40th INTERNATIONAL CONGRESS ON ASSESSMENT CENTER METHODS

BUILDING NEXT GENERATION LEADERSHIP

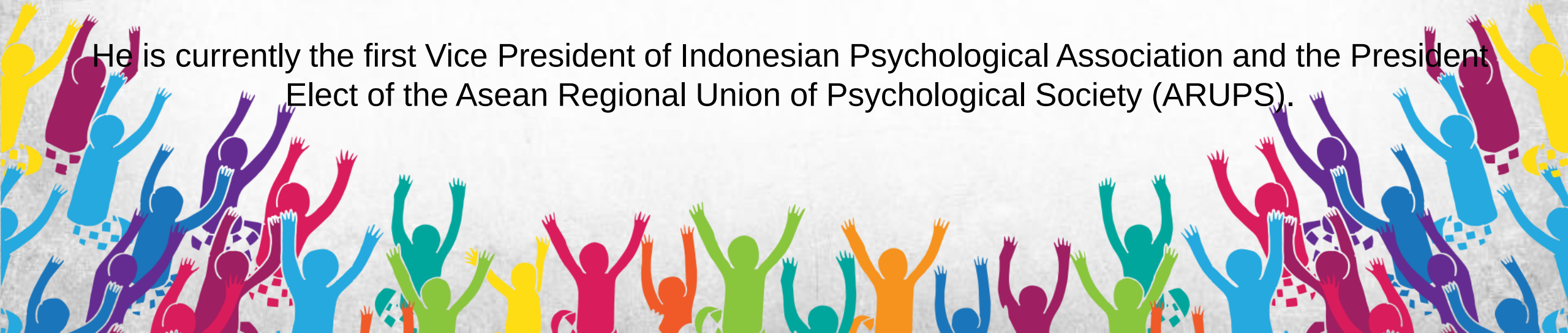
Using Generalizability Theory to Investigate the Psychometric Property of an Assessment Center in Indonesia

40th INTERNATIONAL CONGRESS ON ASSESSMENT CENTER METHODS

BUILDING NEXT GENERATION LEADERSHIP

Urip Purwono received his Ph.D. (psychology) from the University of Massachusetts at Amherst, USA specializing in psychometrics and educational measurement/evaluation.
Founder of the Center for Psychometric Study, UNPAD, his methodological research interest includes test theory, test construction, test adaptation, and structural equation modeling. In addition, he also does many works related to cognitive and educational test development to be used in Indonesia's setting.

He is currently the first Vice President of Indonesian Psychological Association and the President Elect of the Asean Regional Union of Psychological Society (ARUPS).



40th INTERNATIONAL CONGRESS ON ASSESSMENT CENTER METHODS

BUILDING NEXT GENERATION LEADERSHIP

Using Generalizability Theory to Investigate the Psychometric Property of an Assessment Center in Indonesia

R. Urip Purwono, Ph.D.

Padjajaran University

Using

Generalizability Theory

to Investigate the Psychometric Properties of an Assessment Center Program

Urip Purwono, M.Sc., Ph.D., Psi.

The Center for Psychometric Study
Faculty of Psychology
Universitas Padjadjaran Bandung, Indonesia

Background

- Formally first introduced in 1990 by PT Telkom, Assessment Center method become increasingly popular in Indonesia nowadays
  - uses for many purposes
  - by many organization, both public and private sectors
  - has many variations
- However, research is lagged far behind the practices. Psychometric study is even much less so that an appeal to fill this gap is warranted
- Without research, it is difficult to perfect the method and to improve the service

Measurement

- "The process of assigning number to an attribute of an entity according to a certain rule" (Steven, 1951)
- Assessment Center fit with this definition. Using performance tasks,
  - It assigned number
  - To attributes of an individual (namely "competency")
  - The are rules in the assignment of the number (i.e. the "rubrics")
- Another form of measurement: Test, Objective Personality Assessment

Technical Quality Indices

- Paper and Pencil measures use many indices of its technical quality: reliability, test/item information function, item fit, person fit, standard error of measurement i.e.:
  - *What is the reliability of the test? How big is the SEM?*
  - *How is the Item Information Function looks like? (IRT). How about the in-fit and the out-fit (Rasch measurement)*
- What about Assessment Center? Is there any "number" we can use to indicate its technical quality?
- Fall under "reliability concerns" (General Session 1, Willian Byham)

Technical Quality Indices (Cont.)

- The issue of reliability will never be over – as long as we are using number
- Methods are continuously developed
- Cannot interpret the number if we cannot rely to the number
  - The ratings change as the time of assessment change
  - The ratings change from rater to rater
  - The ratings changes due to the exercises

Purpose

- To familiarize us with indices derived from Generalizability Theory framework that can be used as quality indices of an assessment center program
  - Consistent with the assertion of Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Brennan, 1991; Kane, 1982; Shavelson, Webb, & Rowley, 1989)
- To present the basic idea of the Generalizability Theory and its simplest model appropriate to be used in the context of Assessment Center
- To illustrate the use of the approach in the context of Assessment Center using a real Assessment Center data

Assessment Center is a complex performance based assessment

- It involves:
  - Standardized performance tasks
  - Criterion to use in rating candidate performances
  - Rubric
- The Complexity:
  - Multiple raters
  - Multiple dimensions

Indices of technical quality

- Classical Test Theory provides ways to estimate the amount of error contain in in every observed score
- Item response theory provide ways to estimate ability and assign score more accurately
  - The Test Information Function is the inverse function of Standard Error of Measurement
- Inter-rater reliability provide information on the variation of ratings of two or more raters to person performances

Indices of technical quality (Cont.)

- The three approaches yield undifferentiated error of measurement (Feldt & Brennan, 1989), thus provides too gross a characterization of the potential and/or actual sources of measurement error

- The G-Theory, on the other hand, G theory provide information on the sources of error variation, disentangles them, and estimates each one, assuming randomly parallel tasks sampled from the same universe

The Tenets of GT (Cont.)

- *Generalizability (G) theory* is a statistical theory for evaluating the dependability (or reliability) of behavioral measurements (Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Brennan, 2001; Shavelson & Webb, 1991).

- It is an application and extension of the Analysis of Variance procedures applied to solve measurement problem

- It is an extension of the Classical Test Theory

The Tenets of GT (Cont.)

- The building blocks are variance and covariance components for a universe of admissible observations (UAO).
  - In performance based assessment: *a sample of candidate's performance drawn from a complex universe defined by a combination of all possible tasks, raters, and measurement methods, etc.*
  - Tasks, raters, and methods are called "facet"
  - The simplest AC facets: Task and raters

The Tenets of GT (Cont.)

- The individual performances rating is seen as affected by sampling variability due to tasks, raters, and so forth (facets), and their combinations, providing estimates of the magnitude of measurement error in the form of variance components

- it provides a summary coefficient reflecting the "reliability" of generalizing from a sample score or profile to the much larger domain of interest. This coefficient is called a *generalizability* coefficient (G coefficient)

- We use the results of the G-studies to design a measurement procedure that minimizes error

Illustration

- Job target: High rank position in a Public Sector/Local Government office
  - Two exercises, with
  - four assessors

- Design: p x I x r

$$\sigma^2_{X_{pio}} = \sigma^2_p + \sigma^2_i + \sigma^2_o + \sigma^2_{pi} + \sigma 2^2_{po} + \sigma^2_{io} + \sigma^2_{pio,e}$$

Results

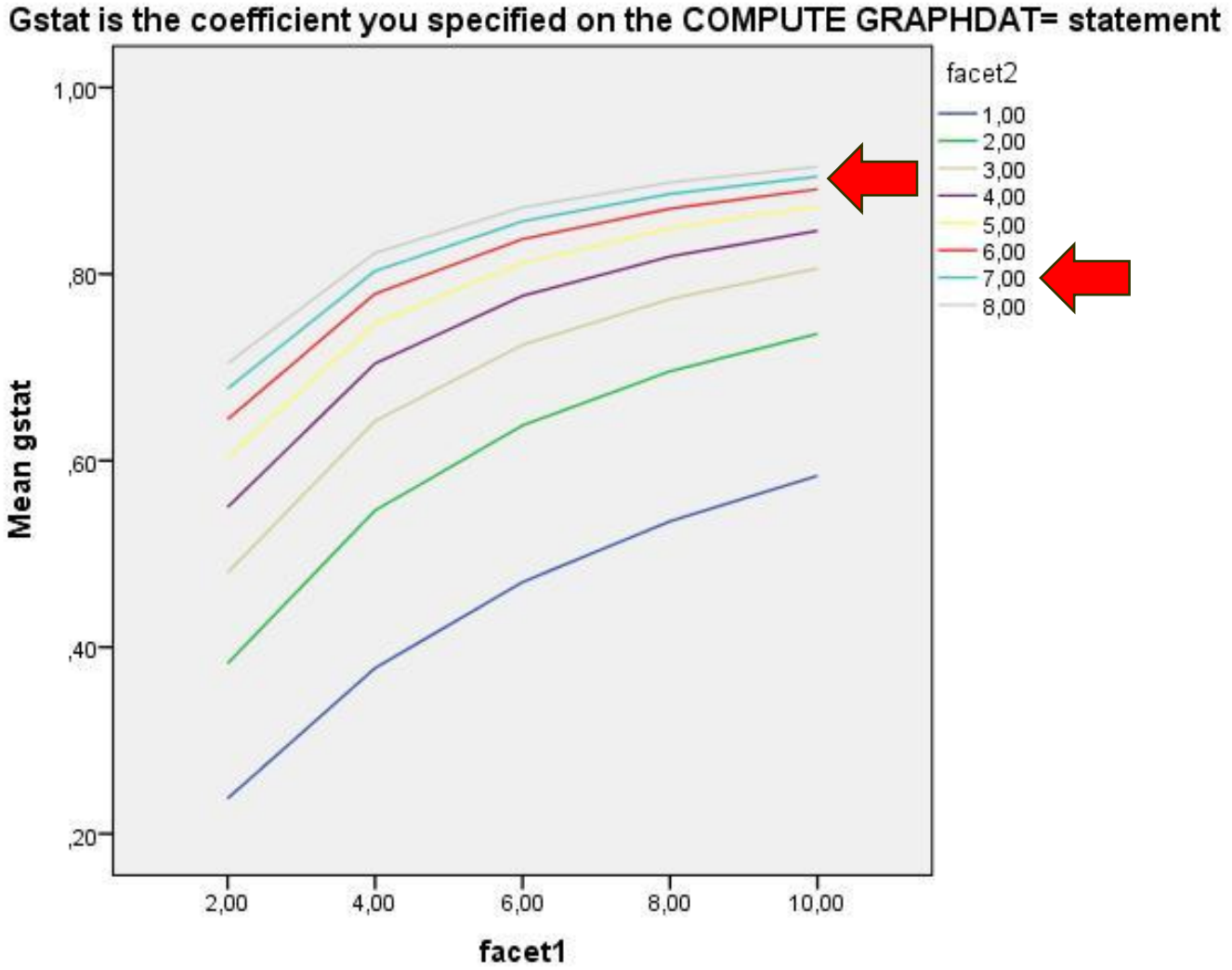
ANOVA Table

	df	SS	MS	Variance	Proport.
P	34,000	152,036	4,472	1,897	,137
F1	2,000	,804	,402	,082	,006
F2	3,000	4,811	1,604	,170	,012
P*F1	68,000	7,821	,115	,000	,000
P*F2	102,000	94,564	,927	,000	,000
F1*F2	6,000	3,211	,535	,000	,000
P*F1*F2	204,000	2395,664	11,743	11,743	,845

G-coefficients:

G	Phi
,564	,550

D-Studies


Gstat is the coefficient you specified on the COMPUTE GRAPHDAT= statement
Mean gstat
1,00
,80
,60
,40
,20
2,00
4,00
6,00
8,00
10,00
facet1
facet2
1,00
2,00
3,00
4,00
5,00
6,00
7,00
8,00

Validity

- The G and D coefficient does not speak to the issue of validity
- Modern theory of validity: Evidences need to be collected with regard to the interpretation and use of the score
  - Source of validity evidences: content, internal structure, relation to other variables, response process, and consequences of the installment of Assessment Center Program
- However, we cannot validate the method unless the results (numbers) have been proved to be reliable. The two goes hand in hand

Conclusion and Recommendations

- Generalizability Theory provides an appropriate framework to investigate the technical quality of Assessment Center, better than the CTT and IRT approach
- These investigation need to be adopted as a common practice if we are to perfect our Assessment Center method and provide a better and responsible service to the society
- It does not address the issues related to validity
- While large data is available, the association can take an initial step by setting up a Research and Development body in the organization

THANK YOU