Expert professionalism as judged by experts: An empirical pilot study

BY THOMAS G. GUTHEIL, M.D., PATRICE MARIE MILLER ED.D., AND MICHAEL LAMPORT COMMONS, PH.D.

We conducted a pilot study about the professionalism of opposing expert witnesses as perceived by subject experts. The investigated topics included opponents' disorganization; forgetfulness; failure to examine a relevant party or review relevant documents; ignorance of the legal standard; giving an opinion or espousing a theory beyond the case facts or the relevant science; and replacement of objectivity with advocacy. We also looked at perceptions of opposing experts as "hired guns" and at treating professionals serving as experts. Significant findings were: opposing experts' lack of professionalism was perceived by half of the sample; lack of professionalism was attributed to espousal of idiosyncratic theories and loss of objectivity; participants professed a high degree of certainty about these views. "Hired gun" status in opponents and treating professionals as opposing experts were viewed as rare events.

KEY WORDS: *Expert professionalism, expert witnesses, forensic psychiatry, forensic psychology.*

CROSS-EXAMINING ATTORNEY (*contemptuously*): Isn't it true, Doctor, that you are a professional witness?

EXPERT: I am always a professional, today I am a witness, but I am not a professional witness; I am a teacher.

—Trial excerpt

In the adversarial model of the American legal system, expert witnesses are exposed to the behavior of opposing experts in a given case through reading their reports and/or depositions, and reading, directly observing, or hearing about trial testimony. Such exposure may permit formation of opinions about the professionalism of that opposing expert.

The present study is a part of a series of largely unprecedented empirical pilot studies (Commons, Miller, & Gutheil, 2004; Gutheil, 2000b; Gutheil, 2001; Gutheil, Commons, & Miller, 2001; Gutheil & Sutherland, 1999; Strasburger, Gutheil, & Brodsky, 1997), in which the Program in Psychiatry and the Law, Beth Israel Deaconess Department of Psychiatry, Harvard Medical School, has surveyed different aspects of expert practice, particularly in the area of expert-attorney relations. Taken as a group, these studies appear to address indirectly various aspects of professionalism of both attorneys and experts.

We assume expert witnesses strive for professionalism in their work. To study perceptions of professionalism more directly, we performed in this case an empirical pilot study of expert witnesses' perceptions of the professionalism of opposing experts.

The Rasch model

In this pilot study, we used a Rasch analysis to show the perceived degree of professionalism in an objective, empirical manner. In order to understand our results, a basic knowledge of Rasch scaling is necessary. Rasch analysis

was originally developed for large-scale achievement testing (Rasch, 1980). Its use has since exploded in a variety of disciplines and for a wide range of topics including issues in psychiatry and the law (see Dattilio, Commons, Adams, Gutheil, & Sadoff, 2006, for a more extensive introduction to the topic). A Rasch analysis, through the use of probabilistic equations, converts raw ratings of items into scales of Rasch scores that have equal intervals. Such a scale can then be used as a type of objective ruler against which to measure the data on survey items as well as on respondents (Andrich, 1988). Statistically speaking, this scale will be linear (Wright & Stone, 1979). As a result, a change of 1 unit represents the same amount of change going from -2 to -1 as going from 0 to +1 (or any other difference of the same size). After analyzing data with a Rasch model, a number of questions can be answered. In particular, where on the scale of perceived professionalism does each item fall? Second, what is the range of scaled values for all items and for all participants?

Method

Participants   After clearance through our human studies committee and approval from the Research Committee of the American Academy of Psychiatry and Law (AAPL), we distributed an Expert Professionalism questionnaire developed by the authors. Some participants in the study were voluntary attendees at one of the Twilight Zone workshops (Gutheil, 2001) held at the annual meeting of AAPL. This workshop was advertised as both research and as an opportunity to discuss in a workshop setting attorney-expert matters that were not often openly addressed, i.e., that existed in an insufficiently assessed Twilight Zone. Some attendees had also attended previous such workshops, and the basic theory and early results had been presented as the 2000 AAPL Presidential Address (Gutheil, 2000a). Thus, some familiarity with the format might be expected from at least

some attendees. The smaller group of participants came from among participants at the Program in Psychiatry and the Law who had not previously seen the questionnaire.

The participants' demographic information is provided in Table 1. This seems to be a fairly typical subsample of those that might attend an AAPL meeting. The majority of participants were physicians (80%), were forensically board certified (53%), were American Psychiatric Association (APA) members (73%), and were AAPL members (66%). This was a forensically experienced cohort fielding an average of 48.82 cases per year.

TABLE 1

Demographic characteristics of those responding to the survey

Total number of participants	45
Males	28 (62%)
Females	17 (38%)
M.D.	36 (80%)
APA members	33 (73%)
AAPL members	30 (66%)
Board certified in forensics	24 (53%)
Mean # of years in forensic practice	11.34 (SD = 9.32)
Mean # of annual cases	48.82 (SD = 79.07)
Median # of annual cases	30

Instrument

Questions for the Expert Professionalism questionnaire were informed by the experiences of members of the Program in Psychiatry and the Law. On this questionnaire, participants were first asked whether an opposing expert had ever failed to perform in a professional manner. If yes, participants were asked to attribute the professional failure to a series of 12 possible reasons, which appeared in the questions that followed. The instructions immediately following the first question asked: "Regardless of how often you observed that failure in those cases, to which of the following do you

attribute the professional failures? Please circle the number closest to your answer." Following that instruction, the first of the 12 questions about reasons or situations appeared, for example: "The opposing expert was disorganized." The participant was to rate how often they had encountered such a case, on a scale from 1 = Never to 6 = Always. The questions used are shown in Table 2.

After responding to the 12 specific queries, participants were asked to choose a sample case in which one or more of the suggested mechanisms was present and to rate their degree of certainty about their ratings. They were then asked if they had done anything about this lack of professionalism. Finally, participants were asked whether they had ever felt an opposing expert was a "hired gun" (supplying an opinion only for money); and whether they had been in cases where the opposing expert was also the treater, after which they were asked how justified was this dual role. Space for explanations or comments was provided. These are shown in Table 2 as well.

Results

Of the 38 participants who responded to the first question of the questionnaire, reporting whether or not they had experienced an expert witness performing unprofessionally, 63.15% (24/38) reported that they had witnessed these kinds of behaviors. When tested against a null hypothesis that experts would not engage in unprofessional behavior at all, this number is very significantly above that rate ($z = 8.0$, $p < .00005$, $d = 2.63$). The probability of obtaining a number deviating this far from 0 is very small and the effect size d is extremely large. Given that this was not at all an infrequent event, but was observed by participants in this sample to have occurred more than half the time, it seemed worthwhile to explore possible reasons that experts attributed to others when there was a lack of professionalism.

TABLE 2
Professionalism questions and participant ratings

Variable Name	Question	Mean	SD	t	p-value*	Rasch Measure	MNSQ	Effect Size
Followup	Whether respondent has ever decided to take action after concluding expert witness on the opposite side has acted unprofessionally.	1.19	.402	15.13	.0005*	2.71	.94	2.96
Treater	Whether respondent has ever been involved in a case where expert witness on opposite side from respondent's has also been the examinee's treater	1.56	.504	18.03	.0005*	1.26	.52	3.09
Profess	Whether expert witness on the opposite side from respondent's has ever failed to perform professionally.	1.63	.489	20.57	.0005*	.94	.33	3.33
Hired-gun	Witness on opposite side from respondent's in a case is a "hired gun"	1.76	.431	23.90	. 0005*	.90	.39	4.08
Forgot	Failing performance by opposing expert due to the latter's forgetfulness.	2.28	1.137	-5.36	.0005*	.55	.71	1.073
Justify	Respondent's assessment of rectitude of opposing expert witness also being the examinee's treater	2.42	1.734	-2.65	n.s.	.37	1.60	0.623
Chaos	Failing performance by opposing expert due to latter's disorganization	2.68	1.492	-2.74	n.s.	.25	.64	1.253
Noexam	Failing performance by opposing expert due to latter's lack of examining a relevant party	2.84	1.748	-1.888	n.s.	.15	.94	0.583

Trialtim	Failing performance by opposing expert due to the latter's failure to read relevant material by trial time	3.44	1.873	-0.160	n.s.	-.18	.79	0.032
Ignorance	Failing performance by opposing expert due to the latter's ignorance of relevant legal standard	3.52	1.661	.060	n.s.	-.25	1.09	.0120
Deptime	Failing performance by opposing expert due to latter's failure to read relevant material by deposition time	3.56	1.878	0.157	n.s.	-.26	.78	0.032
Lacking	Failing performance by opposing expert due to some failure on part of latter not covered by the preceding categories	3.68	1.729	.493	n.s.	-.30	1.62	0.104
Document	Failure of opposing expert due to the latter's failure to review relevant documents/records	3.92	1.765	1.223	n.s.	-.46	.85	0.237
Extra	Failure of opposing expert due to the latter's going beyond the facts of the given case	4.04	1.483	1.852	n.s.	-.57	.94	0.121
Science	Failure of opposing expert due to latter's going beyond the science established for the field	4.19	1.442	2.469	n.s.	-.65	.63	0.478
Wrong	Failure of opposing expert due to latter's espousal of theories not supportable by available facts in case	4.70	1.235	5.066	.0005*	-1.02	.69	0.971
Certain	Respondent's degree of certitude at the time of given events of opposing expert witness lack of professionalism	5.31	1.011	9.120	.0005*	-1.71	2.08	1.790
Partisan	Failure of opposing expert due to expert's objectivity having been replaced by partisanship and advocacy for his/her own side of the given case	5.33	1.177	8.096	n.s.	-1.73	1.24	1.554

* The alpha level for the 18 tests calculated here, using a Bonferroni correction, is .05/18 or .0028.

We asked about 12 different situations, and asked 6 associated questions as well, and expected that some of the situations asked about would be perceived as creating more unprofessional behavior than others. To examine how the situations were compared, we present (in Table 2) both the mean ratings of each questionnaire item, and the scores from a Rasch analysis. The questionnaire items (named and given in detail in the two left hand columns) are ordered in terms of their scaled value in the Rasch analysis. Each type of analysis will be discussed in turn.

The mean ratings ranged from 1.19 (just above the "never" point) to 5.33 (just below the "always" point). Except for the first four ratings, they were all tested against the neutral point of 3.5. The first four ratings, because of their content (asking whether something ever happened) were tested against a value of 0. To a large extent the mean ratings increase as one moves down the table. The items listed further down in the table, such as partisanship (which is espousing theories unsupported by either the facts of the case or by science) are seen as more likely reasons to judge another expert's performance as unprofessional, whereas such factors as forgetfulness or disorganization are less likely reasons to judge an expert's performance in that manner. Only the last two items (using theories not supported by the facts of the case and being partisan) are noted to lead to unprofessional behavior significantly more frequently than factors that are rated at around the midpoint; this can be seen by the fact that the means of these two factors are significantly above the midpoint. The majority of the factors cluster around the midpoint and are not significantly different from it. Effect sizes for each mean are also presented; in this case we use Cohen's d (Cohen, 1992), which is essentially a measure of how many standard deviations the mean is from the hypothesized expected value (which would be 0 for the first four questions, and 3.5 for the remaining questions). As can be seen, the first five

questions have large effect sizes, as do the last three. These are all also statistically significant.

The Rasch analysis answers a different set of questions. It takes all the items and puts them into a scale so that one can specifically assess how much more important a factor is in influencing professionalism than another factor. To carry out the Rasch analysis we put participants' answers to all of the questions pertaining to factors thought to influence professionalism into a data file for analysis. The Rasch analysis retained 39 respondents, those who had completed enough of the items to be included. For a Rasch analysis, this is a small sample. A sample size of 30 assures 95% confidence, but even so, a sample size of 100 is recommended. The results were also analyzed to see if any outlier items needed to be dropped, but there were none. All the items but one clearly fit on a single "perceived-seriousness-of-non-professionalism" scale. The usual criterion for dropping an item is when it has an infit error greater than 2.00. One item in this analysis had an infit mean square of 2.08. That was the item that asked participants how certain they were about the opposing expert's lack of professional behavior. Because this question is not about a specific factor that might be involved in unprofessional behavior it makes sense that it would fit the overall scale less well, however it was so close to the criterion, it was left in.

The Rasch procedure calculates a Rasch score or measure for each of the questionnaire items. The values of this measure, as seen in Table 2, range from 2.71 (at the top of the table) to -1.73 (at the bottom of the table). Note that more negative scores, which are toward the bottom of the table, reflect a rating that these items were more likely to be rated as reflecting unprofessional behavior. Therefore, the item that was judged to influence professional behavior the most was the one that stated that the expert's objectivity had been perceived to have been replaced by partisanship

(Rasch score of -1.73). As was also seen when looking at the mean ratings, other situations that led to the perception of unprofessional behavior included situations in which the expert espoused theories that were not supported by the facts of the case or were not supported by science. Other situations, such as being an examinee's treater, being known to have behaved unprofessionally, or being known to be a "hired gun" were perceived as not leading to unprofessional behavior as often.

As stated above, each interval on a Rasch is of equal size; as a result equal differences in scaled measurements between items about professionalism represent equal differences in the resulting perceptions of the importance of this factor. This allows for another dimension to the comparison of the scores that cannot be obtained from mean ratings, which are simply ordinal. Figure 1 allows for a graphic comparison of the distances of items from each other, if one examines the names of the factors shown on the right side of the line down the middle of the figure. For example, finding that someone had become partisan (at -1.73) is about one unit more unprofessional from finding that someone used bad science in their decisions (-1.73 - (-.63) = -1.10, or just slightly over one unit). For another example, lack of professionalism due to being an examinee's treater (Rasch measure 1.26) is about one unit away, toward the positive end of the scale (the less serious end), from lacking professionalism due to not examining a relevant witness (1.26 - .15 = 1.11). These two differences are the same size on the Rasch scaled measure.

Figure 1 also shows that the participants or raters (shown on the left side of the center line in the figure), tended to be clustered within one standard deviation of the mean value (with only four outliers, toward the negative end of the scale). This suggests that most participants were close together in the ratings they gave to the items.

FIGURE 1
Rasch analysis mapping

```
                SUBJECTS MAP OF RANKS
              <more>    |    <rare>
    3                   +
                        |
                        |       followup
                        |
                        |
    2                   +  T
                        |
                        |
                        |
                        |       treater
                        |
    1                   +  S   profess
                   T  |        hiredgun
            XX          |
             X          |       forgot
            XX      S  |        chaos          justify
           XXX         |        noexam
    0     XXXXX        +  M
            XX          |       ignoranc       trialtim
            XX      M  |        deptime        lacking
           XXXX        |        document       extra
           XXX         |        science
            XX      S  |
   -1        X          +  S   wrong
            XX          |
                        |
                   T  |
                        |       certain        partisan
             X          |
   -2                   +  T
                        |
                        |
                        |
                        |
                        |
   -3                   +
                        |
                        |
                        |
                        |
                        |
   -4   XXXXXXXXX       +
              <less>  |  <frequ>
```
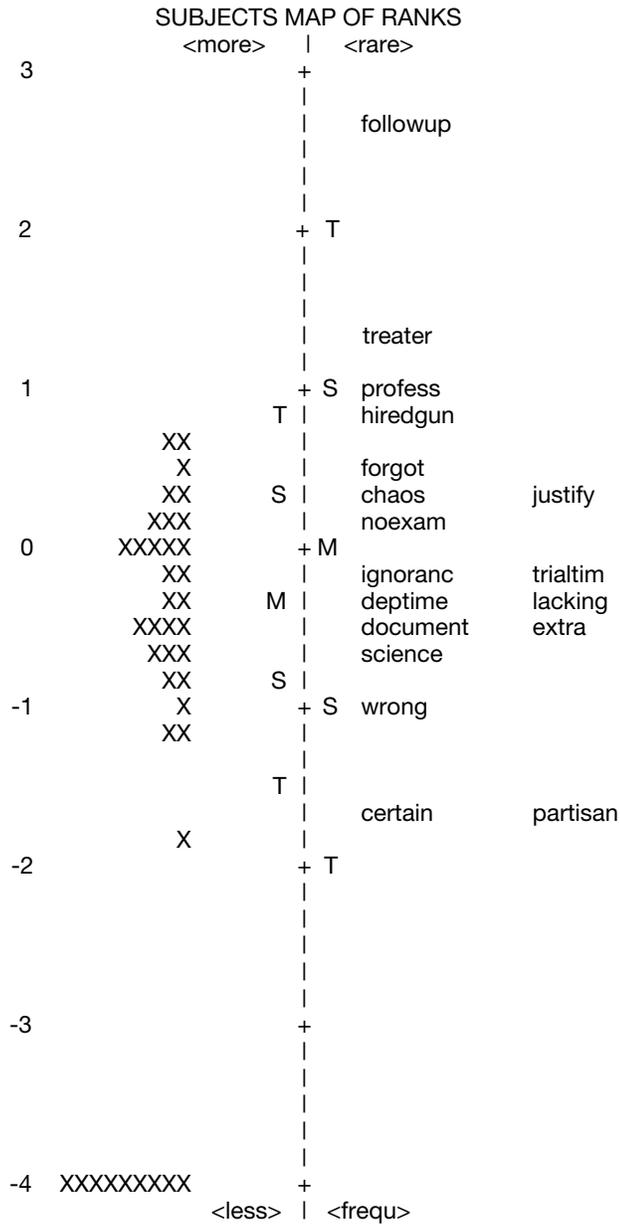
Discussion

Somewhat more than half of the participants in this study reported that they had been involved in at least one case in which the opposing expert was seen to have behaved with a lack of professionalism. An even larger number (78.8%) reported that they had been involved in cases where they saw the opposing expert as a "hired gun." There can be at least two reasons for this finding. One is that there are, in fact, a substantial number of cases in which some form of unprofessional behavior does occur among expert witnesses testifying in court.

The participants did in addition rate how often each of 12 possible factors had led to the observed unprofessional behavior, something that bolsters the idea that the unprofessional behavior did in fact occur. Both the mean ratings and the Rasch scores suggest that one of the major reasons for a judgment of lack of professionalism was giving testimony that in fact contradicted evidence in some way, something that was observed to happen when an expert behaved in an overtly partisan way, espoused theories not supported by the facts in the case, or went beyond the science of the field. Carelessness or a lack of preparation were observed less often, as seen by Rasch scores that were much less negative for items such as the expert's ignorance of the relevant legal standard, failure to read material or failure to have completely reviewed relevant documents or records. Other factors, such as an expert witness having been an examinee's treater, or being a "hired gun," were rated as much less often observed (mean ratings closer to the minimum rating of 1) and obtained positive Rasch scores.

It is important to point out as well that while both the mean ratings and the Rasch scores are ordered in the same way, it is only by looking at the Rasch scores that we can see that certain factors, such as giving testimony that contradicted evidence in some way, were rated as twice as likely to be

included in a judgment of a lack of professionalism than others, such as being the examinee's treater.

Conclusion

The results of this study point to the critical role of freedom from bias and maintenance of objectivity. In those responses that cited going beyond the facts of science of the case and those that critiqued opposing experts as espousing unsupported theories, the underlying message is that our subjects saw opposing experts as bending the reality of the case to meet partisan demands. These findings may be useful to experts seeking to improve their practice and attorneys seeking to cross examine experts.

References

Andrich, D. (1988). *Rasch models for measurement*. Thousand Oaks, CA: Sage.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155-159.

Commons, M. L., Miller, P. M., & Gutheil, T. G. (2004). Expert witness perceptions of bias in experts. *Journal of the American Academy of Psychiatry and the Law*, *32*, 70-75.

Dattilio, F. M., Commons, M. L., Adams, K. M., Gutheil, T. G., & Sadoff, R. L. (2006). Pilot Rasch scaling of lawyers' perceptions of expert bias. *Journal of the American Academy of Psychiatry and the Law*, *34*, 482-491.

Gutheil, T. G. (2000a, October). Presidential address. Speech delivered at the annual meeting of the American Academy of Psychiatry and the Law, Vancouver, BC.

Gutheil, T. G. (2000b). The presentation of forensic psychiatric evidence in court. *Israeli Journal of Psychiatry and Related Sciences*, *37*, 137-144.

Gutheil, T. G. (2001). Adventures in the twilight zone: Empirical studies of the attorney-expert relationship. *Journal of the American Academy of Psychiatry and the Law*, *29*, 13-17.

Gutheil, T. G., Commons, M. L., & Miller, P. M. (2001). Personal questions on cross-examination: A pilot study of expert witness attitudes. *Journal of the American Academy of Psychiatry and the Law*, *29*, 85-88.

Gutheil, T. G., & Sutherland, P. K. (1999). Forensic assessment, witness credibility, and the search for truth through expert testimony in the courtroom. *Journal of Psychiatry and Law*, *27*, 289-312.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.

Strasburger, L. H., Gutheil, T. G., & Brodsky, A. (1997). On wearing two hats: Role conflict in serving as both psychotherapist and expert witness. *American Journal of Psychiatry*, *154*, 448-456.

Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago: Mesa Press.