DOCUMENT RESUME

ED 356 249 TM 019 673

AUTHOR D'Costa, Ayres
TITLE Extending the Sato Caution Index To Define the Within and Beyond Ability Caution Indexes.
PUB DATE Apr 93
NOTE 15p.; Paper presented at the Annual Meeting of the National Council for Measurement in Education (Atlanta, GA, April 13-15, 1993).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Ability; *College Students; *Error Patterns; Factor Analysis; Higher Education; Performance; *Scores; Test Interpretation; *Test Items
IDENTIFIERS *Caution Index (Sato); Caution Indices; Testlets

ABSTRACT

The Sato Caution Index takes into account the number and difficulty of items gotten wrong by a student within his or her ability, as well as the number and difficulty of items gotten right beyond his or her ability. Sato subtracts the two components to define a single Caution Index. In this study, the components are kept separate, defining a Within Ability Concern Index (W) and a Beyond Ability Surprise Index (B). Using data from 10-item testlets taken by 121 college students, the critical information made available by using the B and W Indexes in addition to the Sato Index is identified. The relationships of the three indexes to total score and the number of errors committed by the students within ability-level are examined. Factor analysis reveals that the new indexes can add a new dimension to test performance information provided by the Sato Index. Five tables provide analysis data. (Author/SLD)

Extending the Sato Caution Index to Define the Within and Beyond Ability Caution Indexes

Abstract

The Sato Caution Index takes into account the number and difficulty of items gotten wrong by a student within his/her ability, as well as the number and difficulty of items gotten right beyond his/her ability. Sato subtracts the two components to define a single Caution Index. This study proposes to keep each component separate, defining a Within Ability Concern Index (W) and a Beyond Ability Surprise Index (B).

This study points out the critical information made available by using the B and W Indexes in addition to the Sato Index. It also examines the relationships of the three Indexes to Total Score and the Number of Errors committed by student within ability-level. A factor analysis revealed that the new Indexes can add a new dimension to test performance information provided by the Sato Index.

Key- Words: Sato Caution Index, W Index, B Index
Test performance, Diagnostic feedback.

Extending the Sato Caution Index to Define the Within and Beyond Ability Caution Indexes[1]

Ayres D'Costa
The Ohio State University

Background & Rationale

Most testing programs represent student test performance in terms of the total score earned on the test. Not taken into account is the fact that two students who get the same total score may have earned that score by getting entirely different items correct. In the current test-scoring practice there is usually no distinction made between a student who gets a score of 10 by solving the 10 most difficult items in the test, versus another student who gets the same score of 10 by solving the 10 easiest items in the test. Test items are assumed to be equally difficult.

To deal with this problem, Sato (1980) introduced the Caution Index (SCI). To understand this Index, one must view the items in a test as if Guttman-scaled and in ascending order of difficulty. Let us assume that Guttman scaling is appropriate; and, furthermore, that a student's test score identifies the upper bound of the set of items in the test which he/she should have ordinarily got right. Items at or below this bound would be considered to be <u>within</u> his/her ability level. Items beyond this bound are <u>beyond</u> his/her "true" ability level.

Sato's Caution Index (SCI) takes into account the number and difficulty of items got wrong by a student within his/her ability level, as well as the number and difficulty of items that the student got right beyond his/her ability level. Both types of test performance are presumed to be unusual,

[1] Presented at 1993 convention of National Council for Measurement in Education, Atlanta, Ga.

worthy of caution, but opposite in effect. Sato subtracts the two to define a single Caution Index.

This study began with the belief that these two types of test performances should be reported separately. When a student misses items within ability level it could be indicative of carelessness or gaps in learning. This type of performance would suggest concern to the teacher. But when items beyond ability are got right by a student, the teacher would be hard pressed to label this merely as the opposite of carelessness, or as compensatory of gaps in learning. Indeed, the attentive teacher would be surprized and would look for unrecognized skills and unsuspected learning acquisitions. This paper suggests that the seriousness of items missed within ability be indicated by a "within ability concern index", whereas the latter should be separately indicated by a "beyond ability surprise index". SCI currently combines these two types of concern, thereby becoming vulnerable to a possible washout effect.

Table 1 here

We will begin by illustrating the specific computations required for the SCI. Table 1 presents a Test Outcomes Matrix of 0^s and 1^s for 10 students who took a 20-item test. The 1^s indicate correct responses, the 0^s indicate wrong responses. The 20 items, whose serial numbers are listed in the top row of Table 1, are ordered by difficulty level (p) from easy to difficult. These p values are shown without decimal point in the second row of the Table. The students are listed in descending order of their total score. Column 1 presents Student Id #, while Column 2 presents their Raw Score on the 20-item test.

Tatsuoka & Linn (1983) presented the mathematical and psychological bases for the SCI. Harnisch and Linn (1981, p.135) introduced a modified form of the SCI which yields a lower bound of 0 and an upper bound of 1. Mathematically, the modified SCI (MCI) is defined by them as follows:

$$\frac{\sum_{j=1}^{n_{i\cdot}} (1 - u_{ij})\, n_{\cdot j} - \sum_{j=n_{i\cdot}+1}^{J} u_{ij}\, n_{\cdot j}}{\sum_{j=1}^{n_{i\cdot}} n_{\cdot j} - \sum_{j=J+1-n_{i\cdot}}^{J} n_{\cdot j}}$$

where $i = 1,2,\ldots\ldots..I$ indexes the examinee,
$j = 1,2,\ldots\ldots..J$ indexes the item,
u_{ij} = 1, if student i gets item j correct,
= 0, if student i gets item j wrong.
$n_{i\cdot}$ = total correct for student i,
$n_{\cdot j}$ = total correct responses to item j.

In simple language, this Modified Caution Index (MCI) is defined as follows:

Let p = traditional difficulty index for an item
Let t = student score on test
Let k = number of items in test
Let w = Sum of p's for items got wrong within ability level, i.e., the first t items in Table 1.
Let b = Sum of p's for items got right beyond ability level, i.e., the remaining (k - t) items.
Let H = Sum of p's for t items with highest p values (easiest items).
Let L = Sum of p's for t items with lowest p values (most difficult items).

$$\text{Then MCI} = \frac{w - b}{H - L}$$

Purpose

This paper proposes that the Within ability (Concern) factor and the Beyond ability (Surprise) factor be computed as two different types of caution that teachers could use to help their students in an instructional setting. Instead of combining them into a single Caution Index, this paper extends the Sato concept to define two new Indexes, W and B. The Index W reflects Concern, whereas B reflects Surprise. As a first step, this paper defines W and B mathematically to highlight their construct meaning, and in a manner that standardizes their measured value between 0 and 1. As a second step, we will explore the nature of the constructs associated with these three Indexes by examining their inter-relationships and their relationships with traditional test outcomes, such as Total Score and Number of Errors committed by a student within his/her ability level. Note that the latter is less than or equal to the total number or errors committed by student.

Definition for the New Caution Indexes

Each Index is defined as a ratio derived from the matrix 'U' of 0s and 1s as shown in Table 1. Traditionally, the p value for an item is the proportion of students who got the item correct, and q = 1-p.

$$\text{Let } W = \frac{\text{Sum of p's for items got wrong within ability level}}{\text{Sum of p's for all items within ability level.}} = \frac{\sum_{j=1}^{n_{i\cdot}} (1-u_{ij})n_{\cdot j}}{\sum_{j=1}^{n_{i\cdot}} n_{\cdot j}}$$

where all symbols are as defined earlier.

$$\text{Let } B = \frac{\text{Sum of q's for items got right beyond ability level}}{\text{Sum of q's for all items beyond ability level.}} = \frac{\sum_{j=i.+1}^{J} u_{ij}(1-n_{\cdot j})}{\sum_{j=n_{i}.+1.}^{J} (1-n_{\cdot j})}$$

Intuitively, W is the proportion of *p's missed from the within ability* test items, and B is the proportion of *q's achieved from the beyond ability* test items. Thus the W index can be said to measure the proportion of concern, whereas the B index measures the proportion of surprise in the student's test performance.

With reference to the simplified MCI formula: $\frac{w - b}{H - L}$,

note that while $W = \frac{w}{H}$; $B \neq \frac{b}{L}$.

Table 2 here

Table 2 presents the computed values for the three Indexes. Note the computation of the Sato MCI for Student Id # 6 and observe how this Index could dilute itself by combining the two opposing factors represented by W and B. Let us first illustrate the computation of MCI:

t = 17

w = 0.6 = 0.6

b = 0.2 = 0.2

H = Highest 17 'p' values = 11.3

L = Lowest 17 'p' values = 9.1

$$\text{Sato MCI} = \frac{0.6 - 0.2}{11.3 - 9.1} = \frac{0.4}{2.2} = 0.18$$

This MCI for Student # 6 appears quite low, given that Sato proposed 0.5 and Harnisch (1983) suggested 0.3 as the cut-off for judging 'significant' concern. Most users of the Sato MCI (Blixt & Dinero, 1985) would ignore values this low, and would thus miss out on the fact that this student has responded correctly to Item #7 which is a relatively a very difficult item. While the computation of the W Index is relatively straight-forward, note that the B Index uses 'q' values in its computation.

$$W = \frac{0.6}{11.3} = 0.05$$

$$B = \frac{0.8}{2.3} = 0.35$$

Note that of the three Indexes, the Surprise Index B is most prominent in the case of this student. The same is true for Student Id # 10 whose Sato MCI = 0.21. A somewhat different observation is made for Student #3, whose MCI computes to 0.17, and the W Index is significant.

W = 0.33

B = 0.21

These examples illustrate the fact that whereas the Sato MCI for a student may sometimes appear non-significant because of the wash-out effect; the extended indexes, W and B, may be interesting for a teacher to follow up.

Methodology and Data for Construct Analysis

The second segment of this paper explores the nature of these three Indexes using real data from 10-item testlets taken by various groups of students at a large midwestern state university. The three Indexes were computed using a special FORTRAN 77 program written by the author. Five variables were considered in this study: Total Score, Number of Errors committed by student within ability level, and the three Caution Indexes. All analyses were conducted using MINITAB Version 8.2.

Table 3 here

Results

Table 3 presents a comparison of educational decisions made with the Sato MCI in relation to decisions that could be made if the B Index and the W Index were also available. For 121 students tested, the Sato MCI identified 35 students as Marginally Significant (MCI between 0.26 and 0.45) and only 3 students as Significant (MCI above 0.45). However, of the 83 Non-Significant decisions made with the Sato MCI, 21 could be Marginally Significant and 44 could be Significant if the B Index were used instead. None of the Significant decisions made by Sato are ignored by the B Index, thus showing itself to be a sensitive indicator.

The W Index, on the other hand, is a very conservative index, with generally low values. However, this index appears to be sensitive (or large) when the Total Score is low, whereas the reverse is true of the B Index. Thus, it appears that the W and B Indexes provide helpful diagnostic information when the Sato Index is washed out and shows up as Non-Significant, especially when the Total Score is very high or very low.

Table 4 here

Table 4 presents the inter-correlations among the five variables of interest in this study. These correlational analyses were conducted in order to explore the nature of the constructs associated with the three Indexes. The inter-correlations matrix presented is from two of several groups that were analysed. Note that the upper-right and lower-left triangles of the matrix present correlations from two different groups. The following observations can be made from Table 4, with appropriate reservations given their possible sample dependency:

1) As expected, Test Score is negatively correlated with Number of Errors within ability.

2) The W and B Indexes are each correlated with the Sato MCI. However, W and B are negatively correlated with one another; and understandably, B is negatively correlated with the Sato MCI. Recall that the B component is subtracted from the W component in the MCI formula.

3) Harnisch & Linn (1981) reported that the Sato MCI has a low and sometimes negative relationship with Total Score. This finding was not replicated in this study. Jaeger (1988) reports a similar departure for his data. Note that the W Index is very strongly and negatively correlated with Total Score, whereas the B Index is positively correlated with the Total Score.

A further analysis of the two sets of data revealed that extremely high or extremely low Total Scores result in very high variations in all three

Indexes. This has serious implications for the interpretation of these Indexes and mandates that a valid interpretation include a combination of the three Indexes and the Total Score. As indicated earlier, the W Index is important in the case of low Total Scores, whereas the B Index is important in high Total Scores. The Sato MCI, by itself, would suffer from confounded interpretation when the Total Score has extreme values. For these reasons, the three Indexes, taken together, provide better diagnostic capability than any one of them.

4) All three Indexes correlate positively with the Number of Errors. The W Index shows the best one-to-one correspondence with the Number of Errors, the other two showing very high variations.

Table 5 here

Table 5 presents the results of a principal components factor analysis for the 5 variables with replication across the two groups utilized in this study. Two factors are suggested. Factor 1 is bipolar, with Number of Errors and the W Index at one end, and the Total Score at the other. Factor 2 is defined by the B Index and MCI. This is interesting because, of the three Indexes, the W Index shows up as reflecting an unique type of variance, whereas B and MCI seem to go together. Recall that in this limited study the overall range of values attained by the W Index was somewhat low.

Implications

The use of the Sato Caution Index in educational practice was pioneered by Harnisch (1983) who presented a special computer program to educators interested in this application. The Harnisch computer program provides an elegant printout of the S-P Table, encourages the plotting of the S and P curves by teachers applying the Sato Technique, and facilitates the analysis of student test performance as well as problem (item) performance.

The computer package developed for this study[2] also presents the S-P Table. It outputs the three Indexes for up to 100 Students and 100 Problems. Both Students and Test Problems (Items) are analyzed.

Dinero and Blixt (1990) have presented the Sato Caution Index in lay terms and demonstrated its utility for teachers helping their students on the basis of test-performance. The new W and B Indexes have relevance to classroom instruction and should be used in conjunction with the Sato Index. Students scoring high on the W-Index should be helped to work with greater care, and to deal with a possible test anxiety problem. High W's should be studied further to explore the influence of specific item-content or item-format on test performance.

Students scoring high on the B-Index should be studied in terms of past instructional history or some unusual learning strategies. Difficult items attained by such students should also be examined for possible links to special areas of scholarly interest or intellectual potential. Cross-tabulations of the three Indexes with Total Score, similar to the cross-tabulation suggested by Harnisch (1983) of the Caution Index vs Total Score is also suggested.

The phrase "extended caution indices" was introduced by Tatsuoka & Linn (1983) and utilized somewhat differently when they linked item-response-theory (IRT) and Sato Caution Index approaches to identifying unusual response patterns. No attempt has been made in this study to relate the new B and W

[2] Readers interested in a source copy of the FORTRAN 77 program should contact the author at 1945 N. High Street, Columbus, Ohio 43210 or call 614/292-3239.

Indexes to IRT approaches, and specifically to what are known in IRT literature as "fit" statistics.

There is need to examine the distribution characteristics of the three Indexes to render the interpretive task more meaningful and consistent. Of special concern is the aberrant behavior of these indexes in situations where the Total Score is very low or very high. A Monte Carlo study of this phenomenon is indicated.

References

Blixt, S. L. & Dinero, T. (1985). An initial look at the validity of diagnoses based on Sato's Caution Index. Educational & Psychological Measurement, 45, 293-99.

Dinero, T., & Blixt, S. L. (1990). Deriving information from teacher-made tests using Sato's S-P chart. The College Teacher.

Harnisch, D., & Linn, R. L. (1981). Analysis of item response patterns: Questionable test data and dissimilar curriculum practices. Journal of Educational Measurement, 18, 3, 133-146.

Harnisch, D. (1983). Item response patterns: Applications for educational practice." Journal of Educational Measurement, 20, 2, 191-206.

Jaeger, R. M. (1988). Use and effect of caution indices in detecting aberrant patterns of standard-setting judgments. Applied Measurement in Education, 1,1, 17-31.

Sato, T. (1980). The S-P Chart and the caution index, Tokyo, Japan: NEC Educational Information Bulletin, 80-1, C & C Systems Research Laboratories, Nippon Electric Co., Ltd.

Tatsuoka, K.K. & Linn, R.L. (1983). Indices for detecting unusual patterns: Links between two general approaches and potential applications, Applied Psychological Measurement, 7,1, 81-96.

Table 1. TEST OUTCOMES MATRIX
(10 Students X 20 Problems)

Id#	S / Item#	12	13	15	9	18	19	6	1	8	17	2	3	20	5	16	10	4	11	7	14
	p	100	100	90	80	80	80	70	60	60	60	60	60	60	50	50	40	30	30	20	20
6	17	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	0
1	16	1	1	1	1	0	0	1	1	1	0	1	1	1	1	1	1	1	0	1	1
10	14	1	1	1	1	1	1	1	1	0	1	0	0	1	1	1	0	1	1	0	0
2	12	1	1	0	1	1	1	1	0	1	1	1	1	0	1	1	0	0	0	0	0
9	12	1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	0	0	0	0	0
5	12	1	1	1	1	1	1	1	0	1	1	1	1	0	1	0	0	0	0	0	0
4	10	1	1	1	0	0	1	0	1	1	0	1	1	0	0	0	1	0	0	0	1
7	10	1	1	1	1	1	1	0	1	0	0	0	0	1	0	1	0	0	1	0	0
8	9	1	1	1	1	1	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0
3	8	1	1	1	0	1	0	1	0	1	1	0	0	0	0	0	1	0	0	0	0

Table 2. Comparison of the Three Caution Indexes
for Students arranged by Score

Id#	Score	#Err*	W	B	Sato
6	17	1	.05	.35	.18
1	16	3	.20	.77	.56
10	14	3	.18	.46	.21
2	12	2	.17	.20	.14
9	12	1	.07	.08	.00
5	12	1	.07	.10	.03
4	10	4	.37	.38	.31
7	10	3	.24	.28	.14
8	9	2	.18	.18	.11
3	8	3	.33	.21	.17

*Number of Errors Within Bound

Table 3: Comparison of Decisions using the three Indexes

	W Index			B Index			
Sato MCI	NS	?	S	NS	?	S	Total
Not Sig (NS)	81	1	1	18	21	44	83
Marginal (?)	31	4	0	0	1	34	35
Sig (S)	3	0	0	0	0	3	3
Totals	115	5	1	18	22	81	121

Table 4: Intercorrelations Matrix

		(Group Ht2: N=122)				
		Score	Errors	W Index	B Index	Sato MCI
	Score		-.64	-.84	.43	.03
	Errors	-.62		.81	.06	.57
Group Ht1	W Index	-.77	.79		-.14	.28
(N=177)	B Index	.44	.44	-.01		.66
	Sato MCI	.38	.11	-.04	.87	

Table 5: Principal Components Factor Analysis

Group Ht1					
Eigenvalue	2.60	1.89	0.23	0.16	0.12
Proportion	0.52	0.38	0.05	0.03	0.02
Cumulative	0.52	0.90	0.95	0.98	1.00
Variable	F I	F II	FIII	FIV	F V
Total Score	-0.58	0.02	0.45	0.68	0.00
# Errors	0.44	-0.44	0.73	-0.09	0.28
W Index	0.53	-0.31	-0.31	0.67	-0.30
B Index	-0.32	-0.59	0.07	-0.30	-0.68
Sato MCI	-0.29	-0.60	-0.42	0.04	0.61

Group Ht2					
Eigenvalue	2.70	1.83	0.26	0.11	0.10
Proportion	0.54	0.37	0.05	0.02	0.02
Cumulative	0.54	0.91	0.96	0.98	1.00
Variable	F I	F II	FIII	FIV	F V
Total Score	0.53	0.32	-0.27	0.40	0.63
# Errors	-0.57	0.15	-0.33	-0.48	0.56
W Index	-0.58	-0.07	0.37	0.69	0.24
B Index	0.06	0.69	0.68	-0.26	0.05
Sato MCI	-0.25	0.63	-0.48	0.27	-0.48