DOCUMENT RESUME

ED 356 235 TM 019 644

AUTHOR	Campbell, Kathleen T.
TITLE	Establishing Internal Consistency Reliability of Measurement Data of a New Instrument, the Information Preference Questionnaire.
PUB DATE	Jan 93
NOTE	19p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Austin, TX, January 28-30, 1993).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Assistant Principals; *Cognitive Style; Comparative Testing; Decision Making; Elementary Secondary Education; Estimation (Mathematics); Information Needs; Measurement Techniques; *Principals; Public Schools; *Questionnaires; Test Construction; *Test Reliability
IDENTIFIERS	Alpha Coefficient; *Information Preference Questionnaire; *Internal Consistency; Myers Briggs Type Indicator; Self Report Measures

ABSTRACT

The estimation of the internal consistency of a newly constructed instrument using Cronbach's coefficient alpha is demonstrated. This study sought to determine the relationship, if any, between an individual's cognitive style and the preference for certain kinds of information used in making decisions. Cognitive style was determined with the Myers Briggs Type Indicator (MBTI) and a self-report questionnaire, the Information Preference Questionnaire (IPQ) designed to measure information preference. The IPQ contains 12 scenarios of school-related decisions. The prototype IPQ was field tested with four educators in a study that provided evidence of construct validity. The MBTI and the IPQ were administered to 53 principals and 11 assistant principals from a large metropolitan public school system in the South. The internal consistency reliability of the IPQ's four scales and the test as a whole was determined through the application of coefficient alpha. The IPQ measure was internally consistent, with a reliability estimate of 0.91 and measurement of all 4 scales was reasonable. Six tables present study data, and an appendix lists the Statistical Package for the Social Sciences commands for the reliability estimates. (Contains 21 references.) (SLD)

Establishing Internal Consistency Reliability of Measurement Data of a New Instrument, the Information Preference Questionnaire

Kathleen T. Campbell, Ph.D.

Louisiana State University in Shreveport 71115

Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, TX, January 30, 1993.

Abstract

The paper illustrates how the internal consistency of a newly constructed instrument is estimated, using Cronbach's coefficient alpha. Reliability estimates for the test and its subscales are explained and interpreted. A small data set from a larger study is used to make the discussion concrete, and SPSSX commands are given.

Reliability and validity of instruments are two major aspects of measurement integrity. "If one does not know the reliability and validity of one's data little faith can be put in the results obtained and the conclusions drawn from the results" (Kerlinger, 1986, p. 404). Reliability refers to the degree to which an instrument is consistent and accurate in its measurement, and validity refers to the degree to which an instrument measures what it purports to measure. Thus, generally speaking, "reliability means stability, predictability, dependability, consistency" (Kerlinger, 1979, p. 132), while validity is concerned with content, prediction, and constructs.

Although validity is the more important of the two concepts, an instrument cannot be valid unless it is also reliable, although a test can be reliable but not valid. According to Gay (1992), if a test is valid, it will measure what it purports to measure, and it will do so every time, and thus be reliable; however, a reliable test might consistently measure the wrong thing and thus be invalid. As illustrated by Nunnally (1970), if one were to use the weight of individuals to predict their college grades, the instrument used to measure their weight could be very accurate and consistent and, thus, reliable, but still not be a valid measurement of college grades.

The degree to which an instrument is reliable and valid determines the degree to which the research results will be valuable. Reliability, while not the most important aspect in measuring variables, "is an indispensable feature of measurement" (Kerlinger, 1979, p. 138). Cronbach (1951) asserts that "[e]ven those investigators who regard reliability as a pale shadow of the more vital matter of validity cannot avoid considering the reliability of their measures" (p. 297).

> To be interpretable, a test must be reliable.... High reliability is no guarantee of good scientific results, but there can be no good scientific results without reliability. In brief, reliability is a necessary but not sufficient condition of the value of research results and their interpretation. (Kerlinger, 1986, p. 415)

According to Nunnally (1970), both reliability and validity are concerned with generalizability. Reliability concerns the extent to which one can generalize the results obtained from the application of a measurement instrument to one set of persons in one situation at one point in time to the application of the same or a similar instrument to the same set of persons in a similar situation in another point in time (Nunnally, 1970). Thus, reliability implies repeatability.

Although the importance of reliability and validity in measurement instruments has been reiterated by researchers (Kerlinger, 1979, 1986; Nunnally, 1970; Eason & Daniel, 1989; Meier & Davis, 1990; Willson, 1980), many published studies do not report the reliability and validity estimates. Willson (1980) reported that half of the articles he examined in the American Educational Research Journal did not report reliability information, while

Meier & Davis (1990) reported that the majority of psychological test scales described in the Journal of Counseling Psychology volumes did not include descriptions of psychometric properties. LaGaccia (1991) likewise found that dissertation studies frequently omit reliability estimates. Nevertheless, reliability is considered so important by Willson (1980) that he suggests journal editors reject articles which do not report reliability data. Meier & Davis (1990) recommend that researchers include a numeric estimate of reliability as the minimum information required, no matter how few or how many test items. Eason and Daniel (1989) consider reliability to be a function of a given data set rather than of test items alone, and they, accordingly, advise researchers to compute reliability estimates for each administration of a test rather than merely reporting reliability estimates for a past administration.

All reliability estimates are based on the classical true score model, $X = T + E$; that is, one's observed score (X) is equal to one's true score (T) plus the random error component. The closer one's observed score is to one's true score, the less error there is, and the more reliable the measurement is. "The reliability of a particular test can be defined as the squared correlation between X and T" (Algina, 1989, p. 141).

There are several methods used to determine reliability of instruments: test-retest, split-half, alternate forms, and internal consistency. Test-retest methods involve giving the same test to the same group of subjects after an interval of time and then correlating the results. Split-half methods involve administering a test to a group of subjects, randomly splitting the test in half, and correlating the two halves after applying the Spearman-Brown formula. Alternate forms methods involve administering parallel or equivalent forms of a test to the same group of subjects and correlating the results. Internal consistency involves item correlation of a test with all other items and with the test as a whole.

In an historical overview of the development of reliability estimates, Cronbach (1951) contended that, while the preferred manner of determining the accuracy of measurement is to compare two independent measurements (test-retest), it is impractical for educators and psychologists to recapture their subjects for a second test. For that reason split-half methods were devised. The split-half approach, however, has been criticized on the grounds that split-half coefficients do not deliver the same information as the correlation between two forms given at different times. According to Cronbach (1951), however, the test-retest method actually indicates how stable results are over time and thus produces a coefficient of stability, while the split-half method and the alternate forms method indicate how equivalent two forms measure the same trait and thus produce coefficients of equivalence.

Because the split-half and alternate forms methods fail to produce a single coefficient for the test, Kuder-Richardson developed a formula which computed item total correlation and produced a coefficient of internal consistency. Cronbach (1951) showed mathematically how his alpha coefficient formula subsumes the Kuder-Richardson 20 formula as a special case and how alpha is actually the mean of all possible split-half coefficients. Alpha estimates the proportion of the test variance due to all common factors among the items and reports how much of the test score depends upon general and group factors (Cronbach, 1951).

Although all types of reliability are important, Nunnally (1970) asserts that item correlation methods are superior to methods based on subdividing a test. There are many different ways to subdivide a test, each of which may produce an appreciably different reliability estimate; therefore, methods based on item correlation, which actually estimate the average of all the reliability estimates from all possible subdivisions of a test, are superior (Nunnally, 1970). Since the alpha coefficient of internal consistency determines how each item on the test relates to all other items on the test and to the test as a whole (Gay, 1992), it is a good index of how well the test is put together. According to Cronbach (1951), "If a test has substantial internal consistency, it is psychologically interpretable" (p. 320).

According to Eason (1991), however, reliability is not a characteristic of tests but "a characteristic of data, albeit data generated on a given measure administered with a given protocol to given subjects on given occasions" (p. 84). Sax (1980) reinforces this concept by asserting that it is more accurate to refer to "reliability of *measurements* (data, scores, and observations) than the reliability of *tests* (questions, items, and other tasks)" (p. 261, emphasis in the original). Furthermore, Sax (1980) advises that all references to the reliability of tests should be interpreted to mean the reliability of data obtained from a test.

Accordingly, although the stated purpose of the present paper is to illustrate how Cronbach's alpha was used to determine the internal consistency reliability of a newly constructed instrument, the actual purpose is to determine the internal consistency reliability of the measurement data derived from the application of the new instrument in the present study. All references to the reliability of the IPQ and its subscales should be interpreted as the reliability of the measurement data of the IPQ and its subscales.

The present study sought to determine the relationship, if any, which exists between an individual's cognitive style and the preference for certain kinds of information used in making decisions. In order to determine cognitive style, the Myers-Briggs Type Indicator (Briggs & Myers, 1976) (MBTI) was used. Since there was no instrument available to measure information preference, the author designed a self-report questionnaire, the Information Preference Questionnaire (Campbell, 1990) (IPQ).

The MBTI is grounded in Jung's (1923) theory of psychological type from which is derived cognitive style, the classification predictor used in the present study. Cognitive style refers to two dimensions: the manner in which individuals perceive information

(Sensing or Intuiting) and the manner in which individuals make judgments (Thinking or Feeling). The resultant combinations of the Sensing-Intuiting and the Thinking-Feeling dimensions yield four cognitive styles: Sensing Thinking (ST), Sensing Feeling (SF), Intuitive Thinking (NT), and Intuitive Feeling (NF). Frisbie (1988) asserts that the "core of the Jung/MBTI approach is the functions, which are reflections of cognitive processes. Behaviors characteristic of these functions represent cognitive styles... SF, ST, NF, and NT" (p. 17).

The IPQ was designed to measure subjects' self-reported preferences for certain types of information constructed to correspond to the four cognitive styles. The IPQ contains 12 scenarios concerning school related decisions. The scenarios are followed by four choices of information, each of which was rated by subjects on a blank Likert scale for its relative importance in helping them come to a decision. The four information choices (ST, SF, NT, NF) were arranged in a randomly varied order from one scenario to the next so that respondents would not be aware that they might be establishing a pattern by rating choice A (or B or C or D) consistently higher than the other information choices.

The IPQ scenarios, each of which concerns a different school related decision, were composed after interviewing five school principals of various levels (elementary, middle, high school) regarding decisions which they typically encounter throughout the school year. The information choices were constructed after consulting a wealth of literature describing the Jungian cognitive styles (ST, SF, NT, NF) and the pure functions (S, N, T, F).

Face and content validity of the school related decisions were determined by consulting several experts in school administration, while face validity of the information choices was determined by consulting several other individuals who were knowledgeable of the Jung/MBTI theory. Content validity of the information choices was ensured by a thorough and detailed analysis of each word and phrase of the IPQ information choices by the author.

The prototype of the IPQ was field tested on a group of four educators whose cognitive styles had been assessed by a previous administration of the MBTI unrelated to the present study. Subsequent discussion with the group led to further revision of the information choices. The revised IPQ was pilot tested on two graduate classes in educational administration, the former a group of master's candidates and the latter a group of doctoral students. Discussions with respondents from both groups led to additional rewording of the information choices. A principal components factor analysis of the two pilot test IPQ responses established that pure functions and cognitive styles clearly clustered onto certain factors, thus providing evidence of the construct validity of the IPQ.

Finally the MBTI and the IPQ were administered to a group of 64 administrators (53 school principals, 11 assisant principals). Subjects came from a large metropolitan area public school system in the South. The independent variables were the subjects' scores on the Sensing, Intuiting, Thinking, and Feeling scales of the MBTI, while the dependent variables were their scores on each of the four cognitive style scales of the IPQ. A canonical correlation analysis was performed on the data.

The internal consistency reliability for the four scales (ST, SF, NT, NF) as well as for the test as a whole was determined through the application of coefficient alpha, which is recommended by Cronbach (1951) as one of the most useful estimates. The more general formula for alpha is given by Cronbach (1951) for item consistency (p. 299) as well as for subtest or subscale consistency (p. 321). Each IPQ scale (ST, SF, NT, NF) was considered as a separate subtest or subscale in the analysis. The SPSSX commands are reported in Appendix A. Reliability coefficients for each scale of the IPQ and for the entire test is reported in Table 1. Reliability estimates are reported for the group of 53 principals as well as for the entire group of 64 administrators (principals and assistant principals). Guided by Nunnally's (1967) assertion that .70 or higher can be considered an accepable alpha reliability estimate for testing hypothesized constructs, the researcher concluded that the IPQ measurement was internally consistent with a reliability estimate of .91, and that measurement of all four scales was reasonable, ranging from .72-.78. Since reliability is a function of test length, the reliability estimates of subscales within a test should be lower than the estimates for the entire test (Gay, 1992). Item-total corelations, delivered in the output, also attested to the reliability of the test items in general.

Since, under most circumstances coefficient alpha is actually a "lower bound or floor estimate of reliability" (Crocker & Algina, 1986, p. 142), then it may be concluded that the reliability estimates obtained in the present study are only modest estimates of internal consistency of the measurement data of the IPQ.

Crocker & Algina (1986) summarized the qualities of coefficient alpha. (1) Coefficient alpha is an index of internal consistency, but it implies nothing about the stability of test scores over time or over alternate forms. (2) Coefficient alpha is not a direct estimate of the reliability coefficient but can be considered as the lower bound of this theoretical coefficient. (3) Alpha is the mean of all possible split-half coefficients. (4) A high coefficient alpha does not indicate that the test items can be explained by a single underlying factor. "For a test to be interpretable, however, it is not essential that all items be factorially similar" (Cronbach, 1951, p. 320). (5) Coefficient alpha is generally applicable to any situation where the reliability of a composite is estimated, such as the estimation of reliability of a total score based on subtests or subscales.

As Cronbach (1951) contends, "A reliability coefficient demonstrates whether the test designer was correct in expecting a certain collection of items to yield interpretable statements about individual differences" (p. 297). Based on the internal consistency of the IPQ data as assessed by alpha, the designer concludes that the IPQ does yield interpretable statements about individual differences.

According to Thompson (1989), reliability can be subsumed under the umbrella of generalizability. Algina (1989) asserted that generalizability theory is actually a generalization of reliability theory and illustrated mathematically how reliability coefficients could be derived from generalizability coefficients. Eason (1991) concurred and explained logically "how generalizability theory subsumes all other reliability estimates as special cases" (p. 84). Shavelson and Webb (1991), moreover, attest to the superiority of generalizability theory over rerliability theory. Nevertheless, reliability is an important characteristic of the measurement of data, and reliability estimates will undoubtedly be expected and required by the research community. Thompson (1991), furthermore, asserts that "If reliability as a quality inures to data rather than to tests, then it may be incumbent upon researchers to confirm that a protocol is yielding meaningful data in any given application" (p. 1071).

References

Algina, L. (1989). Elements of classical reliability theory and generalizability theory. In B. Thompson (Ed.), Advances in social science methodology. Greenwich, CT: JAI Press.

Briggs K.C., & Myers, I.B. (1976). Myers-Briggs Type Indicator: Form G. Palo Alo, CA: Consulting Psychologists Press.

Campbell, K.T. (1990). The influence of Jungian cognitive style on information preferences among school principals. (Doctoral dissertation, University of New Orleans, 1991).

Crocker L., & Algina J. (1986). Introduction to classical and modern test theory. Chicago: Holt, Rinehart and Winston.

Cronbach L.J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16, 297-334.

Eason, S.H. (1991). Why generalizability theory yields better results than classical test theory: A primer with concrete examples. In B. Thompson (Ed.), Advances in educational research: Substantive findings, methodological developments (Vol. 1, pp. 83-98). Greenwich, CT: JAI Press.

Eason, S.H., & Daniel, L.G. (1989). Trends and mehodological practices in several cohorts of dissertations. Paper presented at the annual meeting of the Southwest Educational Research Association, Houston, TX. (ERIC Document Reproduction Service No. ED 306 299)

Frisbie, G.R. (1988). Cognitive style: An alternative to Keirsey's temperaments. Journal of Psychological Type, 16, 13-21.

Gay, L.R. (1992). Educational research: Competencies for analysis and application (4th ed.). New York: Merrill Publishing Co.

Jung, C.G. (1923). Psychological types. In H.Read, G. Adler, & W. McGuire (Eds.), The collected works of C.J. Jung, 6, (pp. 3-493). Princeton, NJ: Princeton University Press.

Kerlinger, F.N. (1986). Foundations of behavioral research (3rd ed.). New York: Holt, Rinehart and Winston.

Kerlinger, F.N. (1979). Behavioral research: A conceptual approach. New York: Holt, Rinehart and Winston.

LaGaccia, S. (1991). Methodology choices in a cohort of education dissertations. In B. Thompson (Ed.), Advances in educational research: Substantive findings, methodological developments (Vol. 1, pp. 149-158). Greenwich, CT: JAI Press.

Meier S.T. , and Davis, S.R. (1990). Trends in reporting properties of scales used in counseling psychology research. Journal of Counseling Psychology, 37, 113-115.

Nunnally, J.C. (1970). Introduction to psychological measurement. New York: McGraw-Hill.

Nunnally, J.C. (1967). Psychometric theory. New York: McGraw-Hill.

Sax, G. (1980). Principles of educational and psychological measurement and evaluation (2nd ed.). Belmont, CA: Wadsworth.

Shavelson, R.J., and Webb, N.M. (1991). Generalizability theory: A primer. Newbury Park, CA: Sage.

Thompson, B. (1991). Review of Generalizability theory: A primer by R.J. Shavelson and N.M. Webb. Educational and Psychological Measurement, 51, 1069-1075.

Thompson, B. (1989). Why generalizability coefficients are an essential aspect of reliability assessment. Paper presented at the annual meeting of the Southwest Educational Research Association, Houston, TX.

Willson, V. (1980). Statistical techniques in AERJ articles: 1969 to 1978. Educational Researcher, 9(6), 5-10.

Table 6

Alpha Coefficients for Information Scales of the IPQ

	Total Group (N = 64)	Principals (N = 53)
Total Test	.90	.91
Scale ST	.79	.78
Scale SF	.72	.72
Scale NT	.76	.77
Scale NF	.73	.76

APPENDIX A

SPSS-X COMMANDS FOR ALPHA COEFFICIENT RELIABILITY ESTIMATES

```
TITLE 'RELIABILITY ANALYSIS FOR IPQ'
FILE HANDLE ABC/NAME='IPQ.DAT'
DATA LIST FILE=ABC RECORDS=3/OCCUP 4-5 (A) ST01 TO ST12 SF01 TO
  SF12 NT01 TO NT12 NF01 TO NF12 (24F3.0/2X,24F3.0/)
LIST CASES=5000/VARIABLES=OCCUP ST01 TO S12/FORMAT=NUMBERED
LIST CASES=5000/VARIABLES=OCCUP SF01 TO SF12/FORMAT=NUMBERED
LIST CASES=5000/VARIABLES=OCCUP NT01 TO NT12/FORMAT=NUMBERED
LIST CASES=5000/VARIABLES=OCCUP NF01 TO NF12/FORMAT=NUMBERED
SUBTITLE '***** TOTAL SAMPLE'
RELIABILITY VARIABLES=ST01 TO NF12/
  SCALE(ST)=ST01 TO ST12/SCALE(SF)=SF01 TO SF12/
  SCALE(NT)=NT01 TO NT12/SCALE(NF)=NF01 TO NF12/
  SCALE(TOTAL)=ST01 TO NF12/STATISTICS=DESCRIPTIVE/SUMMARY=TOTAL
SUBTITLE '***** PRINCIPALS ONLY'
SELECT IF (OCCUP EQ 'PR')
RELIABILITY VARIABLES=ST01 TO NF12/
  SCALE(ST)=ST01 TO ST12/SCALE(SF)=SF01 TO SF12/
  SCALE(NT)=NT01 TO NT12/SCALE(NF)=NF01 TO NF12/
  SCALE(TOTAL)=ST01 TO NF12/STATISTICS=DESCRIPTIVE/SUMMARY=TOTAL
```