A SYNOPSIS OF PROFESSOR HANS-MAGNE EIKELAND'S WORK IN PSYCHOMETRICS[1]

Knut A. Hagtvet

Department of psychology
University of Oslo

Professor Hans-Magne Eikeland introduced new psychometric ideas to the Scandinavian countires during 1960s and 1970s. However, parts of his writings have remained unpublished, but were at that time as well as more recently highly assessed and appreciated by scholars within and outside of Norway. He introduced the new ideas in psychometrics that surfaced during the 1960s on the international scene by his ongoing seminars and lectures at the Institute for Educational Research, University of Oslo, as well as at other universities in Scandinavia during late 1960s and early 1970s. During 1968-69 he spent a sabbatical year at the State University of New York at Bufallo. He later presented his lectures in a comprehensive monograph (Eikeland, 1973a). In this monograph Eikeland discussed extensions of classical test theory that moved into the modern generalizability theory. A separate paper on the expected covariance matrix (Eikeland, 1970) also represented the new upcoming ideas in psychometric theory by suggesting an alternative definition of the well known alpha coefficient within a full-fledged random sampling model. The present synopsis provides an overview of Eikeland's delivery of the ongoing reorientation in psychometric theory during this periode as well as his own contributions to this development.

An overview of professor Eikeland's writings in psychometrics is mostly reflecting his ideas about applying ANOVA as a correlational technique in contrast to statistical testing of group differences in experimental designs. Eikeland's interest in correlational applications of the ANOVA framework can be divided into three areas:

[1] This synopsis was written during Spring 2011.

a) Applying ANOVA in the service of psychometric inference in multifacet test designs

b) Applying ANOVA to decompose observed variance in terms of descriptive variance components

c) Applying ANOVA in the framework of factor analysis

Applying ANOVA in the service of psychometric inference in multifacet test designs.

Professor Eikeland has been involved in what is formally called *psychometric inference.* In this type of inference the researcher is concerned with assessing the quality of his/her measurement with respect to how well he/she is able to generalize from a *particular test* to a *universe of tests*. This inference differs from the more common *statistical inference* in which the inference is made from a *sample of individuals* to a *population of individuals*. Both statistical and psychometric inferences are of utmost importance, although most attention has been given to statistical inference on the cost of psychometric inference in the social sciences.

Psychometric inference has mostly been based on one-facet or unstratified test designs where the measurements are classified into one category only. Theoretically this situation is effectively treated within the framework of classical test theory. However, in social science in general, and in psychology and education, in particular, the constructs to be measured or the assessment situations are often of a more complex nature. Manifestations or indicators of a construct can be organized by multifacet test designs into two or more categories often belonging to two or more dimensions or facets. When assessing reliability of scores belonging to multifacet measurement designs, the psychometric theory needs to be extended beyond the simple classical one-facet design. Professor Eikeland has devoted much of his time to elaborate the theory for this type of psychometric inference. This is convincingly demonstrated by his comprehensive discussion of the historic development in psychometric theory that eventually ended in the theory of generalizability which is one of the major theories of psychometrics today (Eikeland, 1973a). His first work coauthored with Rabinowitz (Rabinowitz & Eikeland, 1964), focused on different varieties of two-facet designs. Later he described assessment situations where he generalized the theory for one-facet design into multifacet designs (Eikeland, 1972a). Estimating reliability within the framework of generalizability theory was in the making during 1960's initiated and driven mainly by Cronbach and associates (Cronbach, Rajaratnam, Gleser, 1963; Cronbach, Gleser, Nanda, & Rajaratnam, 1972) at Stanford University. Eikeland made contributions to this development by extending the two-facet mixed model (where subtests/strata were treated as a fixed facet) to a complete random model where also strata were assumed to be ***random*** (Rabinowitz & Eikeland, 1964). Eikeland (1972a) discussed the realism of this model and believed it would come into play in future estimations of generalizability parameters. Rajaratnam, Cronbach and Gleser (1965) discussed estimation of stratified internal-consistency coefficients but restricted themselves to the more common models in which strata were ***fixed***. They referred to Rabinowitz and Eikeland's (1964) extension of the stratified model into a complete random model. Interestingly, support for the type of model suggested by Rabinowitz and Eikeland (1964) was later clearly recognized by Cronbach in a personal communication with Shavelson, Webb and Rowley (1989). According to Shavelson, et al. (1989) Cronbach suggested an alternative interpretation of reliability for test batteries in which several subtests measure a dimension;

> "...if subtests are indicators of a construct (e.g., verbal reasoning), the analysis might better view subtests as *random* (my italics) and evaluate the adequacy of the test score as a representation of the domain of verbal reasoning subtests. It seems to me that an interpretation that a pupil is better in verbal than Abstract Reasoning [sic] is a statement about the domain, not fixed subtests."(Cronbach, personal communication,

July 15, 1987).

This statement appears to be a well founded support for treating strata facets as random as repeatedly discussed by Eikeland (Rabinowitz & Eikeland, 1964; Eikeland, 1972a).

One of his major accomplishments is his extensions of the well known Spearman-Brown prophecy formula. This formula is most often related to the one-facet design. Eikeland (1972a) applied the same rationale but extended its application into a family of different alpha coefficients applicable to a variety of multifacet designs. Even though multifacet designs are more complex in nature, they may be considered more realistic as operational definitions of psychological constructs. Eikeland's involvement in this matter was driven by his observation of the mismatch between complex measurement designs and lack of theory for estimating reliability for composite scores in such designs. As indicated by Eikeland (1972a, p. 75) sophisticated procedures existed for estimating parameters in complex experimental designs by complex ANOVA procedures. However, a corresponding sophistication for measurement designs did not exist in psychometrics at that time. Today we have the rationale and techniques to remedy such challenges. Multifacet test designs are described in advanced and sophisticated textbooks and some few methodological journals out of reach for an average trained researcher in social science. However, the more intuitive approach taken by Eikeland to portray the rationale and technique for multifacet designs is well within reach for applied researchers in the social sciences.

Eikeland (1972a) emphasized strongly the structural aspects as an inherent ingredient of psychometric inference. One of his main suggestions is that " … this general structural theory is but an extension of the long-respected Spearman-Brown rationale. That rationale has so far been restricted to the lowest level in the hierarchy of test designs, the unstratified test. The Spearman-Brown rationale has been the cornerstone in mental test theory for more than sixty years. What seems to come out of multifacet studies conducted so far, is that the

Spearmen-Brown basic thinking in test theory is about to get a much more general formulation. The new perspective for this old formula covers a variety of complex measurement procedures where the hierarchically stratified test design is but one." (p.78).

It seems to me that this perspective on the Spearman – Brown formula introduces a different message than the one offered in text books which restricts the Spearman – Brown rationale to one-facet designs mostly. According to Eikeland the Spearman- Brown rationale has a much wider application. In fact, he extended the generalizability of the Spearman-Brown rationale applied to certain types of multifacet designs.

Another interesting aspect of his work in this area is his focus on covariance matrices to implement the same ideas as behind ***variance*** components in ANOVA (Eikeland, 1970, 1972a). Thus, two different data-analysis languages existed for understanding and estimating generalizability and alpha coefficients as well as variance components.

An illustrative example of the covariance approach was provided by Eikeland (1971a) to offer a perspective on negative variance components. From a definitional point of view a variance cannot be negative. However, from a sampling viewpoint a sample variance can take on negative values caused by random variation. In this context a negative variance component is often set equal to zero in estimating generalizability parameters. Alternatively, a sizeable negative variance component may indicate that the linear model applied to estimate the variance components may have systematic specification errors. Eikeland (1971a) has illustrated by a simple design that a negative variance component, may at times meaningfully be interpreted as a covariance component which obviously makes sense. The covariance terminology offers a more intuitive language to portray the implied psychometric concepts, while at the same time represented a bridge to factor analysis models in which he later got involved (see below).

Applying ANOVA to decompose observed variance in terms of descriptive variance components

Variance components have occupied much of Eikeland's work in psychometrics. As described above variance components were estimated within the framework of generalizability theory. Estimated variance components represent the cornerstones in estimating generalizability coefficients or alpha coefficients. In this framework variance components support the psychometric inference reflecting how well the researcher could generalize from a sample of measures to a defined universe of generalization. Aside from the purpose of psychometric inference, Eikeland's interest in multifacet designs appears to have caused his interest also in the structural properties of such designs and in general the internal structure of complex systems of variation. This involvement brought him beyond test theory into complex multifacet systems of variation in general. The emphasis is now on a descriptive application of variance components – not as estimates of error and true/universe score variance components. This emphasis is explicitly expressed by the following quote;

> "... it is here argued that the most interesting and informative analysis of complex test data is the description of test score variance. The structural analysis is a correlational approach that describes the relationships of the parts going into the hierarchy. The decomposing into variance components is the fundamental basis for making a meaningful interpretation of the observed test scores in terms of the extent to which the battery is measuring one common trait running through all items and less common traits attributable to strata. Even specific traits can emerge, attributable to the substrata." (Eikeland, 1972a, p. 77).

The structural theory was of interest both as part of psychometric inference as well as describing complex systems of variation as illustratively shown in Eikeland (1971b, 1973c). Both the covariance terminology and the structural theory implicit in the a priori multifacet test designs paved the way for applying the ANOVA methodology in the framework of factor analysis.

Applying ANOVA in the framework of factor analysis

Eikeland demonstrated convincingly how the variance terminology of ANOVA on the one hand and the covariance terminology on the other hand could be applied for the same purpose. Eikeland applied the pedagogical potential of this terminological correspondence to present ANOVA as a correlational system. Then it becomes reasonable that variance components can be conceptualized in terms of covariance components. This terminological bridge makes it very instructive to consider factor analysis in terms of a priori covariance structures implicit in the ANOVA system applied to multifacet measurement designs.

Eikeland (1972b) applied the combined variance-covariance methodology in the framework of factor analysis which relied on a priori defined "factors" or linear combinations. By transferring the definition of factors to a priori features of multifacet test designs, a stronger emphasis was put on the conceptual basis for interpreting test score variance than what is often the case in exploratory factor analysis where a naive empiricism is driving the search for factors.

Eikeland (1972b) demonstrated three different ways of analyzing test score variance within multifacet test designs. The test score variance could be described in terms of an observed or manifest variance structure and alternatively in terms of two latent or inferred structures. In Eikeland's conceptualization the manifest structure provides a set of observed orthogonal linear combinations. However, according to Eikeland this manifest structure provides no insight within the score. No suggestions are made with reference to the internal structure of test scores. This is, however, the focus of the two latent variance structures. The first latent variance structure describes the composition of one average test score. In other terms, this structure describes the relative importance of sources of variation that enter into one single or typical test score. Eikeland found most interest in the second latent structure of test score variance which described the variance structure in the linear combination of the ***sum score*** or the test variance. While the variance structure for one average test is composed of unweighted or equally weighted variance components, the latent sum score variance is defined by a weighted sum of variance components. The weights are the a priori given numbers of items within substrata, the number of substrata within strata and the number of strata in the actual multifacet test design. That weighting system led Eikeland to vision a general Spearman-Brown rational within the framework of factor analysis applied to certain multifacet test designs.

Not only the linear combination of a sum score, but also the linear combination of the difference score was subjected to the same type of elaborated analysis (Eikeland, 1973b). The difference score was conceptualized in the framework of multifacet test design. Then both structural properties and inferential features were discussed. Formulas for estimating generalizability of difference scores were derived for seven different test designs. These applications speak to the generality of Eikeland's ANOVA methodology.

Assessment of Eikeland's perspectives in psychometrics

Eikeland's work deserves attention due to his intuitive approach to ***conceptualize*** complex designs and psychometric concepts mostly in the framework of generalizability theory. He estimated a priori variance structures often expressed in the language of correlations and/or covariances. His approach was very well welcomed and appreciated at the time of his teaching and writing. It is even more welcomed today when researchers too often rely on easy accessible modern software to estimate complex models without being required to elaborate the conceptual underpinnings of their estimation. Or, alternatively, too simple models are often estimated to account for realistic complex measurement designs. Unfortunately, we still see an almost a routine application of the one-facet alpha coefficient to estimate reliability even in multifacet test designs. Modern technology, even considered necessary for much psychometric work, still provides opportunities for unelaborated and empirically dominated estimation procedures. Even though the theory of generalizability was first introduced in 1963 and a comprehensive version was published in 1972, estimating reliability in a multifacet design is still not a well known and applied procedure. There is still a substantial impact of naïve empiricism in applied psychometric work. Furthermore, with few exceptions introductory literature in generalizability theory does not exist or exists within a specific context of educational measurement. Generalizability theory is still presented in a rather advanced terminology for the typical main stream researcher. Therefore it is an increased need for Eikeland's intuitive and conceptually founded psychometrics.

Guidelines for the applied researcher can be easily derived from Eikeland's methodological approach. On a general level efforts should be invested in elaborating the conceptual basis for the measurement model *prior* to estimation. His measurement philosphy gives a logical priority to conceptual ideas over and above formal mathematical structures. Secondly, on a practical level, as in test development, this philosophy demands constructs to be conceptually delineated in order to provide rational guidelines for constructing measurement instruments and later defining relevant measurement models.

Eikeland's vision was to emphasize the inherent correspondence between pre-defined constructs, measurement designs and estimation models to ease interpretation of observed empirical relationships. His methodological approach provided a close link to construct definition and the applied measurement context that differs from current expositions that focus on the formal mathematical framework of measurement models. Eikeland defended an a priori construct-based as opposed to an a posteriori oriented psychometric.

Eikeland devoted much of his efforts to what he called structural theory and analysis. Two aspects are essential in generalizability analysis; psychometric inference and structural properties of the measurement design or the universe of admissible observations. Psychometric inference has attracted most attention in the framework of generalizability analysis; that is, how well can we generalize from a sample of obervations to a universe of similar observations? Structural analysis, however, has not been central to generalizability studies probably because generalizability theory does not make any assumption about the content or the dimensionality of the construct domain. Cronbach et al. (1963) stated that "The universe must be unambigiously defined, but it is not necessary that the universe be homogeneous in any other sense" (p. 160).

It is worth noting that Eikeland , as different from the Cronbach school of generalizability, emphasized analysis of structural properties of the construct to be measured. Even though Eikeland did not underestimate the inference aspect of generalizability analysis, his attention was more strongly attracted to the structural properties. Maybe he considered the structural properties to have priority over inference? This involvement may have guided him in the direction of factor analysis way of thinking more than making inference to an extended universe of observations. The terminology of 'a priori latent constructs' or 'latent variables' are not typically applied in the framework of generalizability. However, Eikeland applied the terms 'latent constructs', ' latent variance structure' or ' deep structure' as opposed to manifest or observed structure. His structural terminology was more in line with the development in structural equation modeling/ structural covariance analysis, or more precisely, confirmatory factor analysis with it's a priori structural properties (Bollen, 2002; Jøreskog, 1969) than the conceptual framework of generalizability theory (Brennan, 1992; 2001; Cronbach et al., 1972; Hagtvet, 1998; 2010a,b). Eikeland's contributions to generalizability analysis rather parallelled contemporary writings by McDonald (1970; 1978; 1985; 1999) who launched what he called the 'factor analytic model of generalizability'. This model allows estimation of a generalizability coefficient, called omega, that rests on a unidimensional structure in a one facet design.

As the fields of statistics and psychometrics is developing, an increasing overlap and correspondence between different analytical frameworks has surfaced over the years. It is ample reason to expect further developments that will combine factor theory with generalizability theory to the broader context of multifacet design. Eikeland's perspectives and approaches represent a vision for this development in order to bring psychometric and conceptual theory closer together.

In sum, Eikeland has provided alternative perspectives on central psychometric concepts in multifacet measurement designs, in particular. His vision offered a far more intuitive conceptualization of psychometric models that have been and still are for most applied researchers hidden in mathematical terminology. Eikeland's unpublished work in psychometrics is now available for applied researchers as well as methodologists.

References

Bollen, K.A. (2002) Latent variables in psychology and the social sciences. In *Annual Review of Psychology, 53*, 605-34.

Brennan, R.L. (1992) *Elements of generalizability Theory*. Iowa City, IA, ACT.

Brennan, R.L. (2001) *Generalizability theory*. New York: Springer.

Cronbach, J.L., Rajaratnam, N., & Gleser, G.C. (1963) Theory of generalizability: A liberalization of reliability theory. *British Journal of Statistical Psychology, 16,* 137-163.

Cronbach, J.L., Gleser, G.C., Nanda, H., & Rajaratnam, N. (1972) *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.

Eikeland, H-M.(1970) *Coefficient alpha and the expected variance-covariance matrix of random composite measurements*. University of Oslo.

Eikeland, H-M.(1971a) *Concerning negative variance components in repeated measures designs*. University of Oslo.

Eikeland, H-M.(1971b) *Correlational analyses of school marks influenced by multiple sources of variance. Explorations into internal structures of complex systems of variation.* Paper presented at a symposium on evaluation in Uppsala, Sweden, November .

Eikeland, H-M.(1972a) *On the structure of generalizability theory for hierarchically stratified tests*. University of Oslo.

Eikeland, H-M.(1972b) *Toward a rationalistic factor analysis. Explorations into a priori covariance structures*. University of Oslo.

Eikeland, H-M.(1973a) *Frå spesifikk til generisk reliabilitetsteori*. University of Oslo.

Eikeland, H-M. (1973b) *Generalizability estimates for difference scores: An aspect of the construct validity of tests*. University of Oslo.

Eikeland, H-M. (1973c) *On the change in mental organization with age: an alternative approach*. University of Oslo.

Hagtvet, K.A. (1998) Assessment of latent constructs: a joint application of generalizabilitytheory and covariance modeling with an emphasis on inference and structure. *Scandinavian Journal of Educational Research, 42, 41-63.*

Hagtvet, K.A. (2010a) Generalizabilityteori. In Monica Martinussen (Ed.) *Kvantitativ forskningsmetodologi i samfunns- og helsefag.* Kap. 3: Pp. 89-119. Bergen, Norway: Fagbokforlaget.

Hagtvet, K.A. (2010b) Strukturelle likningsmodeller: Målingsmodell og regresjonsmodell. In Monica Martinussen (Ed.) *Kvantitativ forskningsmetodologi i samfunns- og helsefag.* Kap. 9. Pp. 285-314. Bergen, Norway: Fagbokforlaget.

Jøreskog, K.G. (1969) A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34*, 183-202.

McDonald, R.P (1970) The theoretical foundation of common factor analysis, principal factor analysis, and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology, 22*, 165-175.

McDonald, R.P (1978) Generalizability in factorable domains: Domain validity and generalizability. *Educational and Psychological Measurement, 38,* 75-79.

McDonald, R.P (1985) *Factor analysis and related methods.* Hillsdale, NJ: Erlbaum

McDonald, R.P (1999) *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.

Rabinowitz, W., & Eikeland, H-M. (1964) Estimating the reliability of tests with clustered items. *Scandinavian Journal of Educational Research (Pedagogisk forskning), 2, 85-105.*

Rajaratnam, N., Cronbach, J.L., & Gleser, G.C. (1965) Generalizability of stratified-parallel tests. *Psychometrika, 30*, 39-56.

Shavelson, R.J., Webb, N.M., & Rowley, G.L. (1989) Generalizability theory. *American Psychologist, 44*, 922-932.