AUTHOR	Wittmann, Werner W.
TITLE	The Synthesis of Cattell's BDRM, Cronbach et al. Generalizability Theory and Brunswik's Lens Model. A Framework for Improving Construct and Predictive Validity.
PUB DATE	Apr 86
NOTE	23p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, San Francisco, CA, April 16-20, 1986). Contains small print.
PUB TYPE	Speeches/Conference Papers (150) -- Information Analyses (070)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Construct Validity; Educational Research; Elementary Secondary Education; Foreign Countries; Generalizability Theory; Intelligence; *Mathematical Models; Multivariate Analysis; *Predictive Validity; Psychological Studies; *Psychometrics; *Reliability; Researchers
IDENTIFIERS	Berlin Model of Intelligence; Brunswik (Egon); Cattell (Raymond B); Cronbach (Lee J); West Germany

ABSTRACT

Psychometric theory has been one of psychology's stronger foundations and a major contributor to the recognition of psychology as a scientific discipline. Basic principles of psychometric theory led to the development of respected intelligence tests and large and comprehensive testing and assessment programs. This paper synthesizes major developments in psychometric theory, focusing on the issues of reliability and validity. The synthesis is developed from: (1) Cattell's basic data relation matrix (BDRM) of 10 coordinates for describing psychological events and the reliability and generalizability (Cronbach) coefficients derived from it; (2) the four box conception of data (predictor, criterion, experimental treatment, and nonexperimental treatment) to depict psychological research problems; and (3) a hierachical version of Brunswik's lens model in which correlations between boxes can only be optimal under conditions of good reliability and symmetry. The paper concludes with an empirical example from intelligence and school achievement research: predicting school grades and other kinds of aggregated criteria from the Berlin model of intelligence, the mcst prominent intelligence model in German-speaking Europe. (BS)

THE SYNTHESIS OF CATTELL'S BDRM, CRONBACH ET AL. GENERALIZABILITY THEORY AND BRUNSWIK'S LENS MODEL. A FRAMEWORK FOR IMPROVING CONSTRUCT AND PREDICTIVE VALIDITY [1]

Werner W. Wittmann
University of Freiburg
Federal Republic of Germany

Psychometric theory has been one of psychology's stronger foundations and has contributed a lot toward recognizing psychology as a scientific discipline. Beginning with the London School (Spearman , 1904, see Stanley 1971 or Eysenck, 1981) basic principles of psychometric theory led to the development of respected intelligence tests and gave way to large and comprehensive testing and assessment programs especially in the United States. Reliability and validity issues (Cronbach, 1984a) attracted an uncounted magnitude of researchers and led to a myriad of publications.

Yet in the last decade psychometric theory was the focus of harsh criticism (Lumsden, 1976). Weiss & Davison (1981) stated:

> "Somewhere during the three-quarter century history of CTT (classical test theory, inserted by the author) the real purpose of reliability estimation seems to have been lost. Reliability coefficients in and of themselves have little utility for practical situations except for comparing their magnitudes in order to justify the use of measuring instruments." (p.633)

Validity, especially construct validity also was often labeled as "confusing" (Cronbach, 1985).

In every science there is a time for analysis and a time for synthesis. In psychometric theory time is ripe for synthesis.

A comprehensive theory should synthesize most of the major developments up to the present. It is not my intention to deemphasize contributions by eminent researchers by not mentioning them. But in my eyes, when talking about reli-

(1): Paper delivered at San Francisco, CA. AERA ANNUAL MEETING,1986, April 16-20: Invited Symposium. New Developments in Generalizability Theory. International Perspectives. (Chair: Noreen Webb).

ability and validity, the pillars of the psychometric mansion are Cattell's (1966) basic data relation matrix (BDRM), Cronbach et al. (1972), generalizability theory and Brunswik's (1956) representative design.

If we try to understand change and validity better than through paradoxes, take Cronbach's (1957, 1975) complaints about the two disciplines of scientific psychology seriously, and really want to go "beyond the two disciplines of scientific psychology" we have to integrate another pillar of the research methodology edifice, namely that erected by the so-called Northwestern School (Campbell, Cook, Boruch, see Glass, 1983).

BDRM AND GENERALIZABILITY THEORY.

Cattell (1966, p.78) claimed that every psychological event is completely described by embedding it in the 10-dimensional BDRM. These 10 coordinates are: (1) person, (2) stimulus, (3) response, (4) situation-occasion, (5) observer, (6) states of the person, (7) variants of stimuli, (8) styles of response,(9) phases of the environmental background, (10) conditions or states of the observer. Cronbach (1984 b) in honoring Cattell acknowledged the influence of the BDRM in developing generalizability theory. In Cronbach et al's. (1972) terminology the coordinates are facets of generalization which are partitioned according to ANOVA principles. Cattell (1966), using for reasons of simplicity, only three coordinates also partitions the data box, thus demonstrating different factor analytic techniques. Fig. 1 shows the "unfolding", comparable to a computer-printout according to a three dimensional data box consisting of coordinates, persons, variables and occasions.

Insert Fig. 1 about here.

Such a data box contains only one observation per cube cell, but is easily conceivable as a replicated data box, containing more than one observations per cube cell. One can always argue how many coordinates are sufficient for a concrete research problem. I found it most essential that every researcher have at least a hunch about what s/he wants, what s/he does not want, and what is simply random error. Every BDRM is decomposable into variance / covariance "between wanted" facets (bw), "between unwanted" facets (bu) and "within wanted and unwanted" facets (wuw) of generalization.

For BDRM's with one observation per intersection of wanted and unwanted facets, it is not possible to isolate the interaction wanted x unwanted (wxu) from the random error (e) variance. But this is possible for replicated BDRM'S, i.e. with more than one observation per intersection.

Unreplicated BDRM:

(1) $C_{tot} = C_{bw} + C_{bu} + C_{wuw}$

(2) $C_{wuw} = C_{wxu} \cup C_e$

The symbol for set union means, that in this case interaction and random error are confounded.

Replicated BDRM:

(3) $C_{tot} = C_{bw} + C_{bu} + C_{wxu} + C_e$

How is reliability tied to the BDRM? This is very easy. Reliable variance is true variance. True variance is systematic variance, i.e. not random error variance. Thereforewe add over all sources of systematic variance, i.e.:

(4) $C_{true} = C_{bw} + C_{bu} + C_{wxu}$

Sum over these variance / covariance matrices and sum over total matrix, to get the true and total variance of the first centroid. Then simply divide the two to get the reliability coefficient for the first centroid of such a BDRM:

(5) $$r_{tt_{overall}} = \frac{1'(C_{bw} + C_{bu} + C_{wxu})\, 1}{1'\, C_{tot}\, 1}$$

Overall reliability is always the sum of reliability for wanted, unwanted, and the interaction wanted x unwanted coordinates or facets.

(6) $$r_{tt_{overall}} = r_{tt_w} + r_{tt_u} + r_{tt_{wxu}}$$

Generalizability coefficients are easily derivable.With these coefficients we want only to generalize over the wanted, unwanted , interaction or combinations of these facets. The base for comparing generalizable variance is then only wanted or unwanted variance etc., plus random error variance:

(7) $$\mu_w = \frac{1'\, C_{bw}\, 1}{1'(C_{bw} + C_e)\, 1}$$

(8) $$\mu_u = \frac{1'\, C_{bu}\, 1}{1'\,(C_{bu} + C_e)\, 1}$$

(9) $$\mu_{wxu} = \frac{1' C_{wxu}\, 1}{1'\,(C_{wxu} + C_e)\, 1}$$

(10) $$\mu_{w,u} = \frac{1'(C_{bw} + C_{bu})\, 1}{1'(C_{bw} + C_{bu} + C_e)\, 1}$$

(11) $$\mu_{w,wxu} = \frac{1'(C_{bw} + C_{wxu})\, 1}{1'(C_{bw} + C_{wxu} + C_e)\, 1}$$

$r_{tt_{overall}}$ is always equal $\mu_{overall}$.

(12) $$\mu_{overall} = \mu_w + \mu_w + \mu_{wxu}$$

Generalizability coefficients are incommensurable and cannot easily be summed up, whereas our former coefficients are. Multivariate reliability coefficients are easily developed using Cohen's (1982) set correlation system (Wittmann, in press).

The BDRM as in Fig. 1 can be partitioned according to other facets than persons. This leads to the well known techniques of factor analysis or in generalizabilit theory, to what Cardinet, Tourneur and Allal (1976) have demonstrated.

The Cardinet-, Tourneur- and Allal-symmetry means developing generalizability coefficients for the other possible partitionings of the BDRM.

Reliability is best defined as a set of answers to questions of repeatability. This is nowhere more visible than in the partitioning of the BDRM.

What is won in knowing these reliability or generalizability coefficients ? Reliability is a prerequisite for validity and you have to make sure that the wanted aspects of your research program are repeatable. What should be done when the magnitude of your wanted reliability coefficient or component is not high enough ? Most often we use the Spearman-Brown prophecy formula and try to lengthen our tests. If we do it practically we use aggregation (Epstein, 1983) How does aggregation work ?

$$(13)\quad r_{tt_{(k)}} = \frac{k\, r_{tt}}{1 + (k-1)\, r_{tt}}$$

Substituting Eq. (5) in the Spearman-Brown prophecy formula Eq. (13) gives after some manipulation:

$$(14)\quad r_{tt_{overall_{(k)}}} = \frac{1'(C_{bw} + C_{bu} + C_{wxu})\,1}{1'(C_{bw} + C_{bu} + C_{wxu}) + \frac{1}{k}\,1'C'_e\,1}$$

or for wanted variance:

$$(15)\quad r_{tt_{w(k)}} = \frac{1'C'_{bw}\,1}{1'C_{bw}1 + \frac{1}{k}\,(\,1'(C_{bu} + C_{wxu} + C_e)\,1\,)}$$

We must always be sure to have parallel tests for our wanted components. These tests should be as heterogenous as possible with respect to unwanted variance.

Eq.(15) can be summarized as follows: Reliability conceived as a general concept of science depends on variability of wanted, unwanted and error facets. We can influence reliability through

(1) Choosing greater heterogenity with respect to wanted variance: $1'C_{bw}\,1$. Humphreys (1962, 1979) has always made this proposal.

(2) Greater homogenity of unwanted variance, arrived at through cancellation of different heterogenous components, to decrease $1'C_{bu}1$.

(3) Increasing the test k times with perfect parallel tests, and thus decreasing $1'C_{bu}1$, $1'C_{wxu}\,1$ and $1'C_e1$ by factor k.

(4) Minimizing the ratio of unwanted to wanted variance $1'(C_{bu}+C_{wxu})1\ /\ 1'\,C_{bw}\,1$.

So far nothing has been said about validity, though often, falsely in my eyes, generalizability to a universe of content is related to validity (i.e. construct validity). The relationships between reliability and validity can best be conceptualized with four separate data boxes.

THE FOUR BOX CONCEPTION

Four data boxes are necessary to depict most of psychology's research problems. The first box is a predictor box. Box number 2 is the criterion box. Box number 3 is an experimental treatment box, where persons are randomly assigned to planned treatments. Box 4 is the nonexperimental treatment box, containing treatments where randomization is not possible or containing unplanned treatments with positive or negative influence for criterion boxes. Fig. 2 gives a pictorial representation of the four box conception.

Insert Fig. 2 about here.

It is immediately evident that questions of reliability for the predictor box and the criterion box can be asked, although we will also recognize that questions of reliability of criteria are not so much emphasized as the reliability of predictors. Partitioning leads to reliability for the predictor box:

$$(16)\quad r^{pr}_{tt_w} = \frac{1'\,C^{pr}_{bw}\,1}{1'(C^{pr}_{bw} + C^{pr}_{bu} + C^{pr}_{wxu} + C^{pr}_{e})\,1}$$

and for the criterion box

$$(17)\quad r^{cr}_{tt_w} = \frac{1'C^{cr}_{bw}\,1}{1'(C^{cr}_{bw} + C^{cr}_{bu} + C^{cr}_{wxu} + C^{cr}_{e})\,1}$$

As a first result we can develop a canonical correlation corrected for attenuation with these reliabilities, i.e. for the first canonical component of wanted variance:

$$(18)\quad \text{Cancor}\,(pr,cr)_{1.C} = \frac{\lambda^{1/2}_{1.C\ (pr,\ cr)}}{\sqrt{r^{pr}_{tt_w}\cdot r^{cr}_{tt_w}}}$$

This reasoning immediately leads to the question of reliability of a planned (etr-box) or an unplanned (ntr-box) treatment. How can one develop a formula for it ? In ANOVA notation we partition total variance (V_{tot}) in variance between (bq) and within groups (wq). To obtain a reliability formula for an etr-box we divide the systematic (true) variance (bq) by total variance, i.e.:

$$(19) \quad r_{tt_w}^{etr} = \frac{1' V_{bq}^{etr} \, 1}{1' (V_{bq}^{etr} + V_{wq}^{etr}) \, 1}$$

However, within groups, all subjects are assigned the same score (either one or zero when using dummy coding). Thus the variance within is always a zero matrix and the reliability of an experimental treatment box is always one !

In the ideal experiment it is assumed that through proper randomization and operationalization each subject receives the same amount of treatment or no treatment. In reality however, the variability within groups can be enormous and the reliability of the planned treatment approaches zero. Treatment as intended has no effect then, i.e. does not correlate with criteria. F.e. a modification of eq. (18), i.e.

$$\lambda_{etr,cr}^{1/2} = 0, \text{ because } r_{tt_w}^{etr} \text{ is zero}$$

or so low, thus attenuating a true cancor (etr,cr) so much, that it does not reach significance.

This does not mean that variability within groups cannot correlate with criteria, this is an independent possibility. To illuminate this problem further, we can compare it to Murray's concept of α- and β- press.

Regressing boxes on each other can be done via set correlation. Cohen (1982; Cohen & Cohen, 1983) have already developed, prepared and sharpened that tool.

Regressing criterion boxes on treatment boxes and predictor boxes is the route experimentalists prefer. This arrangement of the "four-box-conception-map" shows the Northwestern passage. Campbell, Cook, Boruch et al. from Northwestern University recommended it over all these years as the best research design f.e. in evaluation research if this randomization is feasible. Cronbach (1982), in the same area, preferred the Southwestern passage - regressing criterion boxes on nonexperimental treatment and predictor boxes, with the assumption of obtaining more generalizable results for real world problems.

The implicit continuum in Fig. 2 between etr-box and ntr-box can be regarded as the gradual flow from true experiments to quasi-experiments.

Relating these four boxes means trying to answer questions of validity, which include prediction and explanation. To obtain successful prediction and explanation we urgently need a principle of symmetry.

BRUNSWIK SYMMETRY; PRINCIPLES FOR SUCCESSFUL PREDICTION AND EXPLANATION.

In science, principles of symmetry often lead to new discoveries and solution of long standing problems. Brunswik's lens model has this built in beauty. I have used a hierarchical version of the lens model to illustrate how successful prediction works (Wittmann, in press). Fig. 3 illustrates how predictor boxes and criterion boxes should look if we use hierarchical (f.e. personality) models. Only symmetrical relationships are fair tests of validity.

Insert Fig. 3 about here.

An unfair test is f.e. relating a broad secondary factor with a single act behavioral criterion either in personality or aptitude research and hoping for a high correlation. Correlations between boxes can only be optimal under conditions of good reliability and symmetry. The boxes have to contain symmetrical components. In terms of the lens model equation (Tucker, 1964) the correlation between boxes f.e. pr and cr is:

$$(20) \quad r_{pr,\,cr} = G_{pr,\,cr} \, R_{pr} \, R_{cr} + C_{pr,cr} \sqrt{1-R_{pr}^2} \; \sqrt{1-R_{cr}^2}$$

$G_{pr,\,cr}$ is the correlation between linearly predictable pr-boxes and linearly predictable cr-boxes. R_{pr} and R_{cr} are multiple correlation coefficients mapping the linear predictability of pr or cr components, respectively.

So far nothing in Eq. 20 is said about reliability. $G_{pr,cr}$ can be attenuated as every coefficient of validity.

$$(21) \quad G_{pr,cr}^{true} = \frac{G_{pr,cr}}{\sqrt{r_{tt_w}^{pr} \cdot r_{tt_w}^{cr}}} \quad \text{, which means}$$

$$(22) \quad G_{pr,cr} = \sqrt{r^{pr}_{tt_w} \cdot r^{cr}_{tt_w}} \quad G^{true}_{pr,cr}$$

$G_{pr,cr}$, compared to the true relationship, is attenuated by the square root of the product of the reliability coefficients of wanted variance in predictor and criterion box. Substituting eq. (22) in eq. (20) and dropping the non-linear term for convenience and clarity, gives:

$$(23) \quad r_{pr,cr} = \sqrt{r^{pr}_{tt_w} \cdot r^{cr}_{tt_w}} \quad G^{true}_{pr,cr} \cdot R_{pr} \cdot R_{cr}$$

Eq.(23) shows how a true correlation can be attenuated. First through lack of reliability of the predictor box variables, second through lack of reliability of criterion box variables, third through lack of construct reliability of the predictor box variables and fourth through lack of construct reliability of the criterion box variables (or construct indicators).

I prefer to label R_{pr} or R_{cr} as construct reliability, meaning the amount of overlap of our indicators with intended or wanted constructs. The term validity should be reserved for relating different constructs.

Boruch & Gomez (1977) have proposed overlap indices between intended and actually measured aspects of treatment and response variables. Leinhardt & Seewald (1981) discuss in instructional research such overlap between what is taught and what is tested as a crucial point in evaluating curricula. In the lens model equation these overlap indices already have a numerical solution. I must quickly emphasize that these indices are only realistically measurable or estimable if we know what we want. Should we want to measure a construct at a secondary or a primary level, how large or narrow should the breadth of the construct be ? More general constructs have a lower family resemblance of constituting elements (Wittgenstein, 1953) than constructs at a lower level of generality. The popular prototype approach is in danger of loosing breadth, leading possibly to constructs of a lower level. Overlap or construct reliability can be lowered by too few or by too many indicators for a wanted construct. In the first case relevant indicators are missing, in the second case unwanted indicators are included. Remember that asymmetry is symmetrical in both directions ! Under ideal circumstances $R_{pr} \cdot R_{cr} = 1$ should hold and the

same must be true for all other pairwise relations between boxes. But pay attention also that this does not automatically tell you at what level of generality you are.

We could draw many interesting relationships between all four boxes with insights in measurements of change and how most paradoxes of classical test theory vanish under such an approach (see Wittmann, in press). But there is no place to demonstrate all that here.

Let us conclude with an empirical example from intelligence and school achievement research, i.e. relating the pr-and cr-boxes.

PREDICTING SCHOOL GRADES AND OTHER KINDS OF AGGREGATED CRITERIA FROM THE BERLIN MODEL OF INTELLIGENCE.

The Berlin model of intelligence (Jäger, 1982, 1984) nowadays is the most prominent model concerning structure of intelligence in german speaking Europe. A short description of the model and how it had been developed by principles of systematic aggregation is given in Wittmann (in press). The model has as many psychometric models a hierarchical structure. At the bottom 48 test are grouped in four factors of an operation mode and three factors of a content mode. The content mode(CN)factors are Number (NCN), Figure (FCN) and Verbal (VCN). The operation mode (OP) contains aggregates of more process-like aspects of intelligence:
Speed on tasks (SOP), Memory (MOP), Creativity (COP) and Processing capacity for complex information (POP). These seven factors constitute a kind of primary factor level. The two modes as broader factor classes can be regarded as a secondary level and g-intelligence as an aggregate over all seven factors at the top of the pyramide. We are already accustomed to laying the hierarchy on the side in the lens model framework. Fig. 4 shows this model.

Insert Fig. 4 about here.

Perception immediately forces us to look for a symmetric structure at the criterion side (cr-box) and to ask what kind of criteria we can predict with such a model. Admittedly we do not know. But Jäger (1982) in addition to measuring his model components, had gathered school grades, interests, and factor-analytically derived scales of self-evaluated abilities in different areas of the content and operation mode. The sample analyzed here consisted of 545 (289 male and 256 female) "Gymnasium"-pupils (age 16-21 with arithm. mean 17,6 years) from Berlin schools.

With these cr-box variables we can try differently aggregated variables to find from which level of generality they are best explained. In relating different sets of variables from pr-and cr-boxes we used set-correlation (Cohen, 1982; Cohen & Cohen, 1983) as multivariate correlation and in relating single variables or single aggregated indices with sets of variables we used hierarchical multiple regression analysis.

The squared and unsquared (in parenthesis) set correlations between school grades (S) as a set (see Tab. 1 for the list of grades used, total grade not included) were as follows.
With the set of four operative factors (OP):

$R^2_{S,OP}$ = .2800 (.5292),

with the set of three content factors (CN):

$R^2_{S,IN}$ = .3318 (.5760) ,

with the set of four operative and three content factors (OPIN):

$R^2_{S, OPIN}$ = .4291 (.6551)

Partial set correlations of school grades with content factors, operative factors partialed out of each set:

$R^2_{S\cdot OP,CN\cdot OP}$ = .2071 (.4551) and

partial set correlation of school grades with operative factors, content factors partialed out of each set:

$R^2_{S\cdot CN,OP\cdot CN}$ = .1456 (.3816)

These set correlations indicate that the Berlin model is a good predictor of school grades (as we have known for a long time from good intelligence tests). From the perspective of differential validity, the content factors show a higher validity then the operative factors. We do not know why,but can speculate that through the very process of grading, operative factors are eggregeted out,diminished or these aptitudes are not graded. Regarding the restriction of range with respect to intelligence in a "Gymnasium"-sample all corre-

lations are severely underestimated.

Table 1 shows the result of single grades, total (cumulative) grade and two more complex composites using multiple regression. Operative factors (OPFAC) and content factors (CNFAC) are correlated. We used commonality-analysis (Cooley & Lohnes, 1976) to orthogonalize both factor classes in unique OPFAC (U^2_{OP}), in unique CNFAC (U^2_{CN}) and the commonality between the two sets (C_{OPCN}). We now see more clearly the predictive validity of each level of generality regressed on these cr-variables. CN-factors have the highest coefficients, although in some grades there is an advantage of OP-factors.Interestingly this is most obvious in politics/history. Separate analysis for that grade showed that creativity (COP) was the single best predictor.

The three highest validities were found for the two complex composites and grades in mathematics. The numbers in parenthesis are computed according to McNemar (1949,p.126)under the very conservative assumption that the sample of "Gymnasium"-pupils are restricted in range compared to the total population of 2/3 with respect to standard deviation. Approximately 20% of an age cohort visit "Gymnasium" in Germany.

Insert Table 1 about here.

The two composites AGGNAWI and AGGEIWI represent European philosophers beloved distinction between "Naturwissenschaften" and "Geisteswissenschaften" (see Legend Tab. 1).

Obviously most symmetrical to the seven factors of the Berlin intelligence model is a complex compound of achievement in science classes,interests and personality. g-intelligence also predicts this index best, though to a much smaller degree than the seven factors. Total average grade is the third best predicted criterion from g and here we see no difference(after considering shrinkage) compared to the prediction from the seven factors.

Very astonishing is the zero correlation of g with AGGEIWI.

Tab. 2 a and b show the hierarchical regression of AGGNAWI and AGGEIWI on the seven factors. In Tab. 2 a.we see that Number (NCN) correlates with a different sign than Verbal (VCN). Thus aptitudes cancel each other and the zero correlation becomes explainable. We also see that Number (NCN) and

Figure (FCN) are suppressor variables.

Both suppress irrelevant variance in creativity (COP). Partialed creativity, which means creativity in numerical and figural tasks partialed out, is the best predictor of AGGEIWI. This partialed creativity is than best described as verbal creativity. Suppressors always give hints that a more symmetrical relationship is obtainable at a lower level of generality where irrelevant (unwanted) components are removed. Looking at AGGNAWI we need at least four from the seven factors in an additive linear equation to explain the criterion variable. These factors stem from both modes, i.e. processing capacity (POP), speed on tasks (SOP) from the operation mode and number (NCN) and figure (FCN) from the content mode.
The results are embarassing to the beloved distinction mentioned above between "naturwissenschaftliche" and "geisteswissenschaftliche" educational achievements. One can often hear that the latter are more differentiated and need more complex aptitudes constellations, in contrast to the more "narrow-minded" science adherents, advocates who only deal with numbers and figures. From the point of view of this exploratory analysis quite the opposite seems to be true.

I hope I was able to whet your appetite for these concepts and at the same time convince you that the synthesis proposed really is an improvement of construct and predictive validity.

REFERENCES

BORUCH,R.F. & GOMEZ,H. (1977). Sensitivity, bias, and theory in impact evaluations. Professional Psychology, 8, 411-434.

BRUNSWIK,E.(1956). Perception and the representative design of psychological experiments. Berkeley: University of California Press.

CARDINET,J., TOURNEUR,Y. & ALLAL,L. (1976). The symmetry of generalizability theory: Applications to educational measurement. Journal of Educational Measurement, 13,119-135.

CATTELL,R.B. (Ed.) (1966). Handbook of multivariate experimental psychology. Chicago: Rand McNally.

COHEN,J.(1982). Set correlation as a general multivariate data-analytic method. Multivariate Behavioral Research, 17,301-341.

COHEN,J. & COHEN,P. (1983). Applied multiple regression/correlation analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

COOK,T.D. & CAMPBELL,D.T.(1979). Quasi-experimentation. Design and analysis issues for field settings. Chicago: Rand McNally, 1979.

COOLEY,W.W. & LOHNES,P.R. (1976). Evaluation research in education. New York: Irvington.

CRONBACH,L.J.(1957). The two disciplines of scientific psychology. American Psychologist,12,671-684.

CRONBACH,L.J.(1975). Beyond the two disciplines of scientific psychology. American Psychologist,30,116-127.

CRONBACH,L.J.(1982). Designing evaluations of educational and social programs. San Francisco:Jossey-Bass.

CRONBACH,L.J. (1984 a). Essentials of psychological testing (4th ed.). New York: Harper and Row.

CRONBACH,L.J. (1984 b). A research worker's treasure chest. Multivariate Behavioral Research, 19,223-240.

CRONBACH,L.J. (1985). Construct validation after Thirty years. Paper delivered at Urbana, Illinois, May 1,1985, as part of a symposium on intelligence honoring Professor Lloyd G.Humphreys and to be published as part of the Symposium volume (mimeo).

CRONBACH,L.J., GLESER,G.C., NANDA,H. & RAJARATNAM,N.(1972). The dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: Wiley.

EPSTEIN,S.(1983). Aggregation and beyond. Some basic issues on the prediction of behavior. Journal of Personality,51,360-392.

EYSENCK,H.J. (ed.) (1981). A model for personality. New York:Springer-Verlag.

GLASS,G.V.(1983). Evaluation methods synthesized. Review of L.J.Cronbach (1982). Designing evaluations of educational and social programs. San Francisco: Jossey-Bass. Contemporary Psychology,28,501-503.

HUMPHREYS,L.R.(1962). The organiz. on of human abilities. American Psychologist,17,475-483.

HUMPHREYS,L.G.(1979). The construct of general intelligence. Intelligence, 3, 105-120.

JÄGER,A.O. (1982). Mehrmodale Klassifikation von Intelligenzleistungen: Experimentell kontrollierte Weiterentwicklung eines deskriptiven Intelligenzstrukturmodells. Diagnostica,28,195-225.

JÄGER,A.O. (1984). Intelligenzstrukturforschung: Konkurrierende Modelle, neue Entwicklungen, Perspektiven. Psychologische Rundschau,35,21-35.

LEINHARDT,G. & SEEWALD,A.M. (1981). What's tested, what's taught? Journal of Educational Measurement,18,85-96.

LUMSDEN,J.(1976). Test theory. Annual Review of psychology,27,252-280.

McNEMAR,Q.(1949). Psychological statistics. New York: Wiley.

SPEARMAN (1904). The proof and measurement of association between two things. American Journal of Psychology, 15,72-101.

STANLEY,J.C.(1971). Reliability. In R.L. Thorndike (ed.), Educational measurement. 2nd.Edition. Washington,D.C.: American Council on Education.

TUCKER,L. (1964). A suggested alternative formulation in the developments by Hursch, Hammond, and Hursch, and by Hammond, Hursch, and Todd. Psychological Review, 71,528-530.

WEISS,D.J. & DAVISON,M.L.(1981). Test theory and methods. Annual Review of Psychology, 32, 629-658.

WITTGENSTEIN,L.(1953). Philosophical investigations. New York:Mac Millan.

WITTMANN,W.W. (in press). Multivariate reliability theory: Principles of symmetry and successful validation strategies.
In R.B. Cattell and J.R. Nesselroade (Eds.). Handbook of multivariate experimental psychology (2nd ed.). New York: Plenum Press.

Author's address:

Prof. Dr. Werner W. Wittmann
Psychologisches Institut
der Universität Freiburg
Abteilung Persönlichkeitspsychologie
Niemensstr. 10

D 78oo Freiburg i. Br.
Federal Republic of Germany

Table 1: Predictive validities for various criteria from the Berlin model of intelligence factors.

Grades	R	R^2	R^2_{OP}	R^2_{CN}	R^2_G	U^2_{OP}	U^2_{CN}	$C_{OP\ CN}$
German	.3414 (.478)	.1165	.0420	.0998	.0233	.0167	.0745	.0253
English	.2207 (.321)	.0487	.0137	.0460	.0088	.0027	.0350	.0110
2nd foreign language	.1165 (.173)	.0136	.0121	.0012	.0000	.0124	.0015	-.0003
Mathematics	.4513 (.604)	.2037	.1583	.1747	.1156	.0290	.0454	.1293
Chem/Bio/-Physics	.3532 (.493)	.1247	.1198	.1085	.1011	.0162	.0045	.1036
Politics/History	.2847 (.407)	.0792	.0661	.0357	.0240	.0435	.0131	.0226
Arts/Music	.2063 (.302)	.0426	.0170	.0124	.0069	.0302	.0256	.0132
Sports	.1774 (.261)	.0315	.0262	.0150	.0133	.0165	.0053	.0087
Total ave. grade	.3309 (.466)	.1095	.0913	.1017	.0861	.0078	.0182	.0835
AGGNAWI	.6048 (.752)	.3658	.2151	.3104	.1354	.0554	.1507	.1597
AGGEIWI	.4267 (.578)	.1821	.0476	.1639	.0084	.0182	.1345	.0293

Legend: AGGNAWI is a composite of grades in Mathematics, Chemistry/Biology/-Physics, interests in vocational areas of natural sciences and self-evaluated abilities in these classes and areas.
AGGEIWI is a composite of grades in Languages (German, English, second foreign language), interests in Arts, Humanities, and self-evaluated abilities in these domains.

a) DEPENDENT VARIABLE AGGEIWI

Variable	Multiple R	R Square	RSQ Change		Simple R
COP	.21198	.04494	.04494	(.018)	-.21198
MOP	.21278	.04528	.00034	(.002)	-.01843
SOP	.21300	.04537	.00009	(.000)	.00960
POP	.21808	.04756	.00219	(.000)	.04681
NCN	.35192	.12384	.07628	(.033)	.18288
FCN	.42562	.18115	.05731	(.000)	-.00347
VCN	.42667	.18205	.00090	(.130)	-.36115

b) DEPENDENT VARIABLE AGGNAWI

Variable	Multiple R	R Square	RSQ Change		Simple R
COP	.10899	.01188	.01188	(.000)	-.10899
MOP	.13912	.01935	.00747	(.015)	-.08646
SOP	.19496	.0381	.01866	(.037)	-.13659
POP	.46378	.21509	.17708	(.002)	-.42081
NCN	.58395	.34100	.12590	(.263)	-.51303
FCN	.60350	.36422	.02322	(.038)	-.19493
VCN	.60479	.36578	.00156	(.009)	.09576

Table 2: Multiple Regression analysis with Jäger's seven intelligence factors on two complexly aggregated indices.

Legend: Factor scores were used for prediction - operative factors first, then content factors. The numbers in parentheses give the squared multiple correlations with content factors first and operative factors second in a hierarchical regression analysis.

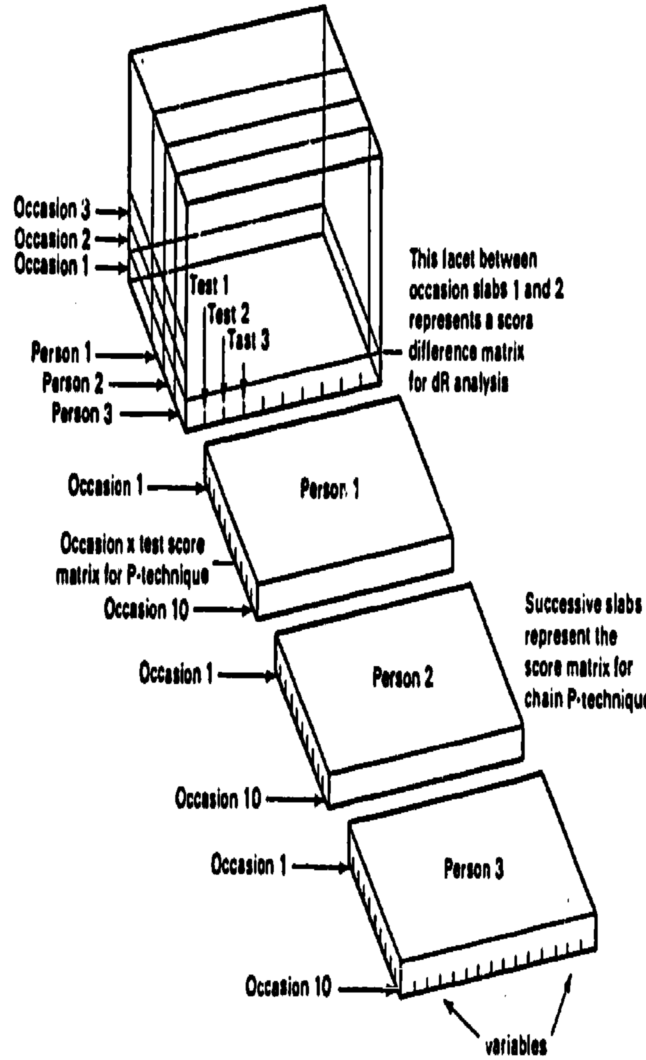
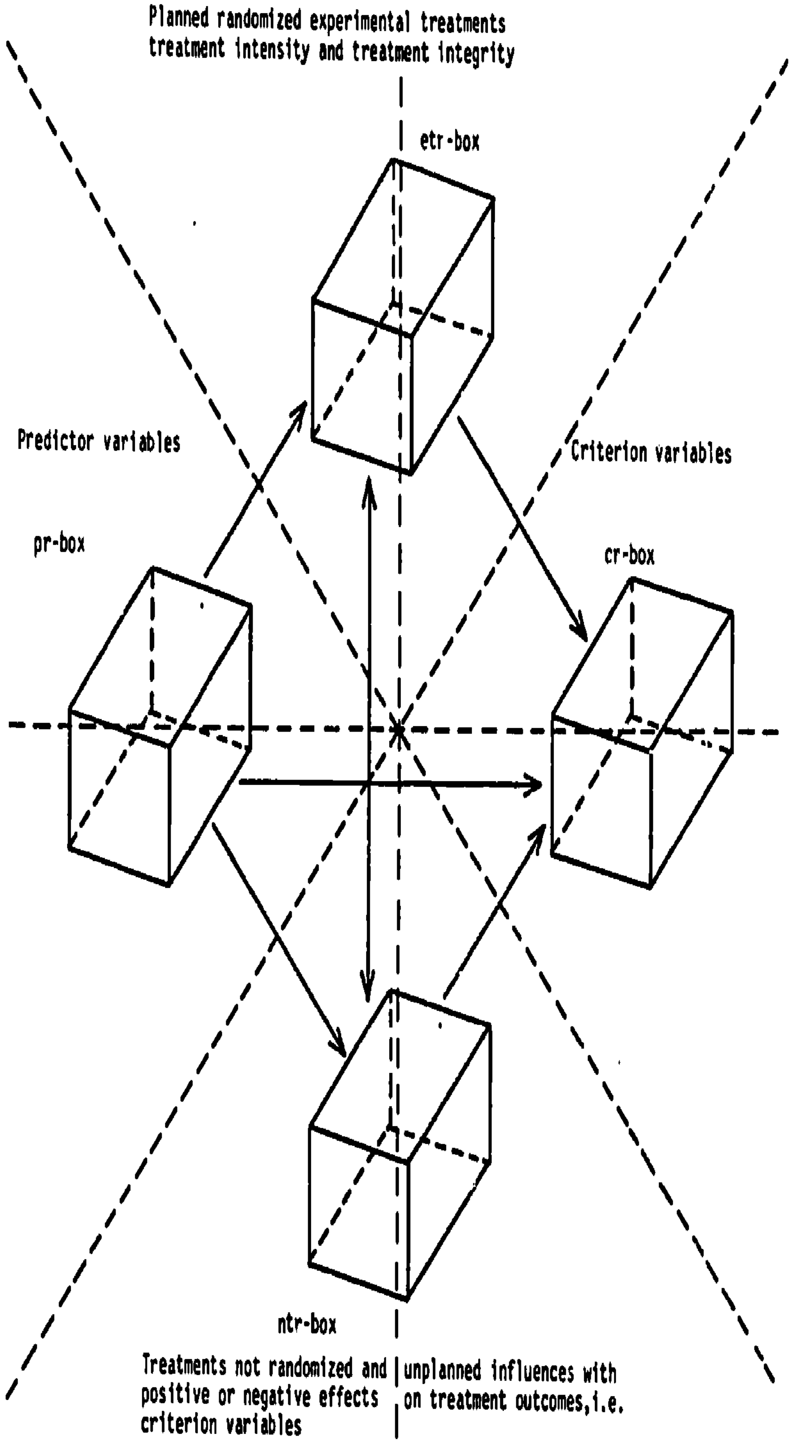
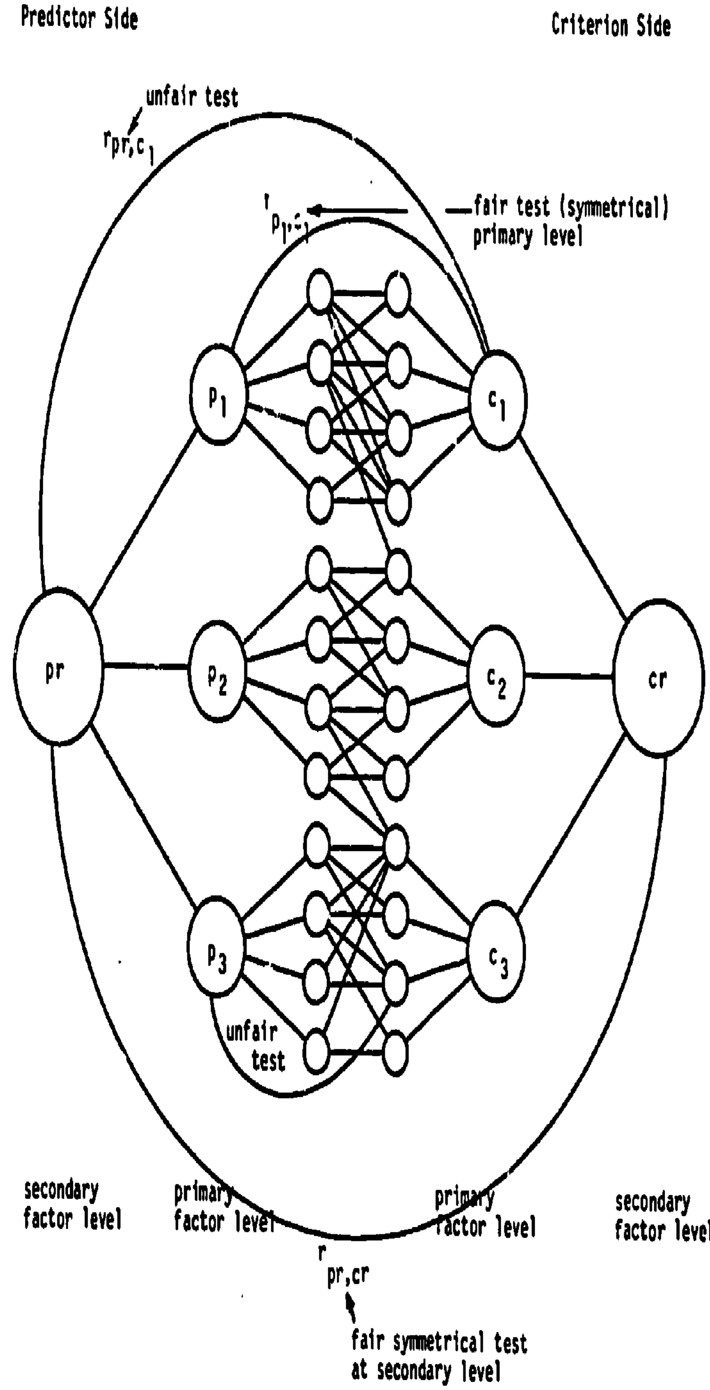


Figure 1 : Unfolding the three-dimensional data box



<u>Figure 2 :</u> The four box conception



<u>Figure 3 :</u> Hierarchical version of Brunswik's lens model for denoting principles of symmetry between predictors and criteria.

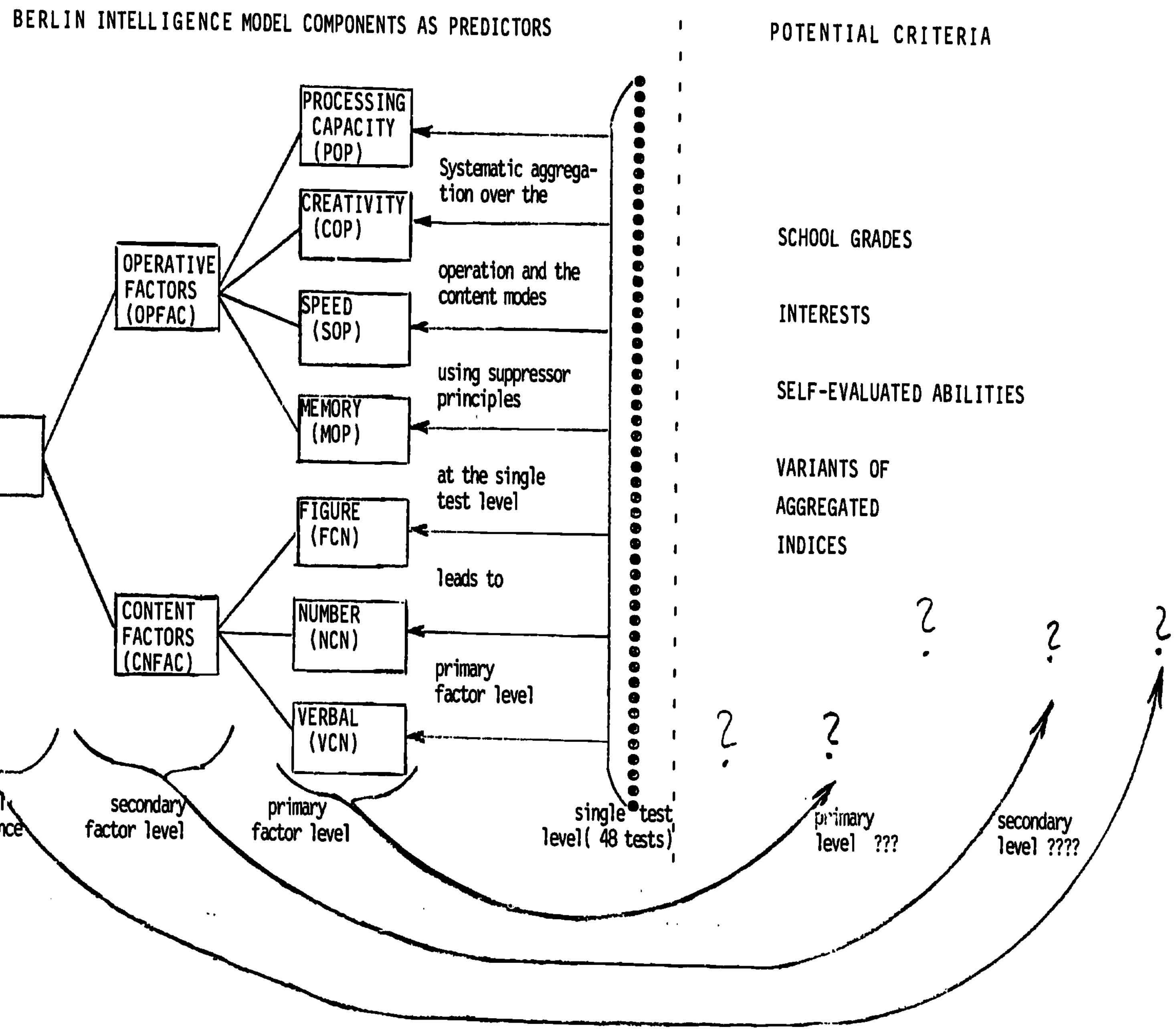


Figure 4 : The Berlin intelligence model in lens model framework