DOCUMENT RESUME

ED 104 928 TM 004 384

AUTHOR	Gagne, Francoys; Allaire, Denis
TITLE	Reliability and Construct Validity of a Discrepancy Approach to Measuring Student Perceptions of Teaching. Document No. 0874-01.
INSTITUTION	Institut National de la Recherche Scientifique, Quebec (Quebec).
REPORT NO	0874-01
PUB DATE	74
NOTE	31p.; Paper presented at the Annual Meeting of the American Psychological Association (82nd, New Orleans, Louisiana, 1974)
EDRS PRICE	MF-$0.76 HC-$1.95 PLUS POSTAGE
DESCRIPTORS	College Teachers; Effective Teaching; Faculty Evaluation; Higher Education; Measurement Techniques; *Participant Satisfaction; Perception; Questionnaires; *Rating Scales; Statistical Analysis; Student Attitudes; Teacher Evaluation; *Teacher Rating; *Test Reliability; *Test Validity
IDENTIFIERS	Canada; *Discrepancy Approach; Students Perceptions of Teachers Questionnaire

ABSTRACT

The discrepancy approach, with a nonmonotonic function, was used in a questionnaire designed to measure the difference between reality as perceived and desires. Reality is described as a student's perception of teacher performance and notion of an ideal teacher performance. The questionnaire deals with both the intensity and direction of a student's dissatisfaction with his/her teacher. An experiment, utilizing 1,059 students from 52 classes in seven different French colleges, was designed to measure the construct validity of this questionnaire. Results are presented in tables entitled: (1) Content descriptions, means, and standard deviations of five constructs for each of 15 items, as well as the percentage of excess discrepancies; (2) Precision coefficients for a direct measure of dissatisfaction and for four constructs associated with a derived measure of dissatisfaction; (3) Observed correlations between five constructs used in the derived approach and direct measure of the satisfaction-dissatisfaction continuum; (4) Corelations for reliability attenuation; and (5) Observed mean within groups correlations between four constructs used in the derivation procedure and a direct measure of the satisfaction-dissatisfaction continuum. It is concluded that desires should be thought of as a continuum with limits being respectively the ideal and the minimally acceptable. (BJG)

INRS - EDUCATION

Reliability and construct validity of a discrepancy approach to measuring student perceptions of teaching

François Gagné, Ph.D.

and

Denis Allaire, M.A.

Paper presented as part of a symposium, entitled "Real-Ideal Discrepancy Estimation in Student Evaluation of Teachers: Psychometric Issues", held during the 82nd Annual Convention of the American Psychological Association, New Orleans, Louisiana.

Document 0874-01

Contents

INTRODUCTION

In 1967, one of us, then a professor of psychology in a Montreal college, used parts of his free time to construct with some of his students a course evaluation questionnaire intended for local use. The idea of using a discrepancy measure came neither from the litterature nor from a desire to verify a particular theoretical model, but from a pragmatic problem the group encountered.

The research team initially considered using a 5-point scale to measure directly the satisfaction-dissatisfaction continuum. But, since the questionnaire had a fair number of items in which satisfaction appeared related non-monotonically to reality, the direction of the dissatisfaction feeling, in terms of deficiency or excess, was deemed as important to measure as its intensity. This was the case for instance with items measuring the frequency of examinations, the quantity of compulsory readings, the speed in speech delivery, etc. It was deemed even plausible that some teachers could be judged too patient, too humble or modest, too prone to humor, or else that some courses could be judged excessively structured.

Looking for a simple procedure that would allow us to measure simultaneously the intensity and direction of student dissatisfaction, we devised an indirect or "derivation" approach, which rests essentially upon a measure of the discrepancy between reality as perceived and desires. From the relative position on a same scale of both measures we would identify the direction we were looking for, while the distance between them would give us the intensity of the dissatisfaction feeling. This procedure, developed intuitively by the research team, had already been tried, as we discovered later. In fact, well

before us, a few researchers had proposed or experimented discrepancy models similar to ours, in order to measure job satisfaction in industrial settings. Locke (1969) as well as Wanous and Lawler (1972) have reviewed extensively these efforts. Closer to our interests, Savage (1957), Gage, Runkel and Chatterjee (1960) and Thomas (1969) have used a fairly similar procedure in experiments designed to verify the influence of feedback as a means to change the behavior of teachers. But none of these authors seems to have noticed the advantages of this approach as applied to items in which satisfaction is a non-monotonic function of reality.

DESCRIPTION OF THE DERIVATION PROCEDURE

How does our procedure work? Figure 1 shows a typical item, as presented in the most recent form of the questionnaire, together with the instructions given to the respondents.

A short description of the attribute to be measured is accompanied by a 7-point scale. Our research (Allaire, 1974) has shown this

Question A: Where do you place this course (this teacher) on the evaluation scale?

If you are satisfied, mark question B at the same point on the scale as you have chosen for question A.

Question B: IF YOU ARE NOT SATISFIED, where should the course (the teacher) be placed on the scale so that you would be satisfied?

	extremely little		a little		to a certain degree		to a very great degree
31. This course is structured (the teacher follows a detailed plan).	1	2	3	4	5	6	7

Figure 1. Typical format and answering instructions in the SPOT questionnaire.

scale length to be significantly more reliable than a 5-point scale (cf. data sources H-I and H-II in Table 4). Four of the seven points are anchored with frequency (v.g. never - always) or intensity (v.g. very little - to a very great degree) qualificatives, depending on which are better adapted to the specific contents. As shown in the same figure, respondents are asked to answer two questions, item after item; we first ask about their perception of reality and, second, about the level deemed desirable for that particular attribute.

From the answers given to these only two questions, we compute at the class level six measures for each item, as shown in table 1. The first construct, called 'MA', that is the mean of the answers to question A, represents the teacher's actual behavior as perceived by his group of students. The construct labelled 'MB' reflects the students' desires and, by its relative position to MA, signals the direction of desired change.

The last four constructs are related to dissatisfaction. First, the construct labelled '%' simply gives the percentage of students assumed to be dissatisfied because their A answer differs from their B answer. Secondly, the construct called 'MAB' cumulates the absolute values of the discrepancies; thence, it indicates only the magnitude of the dissatisfaction expressed by students. On the other hand, the construct labelled 'MAL' (AL for algebraic) takes into account the direction of the discrepancies, so that, within a class, discrepancies in opposite directions will cancel each other. For this reason, we call this measure the "NET dissatisfaction score". Its sign specifies if the residual majority dissatisfaction corresponds to a deficiency or an excess. Finally, the construct called 'PCD' only makes clear, in terms of percentages, the difference between MAB and MAL, that is the degree of contradiction within the sub-group of dissatisfied students.

In order to give a more lifelike picture of these constructs, we present in Table 2 an interpretation of fictitious scores obtained

Table 1

Description of the six main constructs computed for each item in the PERPE-SPOT Questionnaire

Code	Formula	Construct description
MA	Σ(A answers)/N students	Level of teacher behavior as perceived. Interpreted by means of scale anchors.
MB	Σ(B answers)/N students	Level of student desires. Interpreted by means of scale anchors.
%	$\Sigma f(A \neq B) * 100/N$	Percentage of dissatisfied students, that is those students whose perceptions of reality differ from their desires.
MAB	$\Sigma(\lvert A - B \rvert)/N$	Gross dissatisfaction score. It is the sum of the absolute discrepancies, divided by all respondents, dissatisfied or not. It indicates the intensity level of the dissatisfaction felt in the group as a whole. It may vary from 0.00 to 6.00, but rarely exceeds 3.00.
MAL	$\Sigma(A - B)/N$	Net dissatisfaction score. It is the algebraic sum of the signed discrepancies, divided by all respondents dissatisfied or not. When all subjects have answered A and B, MAL = MA - MB. Also, MAL $\leq$ MAB, since dissatisfactions in opposite directions within the class will cancel each other. Since most item scales are ascending from 1 to 7 in terms of frequency or intensity, the "-" sign of the MAL will usually correspond to a deficiency and the "+" sign to an excess of the attribute measured. Same range as above.
PCD	$(MAB - \lvert MAL \rvert) * 100/MAB$	Percentage of within-group contradiction. 0% = all discrepancies in the same direction 100% = half of the discrepancies in each direction 50% = one-fourth of the discrepancies in a direction opposite to the other three-fourths. Note: PCD has been up to now the most "forgotten" construct in our research.

Table 2

Analysis of fictitious results obtained by
two instructors on the sample item presented in Figure 1

RESULTS

Instructor or class	MA	MB	%	MAB	MAL	PCD
X	4.15	4.80	45%	.70	-.65	07%
Y	6.50	6.20	63%	.70	+.30	57%

In the light of data gathered in thousands of classes, we can certify that these fictitious results are quite plausible. By placing the intensity construct (MAB) at the same level in both classes, we wish to show how differently two identical dissatisfaction measures can be interpreted through our other constructs.

a) To a same degree of dissatisfaction correspond quite different reality (MA) and desires (MB) measures. The degree of structure in course Y is perceived as much higher than in group X. Moreover, students in group X appear also to desire a higher degree of structure in their particular learning situation.

b) To a same degree of dissatisfaction correspond different percentages of dissatisfied students: this means that among the dissatisfied students more of them give discrepancies greater than $\pm$ 1 in group X (where % is lower) than in group Y. Since % (in decimals) is always equal or smaller than MAB, the larger the difference between them, the greater the frequency of discrepancies larger than $\pm$ 1 among dissatisfied students. In other terms, by comparing these two measures, we can tell if a given difference is due to a large goup of midly dissatisfied students or to a smaller but more dissatisfied group.

c) To a same degree of dissatisfaction correspond quite different NET dissatisfaction scores (MAL). In fact, group X signals a deficiency in course structure (MAL = -.65) as compared to an excess signalled by group Y (MAL = +.30). The lowering of the dissatisfaction intensity from MAB to MAL in group Y is due to a large degree of contradiction between dissatisfied students pertaining to the direction of the change desired; in fact, PCD shows that more than a fourth of these students would like a more structured course, even though the majority judges the course as already too structured.

In conclusion, without the information given by these measures, and taking into account the fact that students generally desire "well"-structured courses, the instructor in group Y, given only an evaluative judgment (v.g. MAB), might believe that there is a deficiency in course structure. Moreover, he would not be aware of a lack of agreement between his students as to the direction of the desired change. Finally, he would not know the actual degree of structure of his course as perceived by his students.

by two instructors on the item shown in Figure 1.

To the best of our knowledge, the PERPE-SPOT Questionnaire[1] appears to be the very first, and still the only, widely circulated questionnaire using a discrepancy approach to measure student perceptions of courses. Approximately four thousand college instructors in the province of Quebec have used the french form in the past seven years, and a few hundreds more, in Quebec and other provinces, have administered the english form. The richness of the information presented seems to allow for a more elaborate pedagogical diagnosis, which helps to make up for the longer administration time and the greater effort required on the part of the instructor to master the meaning of the different constructs.

THE MAIN PROBLEM: CONSTRUCT VALIDITY

Because of repeated requests from teachers, the PERPE-SPOT Questionnaire was offered in all Quebec colleges well before we had completed our program of research. Among other things, we did not have definitive empirical evidence pertaining to the underlying assumptions of this measuring technique. Since our procedure appeared to be unique as compared to similar questionnaires used elsewhere, it was to be expected that we would take special interest in the validation of the constructs presented above, specifically those related to the concept of satisfaction. The main problem was, of course, to verify if our discrepancy measure did in fact correspond to a direct expression by students of their degree of satisfaction.

---

[1] PERPE stands for "Perceptions Etudiantes de la Relation Professeur-Etudiants"; SPOT stands for "Students' Perceptions of Teachers".

As it became clear to us that it would not be possible to present in sufficient detail more than one of the studies made in the recent years, the present authors decided to focus on our most recent experimental study[1], specially designed to collect data on the above problem.

METHOD

The 1059 subjects in this experiment came from 52 classes in seven different french colleges. All subjects were presented two forms of a course evaluation questionnaire. There were 15 items in both questionnaires, chosen from a group of 75 items that were being tested in the framework of a forthcoming revision of the PERPE-SPOT Questionnaire. In form X, we used the derived approach described above, the only difference being in a 5-point scale anchored with appropriate qualificatives on each of the five points. In form Y, the items were slightly reworded so as to be more congruent with the direct scale used to measure the satisfaction-dissatisfaction continuum (cf. Figure 2). To check for a possible effect due to the sequence

1	2	3	4	5	6
very dissa-tisfied	dissa-tisfied	slightly dissa-tisfied	slightly satis-fied	satis-fied	very satis-fied

To what degree are you satisfied

07. ...of his ability to explain clearly difficult concepts and ideas?

Figure 2. Typical item and instructions from the experimental form using a direct measure of satisfaction.

[1] We wish to thank Miss Lucie Houde, M. René Beauséjour and Professor Richard Brunet for their close collaboration in the planning and data collection stages of this study.

of presentation, two sub-groups of 4 classes were each presented with a particular sequence. The 'item x sequence' ANOVA revealed no particular effect beyond what would be expected by chance. The sequence of presentation was varied randomly in the last 44 classes, with the same sequence for all students of a given class. Data were gathered during the 1974 spring semester.

RESULTS

Basic statistics

As a general information on the "behavior" of each item, Table 3 presents the mean and standard deviation (N = 52 classes) of the five constructs included in the analysis of results. A short description of the item contents also appears. We have placed in the right-hand column of Table 3 the percentage of the observations corresponding to excess discrepancies, that is discrepancies with a '+' sign. This datum was chosen as our index of the degree of non-monotonicity of the items. You will recall that non-monotonic items may generate excess dissatisfaction as well as deficiency dissatisfaction, since the optimally desirable frequency of the behavior measured is anywhere but at the extremes of the scale.

Reliability indices

The reliability coefficients presented in this section were all computed with the same uniform procedure. The basic formula originates from the ANOVA technique (Winer, 1962, pp. 124-132). It gives internal consistency coefficients which can be interpreted in the same manner as those obtained through split-halves on classes. Our procedure also includes a correction for variations in the number of respondents per class (Ebel, 1951, p. 413). Moreover, in order to

Table 3

Content descriptions, means and standard deviations (N=52) of five constructs for each of the 15 items, as well as the percentage of excess discrepancies

Content descriptions	MA		MB		MAB		MAL		D		% of excess discrep.
	M	SD	M	SD	M	SD	M	SD	M	SD	
01 Motivates to investigate further	3.156	.617	3.915	.338	.807	.466	-.755	.459	4.110	.692	2.0
02 Monotonous - lively	3.270	.893	4.253	.241	1.019	.778	-.991	.780	4.183	1.016	1.0
03 Workload (light - heavy)	3.269	.406	3.041	.241	.415	.274	.230	.330	4.543	.553	22.6
04 Student participation in grading	2.363	.573	3.410	.279	1.119	.483	-1.048	.497	3.650	.763	2.8
05 Particular about use of language	4.062	.452	4.077	.346	.220	.171	-.016	.195	5.199	.263	7.5
06 Speed in speech delivery	3.196	.568	3.087	.137	.515	.390	.110	.519	4.534	.658	24.4
07 Clarity of explanations	3.828	.528	4.292	.245	.518	.477	-.468	.458	4.667	.739	2.0
08 Frequency of humour	2.918	.829	3.756	.326	.877	.638	-.840	.647	4.205	.975	1.3
09 Compulsory readings	3.188	.620	2.948	.309	.561	.319	.242	.462	4.379	.559	31.3
10 Questions to verify understanding	2.854	.543	3.343	.387	.690	.364	-.495	.373	4.286	.496	6.7
11 Patience	3.882	.372	4.040	.253	.263	.248	-.155	.250	5.152	.470	4.2
12 Frequency of group-work	2.647	.879	3.324	.493	.820	.501	-.681	.555	4.071	.805	5.3
13 Careful not to hurt students.	3.897	.449	4.057	.317	.279	.316	-.156	.304	5.020	.487	4.3
14 Showmanship	2.414	.824	3.375	.482	1.016	.642	-.959	.646	4.057	.894	1.6
15 Frequency of lecturing	3.991	.533	3.559	.411	.614	.375	.432	.401	4.494	.615	34.8

facilitate comparisons between samples, all coefficients are "standardized" on 25 judges with the Spearman-Brown formula. The coefficients thus computed indicate the degree of precision with which the score of a professor on a given item can be reproduced, whatever the group of 25 student judges invited to evaluate an identical learning situation. Finally, means across items are computed via r z transformation.

The reliability coefficients in the first row of Table 4 were computed on a subsample of the 1059 subjects, randomly extracted in order to carry out an analysis of variance. The other coefficients in Table 4 bear evidence to the similarity of these coefficients as compared to those obtained from earlier data banks.

Except for MB, all coefficients compare quite well with similar indices found in the litterature (Costin, Greenough and Menges, 1971, pp. 512-513). In the present study, the D construct appeared significantly more precise than the corresponding derived construct ($t_{MAB,D} = 5.33$, $p < .01$). However, a much smaller difference would probably have been observed if a 7-point scale instead of a 5-point scale had been used in the derivation approach, as witnessed by the difference in precision between data sources H-I and H-II.

The lower coefficients for the desire measures (MB) could be due to the more subjective character of these measures and to lower between group variances (cf. Table 3). It might seem surprising that these precision coefficients were nevertheless relatively high, considering what we would expect from "subjective" measures. Our results clearly show that student desires, at least as measured by our technique, do vary systematically from one class to another. These results, and others in the same direction, have led one of us (Gagné, 1973) to adopt for the MB measures the terms "circumstancial desires", that is desires pertaining to specific learning situations.

Table 4

Precision coefficients for a direct measure of dissatisfaction
and for four constructs associated with a derived measure of dissatisfaction

Data source code	Date	Number of classes	Length of scale	Number of items	Average number of students per class	$\bar{r}_{25,25}$				
						Reality (MA)	Desires (MB)	Dissatisfaction		
								Gross (MAB)	Net (MAL)	Direct (D)
X - 2	1974	46	5 (D=6)	15	10	.92	.78	.84	.83	.88
A	1970	1196	5	61	27	.88	.68	.81	.80	-
B	1971	728	5	61	25	.90	.71	.82	.81	-
H - I	1973	161	5	53	21	.88	.70	.79	.79	-
H - II	1973	132	7	53	23.5	.91	.79	.85	.86	-

Convergence between constructs

Instructor or class level: observed correlations

As already told, MA, MB, MAB, MAL and D measures were computed for each item in each of the 52 groups, using the formulae presented in Table 1. Then, in order to answer the following question: "Which of the constructs from the discrepancy procedure is most closely related to a direct measure of the satisfaction-dissatisfaction continuum?", we computed, item by item, correlation coefficients between the D measure on one part and each of the other constructs, adding to the group the absolute value of MAL ($|MAL|$). We wanted to verify whether this construct was as good a predictor of D as MAB. If so, we could henceforth forgo the use of MAB, since MAL would serve to convey at the same time the intensity and direction of the dissatisfaction feeling. These five series of correlation coefficients appear in Table 5. In this table, the items are ordered according to their decreasing degree of non-monotonicity. Using Fisher's z transformation applied to the absolute 'r' values, means across items were separately computed for the four clearly non-monotonic items, the remaining eleven and the total group.

The mean correlations appearing on the bottom row of Table 5 confirm, for the total group of items, the superiority of MAB as a predictor of direct satisfaction ($\bar{r}$ = .86). The absolute values of MAL come second ($\bar{r}$ = .82); then, at the same level, we have MA and MAL ($\bar{r}$ = .78) and, far behind, MB with a mean correlation of .38. But, if we compare the respective average correlations for the eleven monotonic and four non-monotonic items, it becomes clear that the general superiority of MAB over the other constructs is explained in large part by a drastic lowering of MA and MAL correlations in the case of non-monotonic items. These low MA and MAL correlations were expected, since with non-monotonic items satisfaction is first an increasing, then a decreasing function of behavior frequency or intensity. As for the MALs in absolute values, their correlations,

Table 5

Observed correlations between five constructs used in the derived approach (MA, MB, MAB, MAL, |MAL|) and a direct measure of the satisfaction-dissatisfaction continuum

Content descriptions	% of excess discrep.	≈ 20 students (N = 52 classes)				
		MA	MB	MAB	MAL	|MAL|
15 Frequency of lecturing	34.8	-.385	.302	-.891	-.814	-.853
09 Compulsory readings	31.3	-.672	-.203	-.806	-.769	-.802
06 Speed in speech delivery	24.4	.091	.229	-.912	.045	-.853
03 Workload (light - heavy)	22.6	-.448	.118	-.718	-.631	-.600
Mean:		.420	.214	.847	.627	.794
05 Particular about use of language	7.5	.577	.548	-.600	.396	-.402
10 Questions to verify understanding	6.7	.712	.364	-.800	.653	-.674
12 Frequency of group-work	5.3	.860	.643	-.893	.803	-.829
13 Careful not to hurt students	4.3	.800	.359	-.804	.807	-.800
11 Patience	4.2	.811	.477	-.808	.735	-.739
04 Student participation in grading	2.8	.676	.051	-.782	.754	-.754
01 Motivates to investigate further	2.0	.918	.604	-.798	.783	-.783
07 Clarity of explanations	2.0	.931	.412	-.913	.880	-.880
14 Showmanship	1.6	.854	.234	-.907	.914	-.911
08 Frequency of humour	1.3	.932	.524	-.926	.932	-.930
02 Monotonous - lively	1.0	.956	.457	-.960	.956	-.957
Mean:		.854	.437	.860	.825	.828
Over-All Mean:		.782	.381	.857	.784	.819

regularly lower than those for MAB, bring us to the conclusion that the cancelling process of opposite sign discrepancies has had a definite effect; consequently, MAL should not be substituted for MAB to measure the intensity of the dissatisfaction feeling.

The mean correlations of the eleven monotonic items, for which excess discrepancies are necessarily rare, show that reality (MA) as well as both discrepancy measures (MAB and MAL) are equally good predictors of direct satisfaction ($\bar{r}$ = .85, .86 and .83). The closeness of MAB and MAL is quite normal, since both measures become perfectly correlated when all discrepancies have the same sign. On the other hand, the equivalence of MA and MAB as predictors of D seems to contradict the theoretical assumption underlying the use of a derivation procedure, namely that satisfaction should be better approximated by taking into account both reality and desires instead of reality alone.

To our judgment, this observed equivalence does not necessarily imply that desires are useless for predicting satisfaction nor that MA and MAB are equally valid measures of the satisfaction construct. Their respective reliability levels (cf. Table 4) rather suggest that MA owes much of its equivalent standing with MAB to its higher precision coefficients.

Instructor or class level: disattenuated correlations

To verify whether the above observations would still hold in the absence of any measurement error, the correlation coefficients presented in Table 5 were corrected for attenuation[1]. The disattenuated

[1] To correct for attenuation, we did not use the precision coefficients presented in Table 4. These coefficients, which consider as error variance systematic differences between students, would be appropriate only if the two measures (direct and derived) had come from different students. In our case, as all students in each class received both questionnaires, the correlation coefficients would be over-corrected.

(continued on page 15)

correlations appear in Table 6, together with the medians of the eleven monotonic and the fifteen items. From these last medians emerges a pattern of relationships between the PERPE-SPOT constructs and the direct measure of satisfaction that is quite similar to the observed correlations presented in Table 5. But, for the sub-group of eleven monotonic items, the previously mentionned equivalence between MA and MAB gives place to a slight superiority of MAB over MA, judging by their respective medians of .97 and .93. This difference seems to confirm the hypothesis we made earlier: MAB is related to satisfaction more intimately than reality, but since the difference is small in the case of clearly desirable items, its better validity may be masked by its lower reliability as compared to MA.

Besides allowing for a comparison between MAB and the other constructs, the data in Table 6 also make possible the estimation of the absolute, rather than relative, degree of convergence between the derived and direct measures of satisfaction. In fact, these disattenuated correlations correspond to the maximum correlations we would obtain if our measures were perfectly reliable; this would be the case for example if the instructors had been evaluated by an infinite number of students. Then, a median disattenuated correlation of .94, as computed between MAB and the direct measure, leads us to conclude that these two measures share on the average 94% of their true variance. The fact that only 6% of the true variance of MAB for a typical item is not reproducible on the direct measure confers on our derived construct a degree of convergent validity which appears, if not perfect, at least quite impressive.

So, the error variance estimates used in the present correction were obtained from the ANOVA (N=460) mentionned earlier; they correspond to that part of the within-instructors variance which is not reproducible from one method to the other (cf. Cronbach et al., 1972, p. 288).

Table 6

Same correlations as in table 5 (except for |MAL|), but corrected for reliability attenuation

Item	MA	MB	MAB	MAL
15	.43	.17	.87	.91
09	.76	.22	.91	.87
06	.07	.25	.92	.01
03	.65	.34	.78	.88
05	.84	.94	>1.00	.64
10	.91	.77	>1.00	.76
12	.96	.66	.96	.89
13	.87	.43	.89	>1.00
11	.93	.55	>1.00	.81
04	.70	.17	.84	.79
01	>1.00	.75	.89	.87
07	>1.00	.45	>1.00	.99
14	.88	.35	.98	>1.00
08	.96	.51	.94	.96
02	.99	.51	.97	.97
$Median_{11}$	.93	.51	.97	.89
$Median_{15}$	.88	.45	.94	.88

Respondent level: within groups correlations

The last series of results, presented in Table 7, are similar to the correlation coefficients given in Table 5, except that, rather than being computed with class means as observations, they were computed within each class, using individual student answers. We wanted to ascertain from these data whether the pattern of relationships between the PERPE-SPOT constructs and the direct measure still reappeared when these constructs were used to measure differences in satisfaction between students of a given instructor, instead of differences in "satisfactoriness" between instructors.

The different groups of average correlations show a pattern strikingly similar to those in previous tables. For non-monotonic items, the reality measure is clearly inferior to the absolute discrepancy measure (.16 vs .48), while the algebraic discrepancy measure occupies an intermediate position (.28). In the case of monotonic items, reality comes out on a par with the absolute discrepancy measure (.42 vs .43), but the algebraic discrepancy measure remains slightly inferior (.40) to both of them. Obviously, these correlations are lower than in Table 5, since they were computed from less precise individual answers. It would have been interesting to see if, when corrected for attenuation, these correlations keep their relative status. We would also have been able to verify with these disattenuated correlations whether we obtain with the individual measures a degree of maximum convergence as reassuring as the level obtained with class scores. Unfortunately, the design of the present experiment did not permit us to estimate the error of measurement in student answers.

Table 7

Observed mean within groups correlations (N=52) between four constructs used in the derivation procedure (A,B,A-B,|A-B|) and a direct measure of the satisfaction-dissatisfaction continuum

Item	A	B	(A - B)	(|A - B|)
	0	0	0	0
15 Frequency of lecturing	-.03	.33	.27	.42
09 Compulsory readings	.22	.06	.21	.42
06 Speed in speech delivery	-.12	.05	-.15	.52
03 Workload (light - heavy)	.24	.03	.24	.43
Mean:	.154	.120	.218	.449
05 Particular about use of language	.36	.25	.12	.18
10 Questions to verify understanding	.26	-.04	.31	.42
12 Frequency of group-work	.34	-.12	.43	.52
13 Careful not to hurt students	.47	.32	.19	.32
11 Patience	.44	.24	.28	.19
04 Student participation in grading	.21	-.10	.29	.33
01 Motivates to investigage further	.55	.20	.42	.43
07 Clarity of explanations	.50	.11	.45	.47
14 Showmanship	.42	-.05	.44	.45
08 Frequency of humour	.49	.06	.42	.50
02 Monotonous lively	.62	.18	.42	.49
Mean:	.431	.153	.347	.396
Over-all Mean:	.362	.144	.313	.411

DISCUSSION AND CONCLUSIONS

We have found three studies[1] in which one of the main objectives was the validation of a discrepancy model through its comparison with a direct measure of satisfaction.

Other studies

Let us first mention Locke (1969), who gives an extensive demonstration, based on the objectivist philosophy of Ayn Rand and Nathaniel Brendon, of the face validity of a perception-value discrepancy model. Then, as empirical proof for his thesis, he describes briefly a few studies, in two of which (Harbaugh-Farr and Mobley Studies) different aspects of work itself are evaluated through direct and discrepancy measures of satisfaction. The derivation procedure uses two questions, presented as we do after each item: "What is your job like?" and "What should it be like?". The correlations obtained appear even more positive than our within groups correlations (cf. Table 7). Indeed, while the correlations between reality and satisfaction are approximately .50 (N = 72) in both studies, the corresponding correlations for the absolute discrepancy measure exceed .70, even though the items used look, from the examples given, generally monotonic[2].

Contrary to Locke's results and ours, Wanous and Lawler (1972), as well as Levinthal (1974), obtain negative results in their attempts

---

[1] Three other studies, Imparato (1972), Sanders and Lynch (1973) and Mitchelmore (1973), while loosely related to this topic, are designed in a manner which precludes any comparison with our results.

[2] These results would be similar to ours if his items were shown to be in fact non-monotonic.

to verify similar hypotheses. On the one hand, Wanous and Lawler, with a questionnaire comprising 23 job facets, most of them clearly desirable, measure in turn: a) reality (Is Now); b) what is deemed equitable (Should be); c) what is deemed desirable (Would Like), d) importance; e) satisfaction. Their results show clearly that, on the average, the direct satisfaction measure (e) is more closely related to reality ($\bar{r}$ = .60) than to either of the two algebraic discrepancy measures, computed from the "Would Like" answers ($\bar{r}$ = .44) or the "Should Be" answers ($\bar{r}$ = .36).

As for Levinthal (1974), he administered in two classes (N = 69 and 123) a course evaluation questionnaire, whose items were mostly monotonic. For all items, he collects, in turn, reality measures, ideal measures (Should Be), as well as an evaluative judgment on a 5-point scale increasing from "1. outstanding" to "5. poor". Averaged over ten items, the correlation between the absolute discrepancies (D measures) and the evaluative judgments (E) is respectively .25 and .22 in each group, while the corresponding average correlations between reality (O) and E are .52 and .46[1].

The "begging of the question" problem

How can we explain such a large divergence between these two couples of studies? The answer which strikes us first focuses on the procedures adopted. On the one hand, Wanous and Lawler, as well as Levinthal, measure the two basic constructs (reality and desires) separately, while Locke and ourselves ask both questions simultaneously for each item. Our procedure goes one step further, since our instructions make clear to the respondents the interpretation which we will give to the discrepancy between both answers.

---

[1] Note that these two pairs of correlation coefficients may be compared "mutatis mutandis" to the average 'A-B vs D' and 'A vs D' correlations given in Table 7.

By so doing, we will probably be "accused" of "begging the question", since we inform the respondents about the derivation process we wish precisely to validate. This objection, logically unassailable, would shift the validation problem from MAB, which is now clearly validated, to MB, the construct most liable to "manipulation" so that the resulting discrepancy will be coherent with the dissatisfaction feeling. However, the above objection would loose most of its weight if it was demonstrated that the two constructs directly measured, that is reality and desires, suffer no discernable bias due to their juxtaposition or to our instructions. Now, a study by one of our colleagues (Houde, 1974), while conducted in only two classes, shows a total lack of significant differences between A answers, measured separately or as part of our derivation procedure; within the B answers, a few significant differences appear, but still not large enough to speak of a net bias. Anyway, we intend to do further research in that direction. Let us note finally that Locke's clearly more "neutral" procedure is nonetheless open to the above objection: it appears quite plausible that the simple fact of placing side by side the two questions could be sufficient to draw the respondent's attention to the "common sense" relationship which unites reality, desires and satisfaction.

Even though the "begging of the question" objection does reduce the significance of our positive results, it does not explain for all that the negative results obtained through a procedure not subject to that objection. Do Wanous and Lawler's as well as Levinthal's results reflect more precisely the validity, or rather the invalidity of a more strictly derived measure? In Levinthal's research report, we were able to look at his data much more throughly than those of Wanous and Lawler. From this analysis, we submit that his I question (ideal or desires) was probably not interpreted univocally by all students, thus seriously jeopardizing his results.

In order to illustrate our forthcoming demonstration, we present in Table 8 some of the results pertaining to the first two

Table 8

D vs E and O vs E correlations for all subjects and for sub-groups at each level of I, as obtained with items 08 and 14 in experiment OEIII (from Levinthal, 1974, pp. 44 and 48)

	Total Sample	Levels of I-score				
		1	2	3	4	5
Should generate excitement about subject being thought						
r (D,E)	.43	.62	.25	-.64	-	-
n	(121)	(74)	(38)	(7)	(2)	(0)
r (O,E)	.61	.63	.50	.65	-	-
n	(121)	(75)	(37)	(7)	(2)	(0)
Should be dynamic and outgoing						
r (D,E)	.21	.57	-.05	-.37	-	-
n	(120)	(41)	(59)	(16)	(1)	(0)
r (O, E)	.62	.57	.61	.63	-	-
n	(119)	(41)	(59)	(16)	(2)	(0)

items of his questionnaire (cf. 1974, pp. 44 and 48, experiment OEIII). The data shown are the correlation coefficients between the "direct" measure (E) and the discrepancy measure (D), as well as the O (reality) vs E correlations. They are given for all subjects and for each sub-group having chosen a different I answer. The number of respondents appears inside parentheses. Taking into account the clear desirability (or undesirability) of most of his items, we were first surprised by the large proportion of students (approximately 50% across items) who placed their "ideal", or more precisely their "Should Be" answer, anywhere but at the endpoint of the frequency scale. For instance, it seems strange that 13% of the respondents (cf. Table 8) answer that

their instructor should be (ideally) "occasionnaly dynamic and outgoing". Levinthal interprets these dispersions as real differences between student ideals (1974, p. 17).

We rather believe that they reflect divergent interpretations of the "Should Be" question. In our opinion, most of the students using point 1 (almost always) interpret the question as "Should be ideally", most of the students using point 3 (occasionnaly) interpret this same question in terms of "Should be at least", while the respondents at point 2 (often) could either use the above interpretation, or even one similar to our own B question, that is "Should be... so that I would be satisfied". As proof of the above assertion, we will use the D vs E correlations given in Table 8, which change drastically in sign and value according as they come from students using points 1, 2 or 3 of the I-scale. The negative correlations appeared, particularly mysterious to Levinthal, while they are quite simply explained with our assumptions. For instance, for those students who have used the "ideal" interpretation and consequently have probably chosen point 1, any observed answer below that point will generate negative discrepancies, that is dissatisfaction. Thus, the greater the discrepancy, the greater the dissatisfaction, which is coherent with the positive correlations observed. But, for those students using point 3 on the I-scale, which we assumed to be a "minimally acceptable" answer, reality answers placed at points 1 or 2 will mean a frequency level higher than the minimum desired; then, these I-O positive discrepancies must be interpreted as a "surplus of satisfaction"; consequently, the correlations will be negative, meaning precisely: "the larger the discrepancy, the greater the satisfaction"[1]. Finally, we can assume

[1] The change in sign but not in value between the D vs E (-.64) and O vs E (+.65) correlations for subgroup I=3 on the first item in Table 8 constitutes a mathematical proof that all discrepancies (O-I) in that particular sub-group had a positive sign.

for point 2 a probable mixing of both interpretations, which would explain the frequently low correlations observed for this point.

In our judgment, these divergent interpretations of the I-scale largely explain the low D vs E correlations computed with total sample data since, according to their sign, the absolute discrepancy measures have opposite meanings.

Even though we do not have much information to discuss the Wanous and Lawler study, because of the similarity in the procedure used, it appears vulnerable to the same kind of criticism. Consequently, from these negative results, as well as from Locke's and our own positive results, we cannot conclude definitely whether a discrepancy model allows for a valid measure, not only of satisfaction, but also of reality and desires.

Conclusion

Most of the above discussions pertaining to our data and those of our colleagues seem finally to converge on the same problem: the nature of the construct, generally called desires, which accompanies the reality measure in a derivation procedure. One thing is clear to us: desires are not to be thought of as a specific point which we are progressively trying to encircle, but as a continuum, its limits being respectively the ideal and the minimally acceptable.

Since the terms "would like", "should be" and "prefer" do not represent specific points on this continuum, but are as many synonyms to describe the continuum itself, with what terms are we left to describe specific points, other than the two extremes? Could it not be possible that there is no choice, if we wish to describe an intermediate point, but to allude more or less directly to satisfaction? In fact, are not the words "minimally acceptable", and even "ideal", related more than implicitly to the idea of satisfaction?

Going a little further, could we not even say that the desires continuum is nothing else but a satisfaction continuum, since it varies from a minimum satisfaction to an infinite degree of satisfaction? This is by and large the gist of our current reflexions?

REFERENCES

ALLAIRE, D. (1974) La fidélité de formes expérimentales en cinq et sept points du test PERPE. Rapport 0374-02, INRS-Education, Montréal.

COSTIN, Frank, GREENOUGH, William T. & MENGES, Robert J. (1971) Student ratings of college teaching: reliability, validity and usefulness. Review of Educational Research, 41, 511-535.

CRONBACH, L.J., GLESER, G.C., NANDA, H. & RAJARATNAM, N. (1972) The dependability of behavioral measurements. New-York: Wiley.

EBEL, R.L. (1951) Estimation of the reliability of ratings. Psychometrika, 16, 407-424.

GAGE, N.L., RUNKEL, P.J. & CHATTERJEE, B.B. (1960) Equilibrium theory and behavior change: an experiment in feedback from pupils to teachers. Urbana: Bureau of Educational Research, College of Education, University of Illinois.

GAGNE, F. (1973) Les mesures d'attentes du test PERPE en tant qu'indices de "désirabilité circonstanciée". Rapport 0573-16, INRS-Education, Montréal.

HOUDE, L. (1974) Etude portant sur l'interinfluence des réponses de réalité et d'attentes due possiblement à la formule du test PERPE utilisant ces deux mesures simultanément. Rapport de stage. INRS-Education, Montréal.

IMPARATO, N. (1972) Relationships between Porter's need satisfaction questionnaire and the Job Descriptive Index. Journal of Applied Psychology, 56, 397-405.

LEVINTHAL, C.F. (1974) An analysis of the teacher evaluation process. Project No. 2B089, Final Report, U.S. Department of Health, Education and Welfare, National Institute of Education, Washington, D.C..

LOCKE, E.A. (1969) What is job satisfaction? Organizational Behavior and Human Performance, 4, 309-336.

MITCHELMORE, J.A. (1973) Development of the Instructor Behavior Description Questionnaire to measure graduate student satisfaction from perceptions of ideal and actual instructor behavior. Unpublished doctoral dissertation, Ohio State University.

SAVAGE, N.L. (1958) Changes in Student Teachers through the use of pupil ratings. (Doctoral dissertation, University of Illinois) Ann Arbor, Mich.: University Microfilms, No. 58-1733.

SANDERS, J.R. & LYNCH, M. (1973) Student evaluation of instruction: the analysis of discrepancies between perceived and ideal conditions. Report no. 8, Educational Research and Evaluation Laboratory, Indiana University, Bloomington.

THOMAS, H.B. (1969) Changing teacher behavior; an experiment in feedback from students to teacher. Unpublished doctoral dissertation, Purdue University.

WANOUS, J.P. & LAWLER, E.E. (1972) Measurement and meaning of job satisfaction. Journal of Applied Psychology, 56, 95-105.

WINER, B.J. (1962) Statistical principles in experimental design. New-York: McGraw-Hill.