

Using the FASB's Qualitative Characteristics in Accounting Policy Choices

EDWARD J. JOYCE*, ROBERT LIBBY†, AND SHYAM SUNDER‡

Financial accounting standard setting involves making choices among alternative accounting methods. Accounting methods are multidimensional in nature, and each has many attributes which may be relevant for making a choice. The latest attempt to articulate relevant attributes appears in *Statement of Financial Accounting Concepts No. 2: Qualitative Characteristics of Accounting Information* (FASB [1980]).¹ Underlying this endeavor is the assumption that identifying and defining the appropriate attributes (qualitative characteristics) of accounting information will help standard setters in selecting financial accounting methods. The purpose of this study was to test this assumption for *SFAC No. 2* (hereafter, *Statement*).

A direct way to test the effectiveness of the *Statement* would be to assess the "quality" of standard setters' judgments, before and after the

* Associate Professor, University of Minnesota;† Associate Professor, University of Michigan;‡ Associate Professor, University of Minnesota. Financial support for this project was provided by the Price Waterhouse Foundation through the Institute of Professional Accounting at the University of Chicago and by the Paton Accounting Center. We are grateful to Bob Mautz, Al Rappaport, Roman Weil, Art Wyatt, and the 26 participants who generously gave their time to the project, and in particular to Sidney Davidson for his overall aid. Useful comments were provided by the participants of various accounting workshops and by members of the FASB. [Accepted for publication February 1982.]

¹ Earlier attempts include *ASOBAT* (AAA [1966, p. 7]), *APB Statement No. 4* (AICPA [1970, pars. 57-60]), *The Corporate Report* (Accounting Standards Steering Committee of the Institute of Chartered Accountants in England and Wales [1975, p. 28]), and the *Discussion Memorandum on the Conceptual Framework for Financial Accounting and Reporting: Elements of Financial Statements and Their Measurement* (FASB [1976, pars. 317-81]).

statement. Unfortunately, we know of no operational ways to measure the quality of their decisions. We tested instead for what we believe are three necessary conditions for the *Statement* to facilitate standard setting. First, the qualitative characteristics should be *operational*. We assessed the operational status of the qualitative characteristics within a multi-trait-multimethod paradigm proposed by Campbell and Fiske [1959]. Second, the qualitative characteristics should be *comprehensive* in that they represent all factors important for making accounting policy choices. Third, there should be a minimum amount of overlap in meaning among the characteristics. Significant redundancies in meaning would suggest that a more *parsimonious* set of characteristics would serve policy makers equally well. These three criteria—sometimes called convergent validity, predictive validity, and discriminant validity—are frequently suggested in the social science literature for evaluating the usefulness of concepts, and the FASB has expressed concern about the degree to which their qualitative characteristics meet each of these criteria.

The FASB characteristics were tested in an experiment using former members of the APB and FASB. Our results suggest that two of the characteristics—verifiability and cost—are operational as defined in the statement, but not the others. In addition, while the qualitative characteristics appear to constitute a comprehensive set of attributes for accounting policy choices, the set is not parsimonious.

Role of Policy Makers

The *Statement* explicitly identifies and defines the qualitative characteristics of accounting information and asserts that they should be the basis for choosing among accounting methods (FASB [1980, par. 5]), but without indicating how they should be used. It would seem that these characteristics could serve as the basis for choosing among accounting methods only if the policy makers can first assign measures to the amounts of each qualitative characteristic possessed by each accounting alternative under consideration and then somehow aggregate these measurements into a preference for one accounting method over its alternatives. Neither task seems easy when considering such abstract concepts as relevance, verifiability, etc.

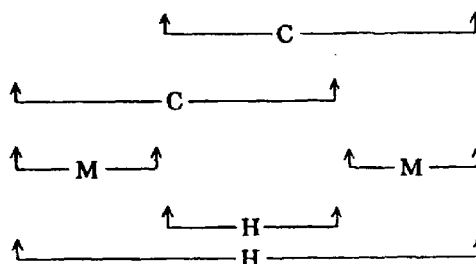
THE MULTITRAIT-MULTIMETHOD MATRIX

The ability of policy makers to assign operational measures to the qualitative characteristics of two (or more) accounting alternatives was assessed using the multitrait-multimethod matrix of Campbell and Fiske [1959].² The choice of operationality as a criterion was suggested by the FASB [1976, pars. 326-28, 380]. The application of the Campbell and Fiske matrix can be illustrated as follows.

Suppose that two accounting policy makers were asked to measure, say, on seven-point scales, the relevance and reliability of the accounting data produced by the alternative accounting methods for oil and gas exploration costs. The methods would include full cost, successful efforts, reserve recognition, and others. In the multitrait-multimethod matrix, relevance and reliability are viewed as traits and the two policy makers as measurement methods. Figure 1 represents the task and highlights some important relationships among the measurements.

For these policy makers' measurements of relevance and reliability to be valid, certain conditions should be met. First, there should be high agreement between each policy makers' ratings of the two characteristics. This agreement is denoted *C* in figure 1 and is called *convergent validity*. If there is little agreement between the ratings of different policy makers on the same qualitative characteristic, the measurements are not likely to be operational.

Accounting Methods	Policy Maker Number			
	1		2	
	Relevance	Reliability	Relevance	Reliability
Full Cost	7	6	6	5
Successful Efforts	3	4	4	3
Reserve Recognition	4	1	5	5
.
.
.



Term	Policy Maker	Qualitative Characteristics
M	Same	Different
C	Different	Same
H	Different	Different

High convergent validity, however, is but one necessary condition for operationality to exist. Valid measuring instruments should measure like things alike and different things differently. In our example, perfect convergent validity could be attained simply by each policy maker assigning a rating of "5" to every accounting method for both characteristics. These measures would be invalid if in fact the different methods varied on these qualitative characteristics and relevance was distinct from reliability. Under these conditions, there should be greater agreement between the measures of the different policy makers on reliability (C) than the agreement between one policy maker's measure of reliability and his (her) measure of relevance (denoted M). Similarly, greater agreement should exist between the different policy makers' measures of reliability (C) than the agreement between one policy maker's measures of relevance and the other's measures of reliability (denoted H). To the extent that $C > M$ and $C > H$, the measurements are said to represent distinct concepts and to possess *discriminant validity*. In developing the *Statement*, the FASB sought to include only those characteristics which added something important to the list and to exclude those which were redundant [1980, pars. 158-60]. Failure to meet the two tests for discriminant validity would indicate that the Board did not meet this goal and that a more parsimonious list could be constructed.

The C , M , and H measures discussed above are usually presented in a multitrait-multimethod correlation matrix. An example of the matrix with three policy makers and three qualitative characteristics is shown in table 1. The upper triangular submatrices along the main diagonal are called monomethod blocks because they contain the correlations between a *single* policy maker's ratings on the three qualitative characteristics. Each diagonal element (R) of the monomethod blocks is the correlation between a *single* policy maker's ratings on the *same* qualitative characteristics at two different times. The R 's thus represent a measure of the reliability of the policy maker's evaluations on each qualitative characteristic. The off-diagonal elements (M) in the monomethod blocks are the correlations between the *same* policy maker's ratings on *different* qualitative characteristics (e.g., relevance and verifiability).

The off-diagonal submatrices are known as heteromethod blocks. Their diagonal elements (C) are the correlations between the ratings of *different* policy makers on the *same* qualitative characteristics. They measure convergent validity—the extent of agreement among policy makers. The off-diagonal elements (H) in the heteromethod blocks represent correlations between *different* qualitative characteristics rated by *different* policy makers.

POLICY PREFERENCES

In addition to operationality and parsimony, the FASB expects the qualitative characteristics will comprise a comprehensive set of criteria

qualitative characteristics will have been omitted. This would then allow one to predict policy choices from assessments of the characteristics. We now discuss how qualitative characteristics can be used to make policy choices.

Although the *Statement* does not indicate how policy makers can move from measures of qualitative characteristics to policy decisions, it does indicate the possibility that they may have to make trade-offs among the qualitative characteristics (FASB [1980, pars. 31, 42, 57, 90, 133]). This suggests that a *compensatory* model of choice can represent policy makers' decisions.

A mathematical expression of a linear compensatory model of choice would be:

$$A_j = \sum_{i=1}^n w_i X_{ij} \quad j = 1, 2, \dots, m$$

where

A_j = the index of preference of a subject for accounting alternative j , $j = 1, 2, \dots, m$.

w_i = importance weight attached by the subject to the i th qualitative characteristic, $i = 1, 2, \dots, n$.

X_{ij} = the amount of qualitative characteristic i in accounting method j , $i = 1, \dots, n$; $j = 1, \dots, m$.

n = number of qualitative characteristics.

m = number of alternative methods.

This model implies that a policy maker behaves as if (s)he applies the following procedure for each of the m accounting alternatives. First (s)he assigns a score to each of the alternatives' n qualitative characteristics (relevance, reliability, etc.). These attribute scores, X_{ij} , are then multiplied by their relative weights, and the sum of these products, A_j , becomes a measure of a policy maker's preference for method j . Each subject then chooses the accounting method with the highest score, A_j .

Although this formulation of how accounting choices are made by individual policy makers seems simplistic, it does permit trade-offs between the attributes in a manner which is consistent with descriptions of the choice process made by accounting policy makers (e.g., Davidson [1969]). It has also been shown that a linear model provides a reasonable representation of the cognitive process employed in choices when the alternatives and the attributes are few in number (Payne [1976]). Its predictive accuracy has also been extensively demonstrated (Libby [1981]).

We estimated the above model for each policy maker from responses supplied in the experiment (see below). We validated each model by using it to predict each subject's accounting policy choices.

Method

The experiment required the following assessments by each subject: (1) the amount of each qualitative characteristic for each accounting alternative (X_{ij}); (2) the importance of each qualitative characteristic (w_i); and (3) each subject's preferences between accounting alternatives. A model of each subject's choices was then estimated using a simple estimation technique developed by Einhorn and McCoach [1977]. The technique has proved to be easy to implement and, in cases where a small number of unrestrictive assumptions are met, it provides accurate predictions of individual choices. The most important assumption is conditional monotonicity, which requires that more (less) of an attribute be preferred to less (more) for any given level of the other attributes (e.g., more relevance or less cost is preferred at any given level of verifiability). The manner in which the qualitative characteristics are written in the *Statement* presumes this condition.

PROCEDURE

Each policy maker was presented with (1) a two-page foldout containing brief descriptions of eight accounting policy issues, (2) a booklet containing instructions, forms for rating the policy issues, and a debriefing questionnaire, and (3) a copy of the *Statement*.³ To simplify the task and the analysis, each policy choice was limited to two major alternatives.

The policy makers were asked first to rate numerically each of the alternatives for the eight accounting policy issues on each of the 11 qualitative characteristics developed by FASB.⁴ These were presented in random order. Each qualitative characteristic was presented on a single page for the eight policy issues (see Appendix A). The definition of each characteristic was taken directly from the *Statement* glossary with the appropriate page reference. A reference to the more extensive explanation in the body of the *Statement* was also provided.

For each of the accounting choices, the policy makers indicated their beliefs about (1) which of the reporting alternatives has more of the qualitative characteristic and (2) how much more it has of the qualitative characteristic (for an illustration, see Appendix A). The policy makers were told they could refer to the attached copy of the *Statement*, if necessary, for further explanation of the definitions. Previous answers could be changed as the subjects proceeded through the task.

Following the above, policy makers then ranked the 11 qualitative characteristics in order of their perceived importance (ties were allowed). Finally, they were asked to choose the accounting method they preferred

³ These materials were extensively pilot tested with two groups of senior academics and practitioners.

⁴ The 11 criteria include the 10 qualities listed in figure 1 of the *Statement* plus the cost portion of the benefits-costs "passive constraint." The exact number of characteristics

(or mark no preference) and to indicate the strength of their preference (see Appendix B).

After they had completed all tasks, they were asked to check whether they had responded to all questions. They then responded to questions about the clarity of instructions, time taken to complete the task, and attitudes toward the *Statement* and the experiment. Space was also provided to allow subjects to indicate the policy issues where the two listed accounting alternatives did not include their most-preferred alternative.

ISSUES

The eight accounting policy issues chosen for evaluation in this study involved measurement issues as opposed to disclosure issues. Each had previously been addressed by either the APB, the FASB, or both. They differed considerably in the controversy they had generated (as measured by number of dissenting votes).

Two popular alternatives were selected for each issue. In the six cases in which the APB or FASB had previously selected a single reporting method, that method was included in the pair. In the two instances in which two reporting methods were allowed (investment tax credit and inflation accounting), both were included. (The issues and alternatives appear later in table 6.)

SUBJECTS

The subject population was chosen to insure experience in dealing with accounting policy choices and with the use of earlier qualitative criteria designed to facilitate accounting choices. The total available population was defined as all living individuals who had previously served on the FASB or its predecessor, the APB, since 1968. These subjects were thought to be closest in background and experience to the current members of the FASB. We tried to contact each of these former board members by phone to secure their participation in the experiment, reaching 29 of them. One declined to participate. Of the 28 who agreed to participate, 26 returned their completed experimental booklets. Of these, 20 had served only on the APB, 3 had served only on the FASB, and 3 had served on both boards. Prior to their service on the Board, 11 were from the Big Eight, 5 were from other CPA firms, 5 were from industry or government, and 5 were from universities.

Results

As indicated above, our tests of the *Statement* were based on three criteria suggested by the FASB. The qualitative characteristics should be (1) *operational*, and they should comprise a (2) *comprehensive*, yet (3) *parsimonious*, set of characteristics for choosing among policy alternatives. Criteria (1) and (3) were tested across policy makers using the

Operationality was tested by assessing the extent of consensus across policy makers on the meaning of the qualitative characteristics (convergent validity) and the extent of overlap in meaning among the qualitative characteristics perceived by the policy makers (discriminant validity). Both comprehensiveness and parsimony were tested by assessing the predictive ability of our compensatory model of each policy maker. If the qualitative characteristics are the basis for choosing among accounting policy alternatives, these choices should be systematically related to the relative amounts of each qualitative characteristic possessed by the alternatives, and the relative importance of the characteristics. An ability to predict each policy maker's preferences accurately from a model containing only the qualitative characteristics would suggest a comprehensive set. But the degree to which preferences can be accurately predicted from considerably fewer than all 11 characteristics would indicate that the set is not parsimonious. We also tested the level of consensus on the relative importance of the characteristics and on the final policy choices.

OPERATIONALITY

The multitrait-multimethod paradigm was used first to evaluate the policy maker's ability to operationalize the qualitative characteristics in the context of specific policy issues. The complete multitrait-multimethod correlation matrix for the 11 characteristics and 26 subjects led to a 286×286 upper triangular *MTMM* matrix containing the following three types of measures for each characteristic:

- (1) *C*, the correlation between *different* policy makers' judgments on the *same* characteristics (convergent validity), $325 \times 11 = 3,575$ observations;
- (2) *M*, the correlation between the *same* policy maker's judgments on *different* characteristics (heterotrait, homomethod variance), $55 \times 26 = 1,430$ observations; and
- (3) *H*, the correlation between *different* policy makers' judgments on *different* characteristics (heterotrait, heteromethod variance), $325 \times 11 \times 10 = 35,750$ observations.

Convergent validity. Table 2 contains summary statistics on the *C* coefficient, which is used to assess convergent validity for the 11 individual qualitative characteristics across all of the issues. The overall mean coefficient across the 11 characteristics was .257, far from a perfect value of 1.00. The individual mean coefficients ranged from .522 down to .099. Only verifiability and cost averaged above .5, suggesting that these two characteristics have some common meaning to the policy makers. However, the other nine do not and one, representational faithfulness, has no common meaning at all.

Convergent validity for each accounting issue was examined across the 11 qualitative characteristics. Summary statistics for the eight issues

TABLE 2
Convergent Validity by Characteristic

	Under.	Verif.	Compar.	Reliab.	Time.	Repr. Faith.	Neut.	Feed. Value	Relev.	Pred. Value	Cost
Mean	0.236	0.522	0.181	0.307	0.139	0.099	0.302	0.184	0.138	0.198	0.520
Median	0.216	0.599	0.164	0.342	0.0	0.099	0.326	0.160	0.137	0.215	0.608
Std. Dev.	0.353	0.330	0.350	0.374	0.368	0.367	0.341	0.365	0.364	0.369	0.331
Range											
Min.	-0.835	-0.568	-0.753	-0.796	-0.778	-0.878	-0.712	-0.806	-0.770	-0.776	-0.610
Max.	0.989	0.989	0.944	0.957	1.000	0.931	0.988	0.982	0.954	0.933	0.990

TABLE 3
Convergent Validity by Issue

	Oil & Gas	Mkt. Sec.	Bus. Comb.	Tax Credit	R+D	Inflation	Devel. Stage	Extng. Debt
Mean	0.136	0.190	0.057	0.070	0.296	0.420	0.111	0.097
Median	0.127	0.253	0.095	0.035	0.310	0.489	0.125	0.0
Std. Dev.	0.372	0.400	0.355	0.322	0.305	0.325	0.426	0.388
Range								
Min.	-0.856	-0.842	-1.000	-0.742	-0.456	-0.867	-0.918	-0.892
Max.	0.935	0.934	0.886	0.865	0.927	0.988	0.941	0.970

inflation (mean: .42), was there even a modest degree of common meaning assigned to the qualitative characteristics. This is consistent with the earlier results.

Discriminant validity. Discriminant validity, or the overlap in meaning of the characteristics, was analyzed in two types of comparisons. The first, less restrictive test compares the correlations of *different* policy makers on the *same* traits (C = convergent validity) with the correlation of *different* policy makers' ratings on *different* traits (H); for discriminant validity, the former should be greater than the latter. In our 286×286 *MMTM* matrix, there are (220×325) 71,500 possible comparisons of this type.⁵ In 40,325 or 56.4 percent of these comparisons, C was greater than the corresponding value of H . Consistent with the earlier result, only cost and verifiability met this test more than 70 percent of the time.

In the second test of discriminant validity, the convergent validity coefficients (C) were compared with the correlations between the *same* policy maker's assessments of different characteristics (M). Overall, in only 36.7 percent of the 71,500 comparisons did the C 's exceed the corresponding M 's, and the condition was met more often than it was violated only for cost and verifiability. These results indicate that the distinctions between the definitions of the characteristics are smaller than the differences between policy makers' judgments. The results for each characteristic are presented in table 4.

COMPREHENSIVENESS

Our test of comprehensiveness involved assessing the ability of the estimated linear model for each policy maker to predict that person's accounting policy preferences.

Two approaches to weighting the attribute scores could be employed: (1) rank weighting and (2) equal weighting. A number of studies have indicated that importance weights derived from simple importance rankings lead to predictions as accurate as more complex direct and indirect ratio weight estimation techniques (see, e.g., Eckenrode [1965], Einhorn and McCoach [1977], Stillwell and Edwards [1979], and Stillwell, Barron, and Edwards [1980]), and that equal weighting of attributes will approximate the results from differential weighting under a wide variety of conditions (see, e.g., Dawes and Corrigan [1974], Einhorn and Hogarth [1975], Newman [1977], Wainer [1976; 1978], and Keren and Newman [1978]). Since our results were similar for rank- and equal-weighted schemes, only the former are reported.

Among the rank-weighting techniques, the rank-sum method has proven as accurate as any. Using this method, the importance ranks (obtained from the policy makers in the second part of the experiment)

⁵ These are comparisons of C 's against the H 's within the same row, column, and block.

TABLE 4
Discriminant Validity by Characteristic

Comparison*	Characteristic											
	Under.	Verif.	Compat.	Reliab.	Time.	Repr. Faith.	Neut.	Feed-back Value	Relev.	Pred. Value	Cost	Total
H < C	53.8	72.2	51.6	56.7	56.3	50.4	55.8	48.8	50.4	52.7	71.6	56.4
M < C	34.6	56.4	31.0	34.1	42.4	23.4	38.3	26.7	25.7	30.9	60.5	36.7

* Percentage.

TABLE 5
*Model Predictability for Seven Weighting Rules**

Number of Characteristics	Weighting	
	Differential	Equal
All eleven characteristics	89.4% (100%) .84 (.89)	86.1% (87.5%) .80 (.85)
Five most important	88.5% (93.8%) .81 (.91)	86.5% (87.5%) .81 (.90)
Two most important	83.6% (87.5%) .79 (.902)	78.4% (81.2%) .79 (.87)
Most important	**	76.0% (75.0%) .72 (.84)

* Percentages are hit rates, decimals are product moment correlations; means are presented outside brackets and medians within brackets.

** There are no weights applied in the single-variable model.

were transformed into weights by using the following formula:

$$W_i = \frac{N - r_i + 1}{\sum_{j=1}^N r_j},$$

where w_i is the weight for qualitative characteristic i , N is the number of qualitative characteristics (11), and r_i is the ranking of qualitative characteristic i .⁶

The qualitative characteristic weights for each policy maker k were then combined with his ratings on the qualitative characteristics, using the following equation:

$$A_{ik} = \sum_{j=1}^N w_{jk} \cdot d_{ijk}, \quad i = 1, 2, \dots, 8.$$

where A_{ik} is the preference score of policy maker k for policy issue i and d_{ijk} is the difference in the amount of qualitative characteristic between the alternative accounting methods on policy issue i as measured by policy maker k (arbitrarily scaled from +3 to -3). (Positive values indicate preference for alternative 1 and negative values indicate preference for alternative 2.)

The predictive accuracy of each subject's model was measured by absolute hit rates for choices and correlations with the nine-point strength of preference measure. Table 5 presents the mean and median hit rate and correlation for each weighting method. The predictive ability of the 11 characteristic models was high, with a mean hit rate of 89.4 percent and the mean correlation .84. For 15 of the 26 participants these models perfectly predicted the choices.

⁶ Ties were assigned the average rank (e.g., in the two-way tie for eighth, rank of $8\frac{1}{2}$ was assigned to both). Similarity of this ranking method to sum-of-the-years digits depreciation makes this method particularly dear to our hearts.

PARSIMONY

The lack of discriminant validity indicated in the *MTMM* analysis suggested that the characteristics comprise a less than parsimonious set. This might be expected, given that the FASB [1980, p. 15] indicates that the characteristics can be represented as a hierarchy with certain lower characteristics being components of others. A further direct test of parsimony was performed. We measured the ability to predict policy preferences accurately from less than all 11 qualitative characteristics by computing predictive ability using rank-sum and equal-weight models which included only the two highest and five highest ranking attributes. As a benchmark we also constructed a single variable model for the most important variable. Table 5 presents the mean and median hit rate for the three models. Retaining only the five most important characteristics in the models results in virtually no change in predictive ability. Predictability declined thereafter as the number of characteristics declined. The single most important characteristic produced a mean hit rate of 76 percent and mean correlation of .72. The number of perfect predictions (8 out of 8 issues) dropped from 15 policy makers for all 11 characteristics to 8 for the single characteristic model. Relevance alone predicted an average of 80.8 percent of the choices. These results again suggest that a more parsimonious list of characteristics can serve policy makers equally well.

SUPPORTING ANALYSIS

Disagreements on policy choices may have their basis in disagreements on the rating and ranking of the characteristics. In this section we report the level of consensus on important rankings of the characteristics and on final policy choices. An assessment of the external validity of the experimental policy choices and the debriefing results are also presented here.

Agreement on importance rankings and policy choices. Agreement among the policy makers on the relative importance of the qualitative characteristics was measured by computing rank-order correlations between all pairs of policy makers' importance rankings. The mean of these 325 correlations⁷ was .301 (standard deviation .295), indicating little agreement on the importance of the characteristics. The range was $-.655$ to $.861$. Table 6 lists the characteristics in order of the sum-of-their-importance ranks and presents the mean, median, standard deviation, and range of the ranks for each characteristic.

Agreement on the accounting policy choices was assessed through paired correlations between all pairs of policy makers' choices, computed both with and without measures of the strength of preferences. Given the low levels of agreement on both the meaning of the characteristics and

⁷ Kendall's coefficient of concordance, which is a linear transformation of the mean correlation, is often used to measure agreement in rankings.

TABLE 6
Summary of Ranks Assigned to Characteristics

Characteristic	Mean Rank*	Median Rank	Std. Deviation of Rank	Minimum Rank	Maximum Rank
1. Relevance	2.19	1.5	2.04	1	10
2. Reliability	3.42	2.5	2.55	1	11
3. Understandability	4.15	4.0	2.46	1	9
4. Representational faithfulness	4.96	4.5	3.17	1	11
5. Comparability	5.31	5.0	2.60	1	10
6. Neutrality	5.62	6.0	3.06	1	11
7. Verifiability	6.23	6.0	2.85	1	11
8. Predictive value	7.00	8.0	3.05	1	11
9. Timeliness	7.27	8.0	2.36	2	11
10. Feedback value	7.50	9.0	2.97	1	11
11. Cost	8.54	9.0	2.79	2	11

* The mean rank was less than 6 because of ties.

their importance, one would expect little agreement on final choices. These expectations were met. The average correlation was .111 (.083) and the standard deviation was .119 (.146) with (without) consideration of strength of preference. The range was $-.678$ ($-.745$) to $.910$ (.942).

Counting votes of no preference with the majority, on average only a two-thirds majority existed among the policy makers. There was greatest agreement on the marketable securities issue, interestingly enough, with 76.9 percent favoring "market" over the generally accepted "lower of cost or market." The greatest disagreement was on the investment tax credit issue. Table 7 lists the policy makers' votes.

External validation of choices. As a partial test of the external validity of our responses, we compared policy choices made by the policy makers in this experiment to choices these same individuals made while serving as members of the policy-making boards. These tests were limited by the fact that only a small proportion of the participants voted on any one issue and not all issues in the experimental questionnaire were formulated in the same fashion as they were when votes in the policy-making boards were cast. Thirty-two actual policy votes could be unambiguously matched to the experimental policy votes. In 28 or 87.5 percent of these cases, the experimental responses were consistent with the actual votes. Further, three of the four disagreements occurred in the case of business combinations where three of the policy makers appear to have voted for the compromise position in *APB 16* in place of their most-preferred alternative.⁸

⁸ This result can be interpreted in a number of ways. First, this may be damaging to the whole project if policy makers' past experience dominates their use of the characteristics. Alternatively, this may be supportive in that it suggests the characteristics include the elements that were important to the policy makers when they served on the Board. Third, this may indicate that the *Statement* will be more beneficial to new members whose preferences are less well entrenched.

TABLE 7
Policy Votes*

OIL AND GAS EXPLORATION COSTS		RESEARCH AND DEVELOPMENT COSTS	
Successful efforts	69.2	Capitalization	42.3
Full cost	30.8	Expensing in current period	57.7
No preference	0	No preference	0
MARKETABLE EQUITY SECURITIES		SUPPLEMENTAL INFLATION ACCOUNTING	
At "market"	76.9	Constant dollar (general price level-adjusted) accounting	42.3
Lower of cost or market on a portfolio basis	23.1	Current (replacement) cost accounting	53.8
No preference	0	No preference	3.8
BUSINESS COMBINATIONS		EARLY EXTINGUISHMENT OF DEBT	
"Purchase" accounting only (no "pooling")	65.4	Ordinary income	23.1
"Pooling" when certain conditions are met, "purchase" otherwise	30.8	Extraordinary item	65.4
No preference	3.8	No preference	11.5
INVESTMENT TAX CREDIT		DEVELOPMENT STAGE ENTERPRISES	
"Deferral"	50.0	Statements like other companies	69.2
"Flow through"	42.3	Special accounting practices	30.8
No preference	7.7	No preference	0

* Presented in percentages; may not sum to 100 because of rounding.

Debriefing. The experimental debriefing inquired about the clarity of the instructions, time taken to complete the task, interest in the experiment, and familiarity with, and attitudes toward, the qualitative characteristics. Responses of the policy makers were generally positive. Only five participants found any ambiguity in the instructions, all of which involved a technical detail concerning the definitions of a single accounting policy alternative. No definition was mentioned more than once. Those who found the instructions unclear tended to find the definitions unclear ($r = .48$).

On average, the subjects spent 94 minutes completing the task, with all but four spending an hour or more. Eighty-three percent of the respondents considered the experiment interesting or very interesting and all 26 requested a copy of the results. The policy choices of those who found the experiment more interesting tended to be slightly more predictable ($r = .278$). For the 26 policy makers over eight accounting issues, there were only 19 cases (9.1 percent) where the participant's most preferred alternative was not presented in the experiment.

Five questions were asked about the qualitative characteristics. All but two policy makers responded that they had read *SFAC No. 2*. Only five found the definitions of any of the characteristics to be unclear. Represented by the first question, which produced the nearest convergent and

discriminant validity, was most often mentioned as being unclear. Consistent with the earlier reported importance rankings, relevance and reliability—the Board's primary criteria—were most often mentioned as aiding the decision process. Yet cost and verifiability, the most concrete of the 11 criteria, were cited as being of no help almost as often as they were mentioned to be helpful. Their usefulness in only selected situations may account for this result.

Summary and Conclusions

Many of the results reported here are not favorable to the *Statement*. Nine of the 11 qualitative characteristics clearly fail the tests of operability. Not only is there considerable disagreement among experienced policy makers on what the qualitative characteristics mean in the context of particular accounting policy issues, there is also considerable disagreement on their relative importance. While the qualitative characteristics appear to comprise a comprehensive set for choosing among accounting alternatives, the set is not a parsimonious one. Thus the *Statement* fails to meet two out of its three desired criteria.

This casts doubt on the ability of the qualitative characteristics defined in the *Statement* to facilitate accounting policy making. Indeed, one of the policy makers participating in our experiment responded on the written debriefing questionnaire: "This experiment does not give me much confidence in the usefulness of the qualitative characteristics in choosing among [accounting] alternatives."

In spite of the above findings, it would be imprudent to dismiss the *Statement* entirely. First, our test may have been premature given the incomplete nature of the conceptual framework project. The conceptual framework is intended to be "... a coherent system of interrelated objectives and fundamentals ..." (FASB [1980, p. i]). Testing only one part of the framework may bias the results in an unfavorable manner.

Second, each subject completed the survey independently of the others. Actual accounting policy choices are, of course, made in a group setting. It is possible that, during the deliberations of the Board, agreement on attribute ratings and importance may be significantly greater than we observed.

Third, the generalizability of our findings to future accounting policy issues and current and future members of the FASB is a matter for future research (as we noted earlier). While the eight accounting policy issues incorporated in the study represent a diverse set of past policy makers, their responses may not be representative of those of future policy makers.

Finally, our research design is a "one-shot case study." We have no measures of convergent or discriminant validity among the policy makers before the *Statement* so we cannot determine the marginal contribution of the *Statement*. It might be that agreement on choices, qualitative

characteristics, and importance weights was even lower before the *Statement* was issued. Further, as these new definitions become better known in the literature, the results may improve.

Nevertheless, the study does suggest that it is still not possible to make qualitative characteristics operational enough to guide policy makers in making accounting choices. For example, while relevance was the most important qualitative characteristic according to the *Statement* and the policy makers who participated in our study, there was virtually no agreement among the policy makers on which accounting alternative within a particular issue provided more of it.

A close reading of published documents relating to the *Statement* indicates that the Board is aware of this problem. For example, the related discussion memorandum alludes to the varying levels of abstraction represented in the list of characteristics and accurately lists verifiability among the most concrete (FASB [1976, para. 322]). Further, these documents reflect successively lower expectations for the immediate output of this part of the conceptual framework. Finally, the Board made its own attempt to assess empirically the operability of the characteristics in the context of more general accounting valuation issues (see FASB [1976, pp. 19-24]). The results have not been publicly reported.

Still, the participants in our study appeared to have little difficulty in expressing their global accounting policy preferences consistently through the qualitative characteristics. This suggests the possibility that the characteristics may prove to be useful for expressing individual preference over accounting issues for the purposes of discussion in Board meetings. Thus the framework may ease the task of developing compromises, a subject for future research.

If the Board wishes to continue to develop the qualitative characteristics, we would recommend that the issue of overlap between the characteristics be addressed with the aim of developing a more parsimonious list. The Board might also investigate whether disagreements on attribute ratings and attribute importance rankings are associated with differences in the goals and backgrounds of the various interest groups which participate in the policy-making process.

APPENDIX A
Example of Task 1

VERIFIABILITY

"The ability through consensus among measurers to ensure that information represents what it purports to represent or that the chosen method of measurement has been used without error or bias." (Glossary, p. xvi, for further explanation see *Statement*, paragraphs 81-89.)

1. EARLY EXTINGUISHMENT OF DEBT

- a. Which alternative provides data which are more *verifiable*?
(Circle one.)
- Ordinary income 1
- Extraordinary item 2
- Verifiability* does not distinguish between the two alternatives (skip to issue 2) 3
- b. How much more *verifiable* are the data from the alternative you have chosen? (Circle one.)
- Slightly* more verifiable 1
- Moderately* more verifiable 2
- Much* more verifiable 3

5. INVESTMENT TAX CREDIT

- a. Which alternative provides data which are more *verifiable*?
(Circle one.)
- "Deferral" 1
- "Flow-through" 2
- Verifiability* does not distinguish between the two alternatives (skip to issue 6) 3
- b. How much more *verifiable* are the data from the alternative you have chosen? (Circle one.)
- Slightly* more verifiable 1
- Moderately* more verifiable 2
- Much* more verifiable 3

2. RESEARCH AND DEVELOPMENT COSTS

- a. Which alternative provides data which are more *verifiable*?
(Circle one.)
- Capitalization 1
- Expensing in current period 2
- Verifiability* does not distinguish between the two alternatives (skip to issue 3) 3
- b. How much more *verifiable* are the data from the alternative you have chosen? (Circle one.)
- Slightly* more verifiable 1
- Moderately* more verifiable 2
- Much* more verifiable 3

6. BUSINESS COMBINATIONS

- a. Which alternative provides data which are more *verifiable*?
(Circle one.)
- "Purchase" accounting only (no pooling) 1
- Pooling when certain conditions are met; "purchase" otherwise 2
- Verifiability* does not distinguish between the two alternatives (skip to issue 7) 3
- b. How much more *verifiable* are the data from the alternative you have chosen? (Circle one.)
- Slightly* more verifiable 1
- Moderately* more verifiable 2
- Much* more verifiable 3

3. SUPPLEMENTAL INFLATION ACCOUNTING DATA FOR INDUSTRIAL FIRMS

- a. Which alternative provides data which are more *verifiable*?
(Circle one.)
- 1 Constant dollar (general price level-adjusted) accounting
 - 2 Current (replacement) cost accounting
 - 3 *Verifiability* does not distinguish between the two alternatives (skip to issue 4)
- b. How much more *verifiable* are the data from the alternative you have chosen? (Circle one.)
- 1 *Slightly* more verifiable
 - 2 *Moderately* more verifiable
 - 3 *Much* more verifiable

7. OIL AND GAS EXPLORATION COSTS

- a. Which alternative provides data which are more *verifiable*?
(Circle one.)
- 1 Successful efforts
 - 2 Full cost
 - 3 *Verifiability* does not distinguish between the two alternatives (skip to issue 8)
- b. How much more *verifiable* are the data from the alternative you have chosen? (Circle one.)
- 1 *Slightly* more verifiable
 - 2 *Moderately* more verifiable
 - 3 *Much* more verifiable

4. MARKETABLE EQUITY SECURITIES

- a. Which alternative provides data which are more *verifiable*?
(Circle one.)
- 1 At "market"
 - 2 Lower of cost or market on a portfolio basis
 - 3 *Verifiability* does not distinguish between the two alternatives (skip to issue 5)
- b. How much more *verifiable* are the data from the alternative you have chosen? (Circle one.)
- 1 *Slightly* more verifiable
 - 2 *Moderately* more verifiable
 - 3 *Much* more verifiable

8. DEVELOPMENT STAGE ENTERPRISES

- a. Which alternative provides data which are more *verifiable*?
(Circle one.)
- 1 Statements like other companies
 - 2 Special accounting practices
 - 3 *Verifiability* does not distinguish between the two alternatives (skip to next page)
- b. How much more *verifiable* are the data from the alternative you have chosen? (Circle one.)
- 1 *Slightly* more verifiable
 - 2 *Moderately* more verifiable
 - 3 *Much* more verifiable

APPENDIX B
Example of Task 3

IV.	A.	Which of the following alternatives do you prefer more for INVESTMENT TAX CREDIT? (Circle one number.)	
		“Deferral”	1
		“Flow through”	2
		No preference (skip to V)	3
	B.	How strong is your preference for the selected alternative over the other alternative? (Circle one number.)	
		Very mild	1
		Mild	2
		Strong	3
		Very strong	4

REFERENCES

- ACCOUNTING STANDARDS STEERING COMMITTEE. *The Corporate Report*. London: Institute of Chartered Accountants in England and Wales, 1975.
- AMERICAN ACCOUNTING ASSOCIATION. *A Statement of Basic Accounting Theory*. Evanston, Ill.: AAA, 1966.
- AMERICAN INSTITUTE OF CERTIFIED PUBLIC ACCOUNTANTS. *Statement of the Accounting Principles Board No. 4*. New York: AICPA, 1970.
- CAMPBELL, D. T., AND D. W. FISKE. “Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix.” *Psychological Bulletin* (March 1959): 81-105.
- DAVIDSON, S. “Accounting and Financial Reporting in the Seventies.” *Journal of Accountancy* (December 1969): 29-37.
- DAWES, R. M., AND B. CORRIGAN. “Linear Models in Decision Making.” *Psychological Bulletin* (January 1974): 95-106.
- ECKENRODE, R. T. “Weighting Multiple Criteria.” *Management Science* 12 (1965): 180-92.
- EINHORN, H. J. “Expert Measurement and Mechanical Combination.” *Organizational Behavior and Human Performance* (February 1972): 86-106.
- , AND R. M. HOGARTH. “Unit Weighting Schemes for Decision Making.” *Organizational Behavior and Human Performance* (April 1975): 171-92.
- EINHORN, H. J., AND W. P. McCOACH. “A Simple Multiattribute Utility Procedure for Evaluation.” *Behavioral Science* (July 1977): 270-82.
- FINANCIAL ACCOUNTING STANDARDS BOARD. *FASB Discussion Memorandum on Conceptual Framework for Financial Accounting and Reporting: Elements of Financial Statements and Their Management*. Stamford, Conn.: FASB, 1976.
- . *Statement of Financial Accounting Concepts No. 2: Qualitative Characteristics of Accounting Information*. Stamford, Conn.: FASB, 1980.
- KEREN, G., AND J. R. NEWMAN. “Additional Considerations with Regard to Multiple Regression and Equal Weighting.” *Organizational Behavior and Human Performance* (October 1978): 143-64.
- LIBBY, R. *Accounting and Human Information Processing: Theory and Applications*. New York: Prentice-Hall, 1981.
- NEWMAN, J. R. “Differential Weighting in Multiattribute Utility Measurement: When It Should Not and When It Does Make a Difference.” *Organizational Behavior and Human Performance* (October 1977): 312-24.
- PAYNE, J. W. “Task Complexity and Contingent Processing in Decision Making: An Information Search and Protocol Analysis.” *Organizational Behavior and Human Performance* (December 1976): 366-87.

- STILLWELL, W. G., F. H. BARRON, AND W. EDWARDS. "Evaluating Credit Applications: A Validation of Multiattribute Techniques Against a Real World Criterion." *Social Science Institute Research Report* (June 1980).
- STILLWELL, W. G., AND W. EDWARDS. "Rank Weighting in Multiattribute Utility Decision Making: Avoiding the Pitfalls of Equal Weights." Working paper, University of Southern California, 1979.
- WAINER, H. "Estimating Coefficients in Linear Models: It Don't Make No Nevermind." *Psychological Bulletin* (January 1976): 213-17.
- . "On the Sensitivity of Regression and Regressors." *Psychological Bulletin* (January 1978): 267-73.