
by Gordon Welty
Wright State University
Dayton, OH 45435 USA

[//129]Abstract
Two major problems in selecting experts for the Delphi Technique in cost sensitive educational forecasting are considered: 1. Differentiation between levels of expertise; 2. The relevance of experts in forecasting areas where the subject matter is cultural values rather than technology. It is hypothesized that where values comprise the forecast subject matter there will be no significant differences between expert and lay predictions. Evidence is presented from a replication of Rescher's study of the future of American values using lay judges in place of experts. Comparison of the mean item scores for experts and laymen reveals no significant differences for 14 of 17 items, and differences for two items which can be explained on grounds other than differential expertise. Therefore the relevance of experts in cost-sensitive, value-laden educational forecasting is questioned. On cost-effective grounds, it appears that experts are used in such forecasts because of their high status rather than demonstrated forecasting superiority.

Introduction
A recent development in long-range forecasting is the Delphi technique of aggregating the forecasts of a number of experts on multidisciplinary issues. This technique has shown promise in educational forecasting and planning. By a process of sequential polling of the experts' opinions, interspersed with feedback to the experts of information on the just previous poll, consensus is generated. Meanwhile anonymity is guaranteed the experts in the polling procedures, so that the social psychological pressures present in the usual committee and face-to-face group approaches to multidisciplinary issues are avoided. Rather than the best talker holding sway over the committee deliberation, presumably the "wisest minds" can prevail, forensic abilities notwithstanding. The efficiency of the Delphi technique is thus greater than the efficiency of the usual committee. Efficiency considerations presuppose cost-sensitivity in forecasting exercises. We assume that the educational planner operates with limited resources. In light of Dr. Sidney Marland's recent "belt tightening" activities in the U.S. Office of Education, this assumption appears warranted.

A crucial factor for the Delphi exercise is clearly that of the selection of experts. Experts require honoraria, etc. that laymen do not. Only in relatively cost-insensitive forecasting exercises, perhaps those conducted by the Department of Defense, would the selection of experts not be a crucial consideration on grounds of cost factors. This selection problem has been confronted on two levels, and we shall consider how this bears on the use of Delphi in educational forecasting and planning.

On one hand, there has been an attempt to distinguish greater and lesser subject-matter expertise among a given group of experts. On the other hand, there has been an attempt to distinguish within given forecasting-subject matter the relevance of expertise to its forecasting. We will briefly note some of the literature on the former consideration, then turn to the latter
consideration and present empirical findings that bear upon it.

Levels of Expertise
While a number of the developers of Delphi have attempted to distinguish among levels of expertise, the problem is far from resolution. Brown and Helmer, for example, propose that one's degree of expertise might be ascertained by a self-assessment of forecast subject-matter knowledgeability by the Delphi participant himself. More recently, an attempt to test this sort of self-assessment technique for selection of experts (or for the weighting of expert opinions) showed equivocal findings. Two British investigators were so dissatisfied with the outcome of self-weighting in a Delphi study that they flatly state they "would not introduce the complication of self-weighting in any future study." These two Delphic studies addressed the subject matter areas of medicine and manufacturing. It seems plausible that any methodological problems that would arise in these areas would also present themselves in the much more complex human activity of education. Suffice it to say that the problem of differentiating among various levels of expertise requires more research. Let us now turn to the problem of distinguishing classes of judgmental forecasting issues for which expertise is required.

Are Experts Necessary?
An important early inquiry into the parameters governing judgmental forecasting exercises was Project Outcomes. As part of this long-term effort, 900 students and 778 legislators in seven countries were sequentially polled on their anticipations of cold-war outcomes. Expertise on internationally and domestically held values is presumably concentrated in the legislators. Nehnevajsa, the principal investigator, noted that the majority of all correlations (78.3%) exceeded .76, which strongly suggests that the difference in anticipations (or predictions of outcomes) are not substantial between the students and legislators.

Further heuristic documentation that expertise in the domain of prediction of values is nonexistent is contained in the Project Outcomes report of Shively, where he states: "There are innumerable minor differences of response between the respondents from the various countries, and, ...this difference was generally greater than that found between the students and legislators." While the Project Outcomes studies were similar to Delphic exercises in many ways, they were not expressly Delphi-type exercises and did not explicitly take into account the possibility of interaction effects. They do, however, strikingly frame the problem under consideration.

It is widely acknowledged that the forecasting of values is as important as the more conventional forecasting of technological breakthroughs, since technological efforts are dependent upon a complex set of value judgments for funding, etc. Most definitions of technological forecasting make reference to "levels of support," which presuppose a set of values. In education, it is plausible that support is more important than technological developments per se. The relatively monolithic institutional framework serves to enhance support, hence value considerations, while the "humanistic" orientation of American education tends to downgrade purely technological considerations.

Quinn's prescription simply to eschew long-range forecasting of global and distal values in favor of short-run projections of proximate goals over a few years avoids a basic question. Notice that the intergenerational concern inherent to education demands a long-range perspective. While
the selection of experts may be less problematic for such short-run forecasting exercises, this has not been demonstrated. Dalkey, a major contributor to the development of the Delphi technique, admits of "the area of evaluating experts... We haven't done a great deal of work on this at RAND and it is clearly a very difficult subject."/10/ Another question remains even if the selection of experts for short-run forecasting is unproblematic. This question relates to the nature of the parameter which specifies the short-run horizon wherein the use of experts in judgmental forecasting is warranted, and is a question that clearly requires further empirical study.

Similarly, Judd's supposition that the selection of experts in the area of educational forecasting is not problematic because, in his words, "in matters pertaining to education...it would be hard not to find `experts' on every hand,"/11/ simply ignores the selection issue entirely. In fact, are there experts on educational matters on every hand? If so, what is meant by expertise in this context?

**Forecasting Value Change**

Let us turn then to an empirical study of global and distal values which pertain to the societal framework within which the educational institution exists. In a Delphi-like study of the future of American values, Rescher polled a selection of future-oriented individuals, focussing in his selection on "high-level scientists and science administrators."/12/ This sort of technocratic approach to long range forecasting has frequently been criticized. Linstone and Slofkosky commented [132/133] on the exclusion of social scientists and other non-engineers from the famous TRW long-range forecasts of the middle and late 'Sixties/13/ but the tendency to restrict participants in this fashion remains. Apparently, some forecasting exercises are based on the North-Pyke TRW assumption that very few (preferably engineers or natural scientists) can forecast, while other forecasting exercises are based on the Judd assumption that just about anyone can forecast. Certainly empirical evidence might bear on such a range of opinion!

We have just completed the replication of a substantial portion of Rescher's study. The items employed here were identical to Rescher's, and comparable procedures were used throughout. In this replication, however, the respondents differed in terms of level of expertise as well as in terms of their academic disciplines. Forty-three undergraduate (predominantly junior) sociology majors constituted the group of lay forecasters.

For each of Rescher's 17 items (representing a value of American culture in the year 2000 A.D.)/14/ an opinion of the probable change in emphasis was elicited on a five-point scale (ranging from 1 = greatly increased emphasis to 5 = greatly decreased emphasis). For instance, the respondents were asked whether American society in the year 2000 would value "Intellect" more highly, less highly, or the same as it values it today. Each item mean was computed and compared by the t-test with the (rescaled) item means reported by Rescher. (For item 10, computational errors in Rescher's report forced us to eliminate it from consideration.) Since the covariance structure of Rescher's data was unavailable, it was not possible to compute the overall F-statistic as we would like to. It is important to bear in mind that the overall F will be underestimated in the case of correlated dependent variables. Instead, we have computed a t-test for each of the 16 items, in the fashion of posterior pairwise analysis.

For 14 of the items we found no significant difference at p < 0.05 between the mean responses of Rescher's respondents and the student respondents. For two of the items we did find significant differences. Rescher's respondents expected significantly more emphasis on "material" values and significantly less emphasis on "spiritual" values than did the student respondents. The table
below gives the mean values and standard deviations for each item responded to by Rescher's experts and the students. [133/134]

Comparative Expert and Student Judgements of Value Changes

<table>
<thead>
<tr>
<th>Values</th>
<th>Experts' Means</th>
<th>Experts' S.D's</th>
<th>Students' Means</th>
<th>Students' S.D's</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Self-regarding</td>
<td>2.90</td>
<td>0.99</td>
<td>2.81</td>
<td>1.14</td>
</tr>
<tr>
<td>2. Other-regarding</td>
<td>2.39</td>
<td>0.19</td>
<td>2.19</td>
<td>0.96</td>
</tr>
<tr>
<td>3. Material*</td>
<td>2.14</td>
<td>0.79</td>
<td>3.00</td>
<td>1.23</td>
</tr>
<tr>
<td>4. Spiritual**</td>
<td>3.04</td>
<td>0.91</td>
<td>2.44</td>
<td>1.30</td>
</tr>
<tr>
<td>5. Aesthetic</td>
<td>2.35</td>
<td>0.81</td>
<td>2.47</td>
<td>1.03</td>
</tr>
<tr>
<td>6. Religious</td>
<td>3.87</td>
<td>0.72</td>
<td>3.74</td>
<td>1.58</td>
</tr>
<tr>
<td>7. Personal</td>
<td>2.67</td>
<td>1.04</td>
<td>2.37</td>
<td>1.09</td>
</tr>
<tr>
<td>8. Social</td>
<td>2.14</td>
<td>0.66</td>
<td>2.16</td>
<td>0.81</td>
</tr>
<tr>
<td>9. Local</td>
<td>3.29</td>
<td>0.87</td>
<td>2.98</td>
<td>1.21</td>
</tr>
<tr>
<td>10. National</td>
<td>--</td>
<td>--</td>
<td>3.72</td>
<td>1.22</td>
</tr>
<tr>
<td>11. International</td>
<td>2.15</td>
<td>0.64</td>
<td>1.93</td>
<td>0.99</td>
</tr>
<tr>
<td>12. Prowess</td>
<td>2.96</td>
<td>0.89</td>
<td>3.00</td>
<td>1.23</td>
</tr>
<tr>
<td>13. Intellect</td>
<td>2.14</td>
<td>0.82</td>
<td>2.02</td>
<td>0.83</td>
</tr>
<tr>
<td>14. Character</td>
<td>2.83</td>
<td>0.94</td>
<td>2.63</td>
<td>1.16</td>
</tr>
<tr>
<td>15. Self-oriented</td>
<td>2.87</td>
<td>0.87</td>
<td>3.14</td>
<td>1.19</td>
</tr>
<tr>
<td>16. Parochial</td>
<td>3.62</td>
<td>0.69</td>
<td>3.63</td>
<td>1.09</td>
</tr>
<tr>
<td>17. Humanitarian</td>
<td>2.69</td>
<td>0.96</td>
<td>2.30</td>
<td>1.17</td>
</tr>
</tbody>
</table>

* Sig. p < 0.001, Student's t, df. 93  
** Sig. p < 0.01, Student's t, df. 93

Conclusion
Thus, we can conclude that our earlier reported findings that expertise may not be relevant to the
forecasting of values are sustained. The two items where a significant difference between the means was observed could best be explained as evidencing the "idealism" which is widely prevalent (and fostered) among students. It is questionable if "experts" have any function other than providing prestige to the cost-sensitive value forecasting exercise.

What does this offer to the educational planner about to embark upon a Delphi-type long range forecasting exercise? On the one hand, he should carefully consider his forecasting subject matter. If it is "value-laden" subject matter rather than technological subject matter, he might care little to expend the time and money to seek out experts. Attitudes in the year 2000 A.D. towards man-machine learning systems would require a different sort of forecaster than the technological possibility by the same year of a cost-effective feasible multimedia learning system.

On the other hand, he should carefully consider his own motives in forecasting. If his needs are other than purely informational, then he may still seek to empanel a "blue-ribbon committee" of experts in education to provide legitimacy to his forecasting exercise.

Notes


