather than predicting that it will or will not occur; to offer information related to ultimate legal questions, but not to offer opinion answers to those questions themselves). Moreover, the professions may encourage their colleagues to rebut in the courtroom the testimony of their colleagues who do not limit their opinions in the manner that the profession believes is warranted.

These arguments, however, are very different from those of Faust and Ziskin. They attack mental health experts by interpreting Frye in a way that holds their testimony to a different standard than the one imposed on other medical and nonmedical experts. That is, they claim that Frye demands accuracy of
opinions, whereas in fact it merely requires the field's general acceptance of a theory, concept, test, or method before an expert can rely on it for arriving at an opinion entered into evidence at trial.

Having begun with a misstatement of the legal standard to be applied, Faust and Ziskin's subsequent analysis of mental health experts' abilities to meet that standard could be dismissed on a logical basis. Nevertheless, it is instructive to consider the strength of some of their later arguments on their own merit. Moreover, we are mindful that some courts, applying standards other than Frye, have at times examined the reliability of scientific evidence as a question separate from general acceptance by the scientific community. Therefore, the reliability of diagnosis and forensic assessments deserves scrutiny in its own right.

Reliability of Diagnoses

Faust and Ziskin argue that mental health professionals' clinical diagnoses are too unreliable to be admitted as evidence at trial. In other words, they would claim that diagnoses are so undependable that the field of mental health should not even be allowed to set its own threshold of general acceptance.

In science, the classification of phenomena is essential for systematic study to occur; within the mental health field this classification occurs when diagnoses are employed. Diagnoses begin as descriptions of signs and symptoms of illness; over time, as experience and empirical studies grow, clinicians' ability to predict the course of a mental disorder also grows; prognostic capabilities develop. Once this level of expertise has evolved, interventions can be tested; the course of treated patients can be compared with untreated ones and effective treatment developed. Eventually enough may be learned to understand why the disorder occurs and develops (etiologic and pathogenesis). Adherence to this model accords the mental health professions the stature of other medical and scientific specialties.

Recent years have seen many attacks on the diagnosis of mental disorders. But mental health professionals' abilities to make accurate diagnoses of mental disorders have never been as unreliable as Faust and Ziskin assert. Even before the advent of modern diagnostic criteria, studies had demonstrated that diagnosticians were able to classify patients very reliably into broad diagnostic categories, such as psychoses, organic disorders, or personality disorders.\(^{23}\)

It was true, however, that clinicians' attempts to make more specific diagnoses (e.g., types of psychoses) were less reliable. This was due primarily to the vagueness of earlier, DSM-II diagnostic standards, rather than the intrinsically flawed observation and judgment of professionals, thus leading to the development and implementation of operational criteria in DSM-III.\(^{24-26}\) With the advent of modern descriptive diagnoses, high interrater reliability has been established for Axis I, DSM-III disorders.\(^{24}\) A lower degree of interclinician reliability has been found for DSM-III personality disorders. When clinicians accumulate sufficient medical and arrest records, as well as other longitudinal information.
reasonable levels of reliability can be obtained as a class. The reliability of diagnosis of specific personality disorders, however, is mediocre.

In contrast, Faust and Ziskin’s claims concerning the unreliability of diagnoses of mental disorders are based almost entirely on older, pre-DSM-III studies, according to their citations of studies supporting their argument. Moreover, their review makes heavy use of studies that have documented the less reliable nature of diagnosis of specific personality disorders. They neglect to cite the research, noted above, that has shown the much better reliability of diagnosis of Axis I mental disorders.

Mental health professionals’ diagnoses, of course, are not infallible. But studies demonstrate that they are no less reliable than physicians’ diagnoses of other medical problems. Moreover, there is evidence that professionals’ understanding of clinical syndromes commonly seen in forensic settings exceeds that of judges, and that forensically trained clinicians have a better understanding of clinical syndromes than do general mental health clinicians.

In summary, Faust and Ziskin unfairly portray the diagnostic accuracy of major mental illness as unreliable. Substantial evidence demonstrates that DSM-III can contribute to high reliability of Axis I diagnoses. Mental health professionals have a voluminous literature at their disposal regarding the assessment, treatment, and prognosis of these disorders, most of which is unfamiliar to judges and juries who must make decisions, such as criminal responsibility and legal competencies, for which diagnostic information is of substantial relevance.

Reliability, Validity, and Utility of Forensic Assessments

Pressing on to further accusations, Faust and Ziskin claim that mental health assessments are not valid for forensic purposes because they lack relevance for the legal questions that are asked. Addressing this irrelevance, they claim that clinical conditions (such as psychosis) do not correspond to legal definitions of concepts such as incompetency to stand trial or lack of criminal responsibility. This lack of correspondence, they claim, renders mental health experts’ testimony invalid and useless.

It is difficult to make sense of this assertion. It is true, of course, that clinical conditions of mental illness are not synonymous with legal constructs of incompetency or criminal responsibility, as virtually all authors of leading textbooks in forensic assessment have pointed out. Many of these same authors, however, find substantial legal and scientific support for the argument that the relevance of mental health testimony is not dependent upon its ability to define incompetency or criminal responsibility. Clinical information about mental functioning and illness are at least relevant, and in many instances essential, to judicial decision making in these, and other, areas of forensic mental health. Mental health experts’ clinical opinions do not have to answer the legal questions of competence or criminal responsibility in order
to provide valid assistance to courts. Indeed, as we have pointed out, many of the authors we cite above contend that mental health professionals should not answer the ultimate legal questions. Their proper role is to describe the relevant abilities, disabilities, symptoms, and diagnostic conditions in clinical and behavioral terms, leaving to the court to weigh these observations in the context of legal concepts and standards.

Even if one were to accept that mental health clinicians should render opinions about ultimate legal questions, there is ample evidence to refute Faust and Ziskin's complaints. Numerous empirical studies have shown a high degree of interclinician reliability, or a correspondence between clinicians' and judges' conclusions, regarding defendants' competence to stand trial and criminal responsibility. The correlations found in these studies indicate a remarkable degree of "accuracy" when one considers that the ultimate legal determinations reflect moral and policy dimensions in addition to clinical ones. Inexplicably, Faust and Ziskin made no mention of this body of research in their article.

**Conclusion**

Faust and Ziskin's article in *Science* has provided the general scientific community, as well as the courts, a misleading and potentially damaging view of the value of mental health professionals' assistance to legal decision making. The scholarly value of the article is easily dismissed on three general grounds.

First, as we have shown, their arguments tend to begin with premises that are either misinterpretations of law or misstatements of the purposes of mental health testimony. In either case, their premises set up inappropriate standards against which to judge the quality and utility of mental health assessments in legal proceedings.

Second, as we have described in this review, Faust and Ziskin cite literature very selectively to support their arguments; their citations often reach into the distant past for studies that criticize methods no longer in use. In contrast, they fail even to mention more recent studies that would make their arguments more difficult to support.

Third, careful scholarship requires that one demonstrate a knowledge of views opposing one's own and an ability to rebut those views. In contrast, Faust and Ziskin's article cites or discusses few works from the past decade (other than their own), despite the fact that most of the major textbooks that currently exist on forensic mental health assessment—as well as most of the empirical research—have been published since 1980.

Other writers have raised more important and cogent criticisms of forensic mental health evaluations than those offered by Faust and Ziskin (e.g., ref. 18). One hopes that this will continue. Forensic psychiatry and psychology should be hard on themselves, demanding that mental health professionals and courts alike be fully aware of any limitations and inadequacies that exist in clinical testimony. But we can find no reason
Accuracy and Expert Testimony

why the fields of forensic psychiatry and psychology, or for that matter science as a whole, should afford the Faust and Ziskin article the "general acceptance" that Frye requires for works by which our assistance to courts should properly be judged.

Acknowledgment

The authors would like to acknowledge the help of Paul Appelbaum, who provided comments on an earlier version of this paper.

References

3. Frye v. United States, 293 F. 1013 (D.C. Cir. 1923)
14. Peterson v. Kansas City Public Service Co., 259 S.W.2d 789 (Mo. 1953)
community hospital. JAMA 250:1177-81, 1983
37. Laboratory of Community Psychiatry, Harvard Medical School: Competency to Stand Trial and Mental Illness. Rockville, MD, National Institute of Mental Health, 1973
40. Roesch R, Golding S: Competency to Stand Trial. Urbana-Champaign, IL, University of Illinois Press, 1980