American Journal of Agricultural and Biological Sciences 3 (1): 342-347, 2008
ISSN 1557-4989

Variables Refinery Process to Ensure Research Unbiasedness (Validity) and Invariance (Reliability) in Agricultural Extension and Education

Iraj Malek Mohammadi
Department of Agricultural Extension Education, College of Agriculture,
University of Tehran, Karaj, Iran

Abstract: Involving many variables, usually uncertain discrete, lowers the level of consistency and certainty in complex human studies like agricultural extension education. This research, as a part of a study entitled: Side by side situational analysis of current versus future situations of agricultural sector in manifesting the Ninth Malaysia Plan, is dealing primarily with the purpose and process of refining a set of uncertain independent variables to higher research validity and reliability. This investigation is quantitative in its nature, Applied in kind and design-wise, it is an Ex-Post-Facto analytical survey research. Eleven personal and professional characteristics of 224 agricultural experts were taken as independent and 158 attitudinal variables, measured based on dual (side-by-side) Likert scale, as dependent variables. Research population consisted of all agricultural experts including extension experts in Malaysia. As a result, 13 items (8.2% of the attitudinal questions) out of 157 were distinguished as impacted, moderate to very high sensitive variables to few personal characteristics and were eliminated in further data analysis. Applying this variable refinery method considerably increased construct and content validity as well as reliability of the research instrument and helped researcher to construct a rather impact-free side-by-side questionnaire.

Key words: Variables refinery, variable reduction, validity and reliability, impact-free questionnaire

INTRODUCTION

Concomitant with score interpretability is the notion that including only carefully crafted items on a test the primary method by which the skilled test developer reduces unwanted error variance, or errors of measurement and thereby increases a test score's reliability[14].

Social as well as behavioral researchers, in almost all academic research (i.e. theses and dissertations) and scientific studies from one hand, deal with multidimensional research or statistical population and have to modify this population to specify the limits at which generalization is being applied and from the other hand deal with many discrete variables to be identified as independent variables with uncertain impact on dependent variable(s). Therefore, a considerable part of the research report is being devoted to statistics describing research population without coming to clear conclusions as:

- Unbiased variables, especially when many variables are being measured as independent variables in the study
- Refining sensitive variables to some population characteristics
- Generalizability of findings by referring to no sensitive independent variables to population characteristics,

These shortcomings of human studies along with the qualitative nature of data that social and behavioral researchers are dealing with create a situation under which this kind of research is being accredited lower than quantitative, specifically true experimental, research. The aforementioned shortcomings are certainly evitable although the qualitative nature of human studies is not.

As Osterlind mentioned; interpretability of a test's scores flow directly from the quality of its items and exercises. If a test's scores are to yield valid influences about an examinee's mental attributes, its items must reflect a specific psychological construct or domain of content[14].

The question being answered in this article is that how can a social researcher come up with certain and bias- free independent variables out of a long list of directly or indirectly influential variables (generally discrete independent variables)? The purpose being pursued in this article is examining an innovative variable refinery process to identify bias-free variables through investigating their probable sensitivity to population characteristics.

To fulfill this purpose requires some theoretical and conceptual basis therefore, limitations along with validity and reliability concepts of research is being discussed bellow, due to the fact that this refinery process is in close conjunction with increasing research validity and reliability feature.

Defining the limits of research is one of the major issues in a research project. Three limits must be considered. Firstly, the subject should be clearly identified to some critical limits that shows what the researcher is dealing with and in which area of specialization he/she is conducting the research. Secondly, the Geographic boundaries dealing with the specific environmental and surroundings of the research should be recognized as; international, national, regional, local, or spatial. Thirdly, the Time either as the duration of study or a specific period of time (date).

What remains is what is often neglected in research method literature, in terms of the population limits, by which the research population is being modified i.e., target population and survey population. A target population is the population outlined in the survey about which information is to be sought and a survey population is the population from which information can be obtained in the survey[12].

The population can be defined as any set of persons/subjects having common observable characteristic[18]. A statistical population consists of the complete set of values that could result from a single measurement.

For a research to be accurate, within the limits of the study, its findings must be reliable and valid. Reliability and validity, these related research issues ask us to consider whether we are studying what we think we are studying and whether the measures we use are consistent[5].Validity is the generalizability of research findings[16], or the sense of unbiasedness where as reliability is the sense of unity. Validity is defined as: the degree to which the data support the inference that are made from the measurement[6], degree to which a procedure measures what it is supposed to measure[16], the degree to which our test or other measuring device is truly measuring what we intend it to measure[4].

In this sense, validity refers to the accuracy of a measurement. A measurement is valid when it measures what it is supposed to measure. Therefore, a basic question to answer through validity investigation is: Does an indicator accurately measure the variable that it is intended to measure?[6]. Other words, validity can be interpreted as the truthfulness of findings.

Statistically, the question being asked is: Are the variables under study related? or Is variable A correlated (does it covary?) with variable B? If a study has valid conclusion, we should be relatively certain that the answer to these questions is yes[18].

Apparently, there are many different threats to validity, but an important early consideration is to ensure internal validity. This means that we are using the most appropriate research design for what we are studying (experimental, quasi-experimental, survey, qualitative, or historical). It also means that we have screened out spurious variables, as well as thought out the possible contamination of other variables creeping into our study[8]. According to Oscher, concurrent validity is correlate test scores with criterion scores obtained at about the same time. The ability of a measure to indicate an individual's present standing on the criterion variable[13]. And as Garson expressed; Content validity, also called 'face validity', has to do with items seeming to measure what they claim to. Studies can be internally valid and statistically valid, yet use measures lacking face validity[1].

In addition to the validity of the research and fundamentally, research instrument, on the other hand, research findings should be reliable i.e. consistently the same if the study were done over and over again. Reliability of research essentially, as Kelly stated, is the degree of consistency within the measurement[6]. Moreover, reliability refers to the replicability of the research. The reliability of the research is assessed qualitatively by scrutinizing the design and methodology employed in the research. For research to be reliable it must be replicable[2]. Embodied in this citation is the idea of replicability or repeatability of results or observations.

Kirk and Miller identify three types of reliability referred to in quantitative research, that relate to: (1) the degree to which a measurement, given repeatedly, remains the same (2) the stability of a measurement over time and (3) the similarity of measurements within a given time period [7].

Joppe in Golafshani, defines reliability as: The extent to which results are consistent over time and accurately represent the total population under study is referred to as reliability and if the results of a study can be reproduced under a similar methodology, then the research instrument is considered to be reliable[3].

It is indicated by the Department of Psychology, Georgetown University that, for all forms of reliability, a quantitative measurement of reliability can be used, applied much like the inter-observer reliability coefficient. It should be .80 or higher. However, the coefficient can be lower for averages in a group because individual scores vary[10]. Kelly recommends that to measure the inter-item reliability (internal consistency) for multiple items used to measure a concept and comparing all possible combinations of these items, the Cronbach's Alpha statistic which is the average inter-item correlation for the set of items[6] should be calculated.

It is believed that anything we do to standardize or clarify our measurement instrument to reduce user error is believed to add to its reliability[11].

Salvucci, believe that the argument of some scholars as; the traditional view of reliability being a necessary but not a sufficient condition of validity, is incorrect. They recognize this school of thought as conceptualizing reliability as invariance and validity as unbiasedness. Moreover, a sample statistic may have an expected value over samples equal to the population parameter (unbiasedness), but have very high variance from a small sample size. Conversely, a sample statistic can have very low sampling variance but have an expected value far departed from the population parameter (high bias). In this view, a measure can be unreliable (high variance) but still valid (unbiased)[15].

Although, some scholars consider reliability and validity as separate issues, but as Trochim says; in fact, they are two related concepts. Therefore, we cannot assume validity, regardless of the reliability of our measurements. That is; a test may be valid (unbiased) but it may not necessarily be reliable (invariance). Conversely, a measure can be reliable (low variance) but still invalid (biased). Therefore, although a researcher may have already adopted a validated research instrument and assumes no need to check the reliability and/or validity of the data being collected, the internal consistency of the data may be threatened by different factors[17].

But, Mitchell and Jolley believe that reliability does not guarantee validity and it is only a prerequisite for validity[9].

Considering the above explanations about validity and reliability and while there is no mathematical procedure and or quantitative technique to measure different types of validity of the research instrument and at the same time what is applicable to calculate reliability coefficient of a test (Cronbach Alpha and/or Kudar Richardson formulas) is suitable for score data (interval and ratio), this research was designed to respond to the question of: other than determining personal and professional characteristics of respondents to identify the research population and allocate a considerable part of the research report to it, can we use these data to decrease variability and increase homogeneity of the variables to have a more efficient research instrument?.

For this very purpose, given that the research instrument is highly reliable, in terms of inter-item reliability (internal consistency); there was no guarantee of sufficient findings when personal characteristics of respondents (considered as independent variables) could be determinants of the responses to some degree. In order to find out the cruciality of each one of these variables creeping into the responses received, the researcher found out how sensitive or dependent were the dependent variables of the study on characteristics of the respondents. Other words, whether they responded the same to the questions?

Taking advantage of these notions (validity and reliability) it can be concluded that in making highly strategic decisions a triangulation of validity, reliability and efficiency should be employed to make decisions more valid, i.e. non-biased, more reliable, i.e. Generalization and more efficient, i.e. Trustworthiness This strategy is called variable refinery in this research and dealt with these three aspects of strategic decision making through an innovative process as described below.

To investigate the possible dependency or contamination of responses on personal characteristics of the respondents and to identify out those impact-free and/or non-biased items in the questionnaire to generalize the findings to the whole population Variable Refinery was applied in this research.

To find the sensitive cases out and exclude them of the main study, personal characteristics of the respondents were tested against each one of the dependent variables (DV). For the sack of unification, it was decided to keep those variables not statistically sensitive (dependent) to the independent variables (personal characteristics) and exclude (refined) those with frequent sensitive (contamination) cases from the further data processing stage.

MATERIALS AND METHODS

The task of constructing good test items is difficult because writing precisely and succinctly is challenging[14]. This study as a part of the research entitled: Side by side situational analysis of current versus future situation of agricultural sector in manifesting the Ninth Malaysia Plan, is dealing primarily with the purpose and process of lowering the variance of a set of measurement and consequently, elevating the level of its reliability by eliminating contaminated responses.

This investigation is quantitative and Casual in its nature; Applied in type and Ex-Post-Facto in design.

A side-by-side research instrument, to measure current versus future situation of the dependent variables was designed to squeeze the questionnaire from one hand and encourage the attendants to enjoy responding to the questionnaire on the other, after it was subjected to the aforementioned refinery process. Eight out of 158 questions were open-ended (basically personal characteristics) and 150 were closed-ended (basically attitudinal questions). Although Malaysia is a bilingual country, the questionnaire was written in Malay Language for optimal understanding by respondents.

Close-ended questions dealt with attitudes of Malaysian agricultural experts and comprised 147 variables in turn. A five set of measurement scale (1 = very low important to 5 = very high important), along with an option for null responses (not sure), were assigned to both sides of the questions as shown in Fig. 1.

Level of importance (Current situation)						Criteria	Level of importance (Future situation)					
5	4	3	2	1	0		0	1	2	3	4	5
						Agricultural contribution to the country's development						
						Reinforcement toward agricultural sector in RMK9						
						Allocation of Agricultural Development funds						
						Agricultural Development Implementation Programs						
						Reevaluation of Agricultural Development Programs						
						Etc.						

Fig. 1: Side by side questionnaire to measure present versus future importance of the Agricultural Sector and Agricultural Extension in Malaysia, 0 = No idea, 1 = very low importance, 2 = Low importance, 3 = Moderate importance, 4 = High importance and 5 = Very high importance

To ensure its content and face validity, the research instrument was reviewed several times by the research group and then implemented in a pilot test to measure its reliability. Cronbach's Alpha was calculated for the three sets of questions with ordinal scales

To come up with impact free items in the questionnaire, 11 personal and professional characteristics of 224 respondents (agricultural experts) were taken into account as independents variables (IV) and 157 attitudinal variables (questions) as dependent variables (DV). Over 1,700 statistical tests .i.e., non-parametric tests, were applied based on the nature of the variables scale and level (number of groups compared) to find out the significance of eventual dependency (sensitivity) of the dependent variables on the independent variables.

The research population consisted of 384 agricultural experts including extension experts and agents in Malaysia attended at the Biannual Agricultural Experts Conference in Johur Bahru (Sothern Malaysia). All attendants received the questionnaire at the time they registered for the conference. Two hundred and twenty four of 384 questionnaires were received back within two weeks after the conference (58%). $SPSS_{Win13}$ was used to process the data collected.

RESULTS AND DISCUSSION

The outcomes of calculating Cronbach's Alpha reliability coefficient for all variables, each with three corresponding sets of items, revealed that $\alpha = 0.968$ (for a set of 36 items), $\alpha = 0.965$ (for a set of 48 items) and $\alpha = 0.871$ (for a set of 12 items), while the whole set of items revealed $\alpha = 0.978$ (for 96 items). Therefore, variables showed high reliability, as well as the research instrument as a whole.

As indicated earlier, there were 11 IVs versus 157 DVs. Results from applying 1727 (11x157) different statistical tests in the refinery process was assigned an X on the related cell in the 11 by 157 refinery matrix when each statistical test was statistically significant. Number of Xs then was calculated as the times each specific DV was statistically sensitive to IVs in this research. Consequently, summing up the number of significant tests for DVs and IVs showed the number of each IV having a significant impact on the DVs.

Following this procedure, the DVs were categorized in Table 1, based on their sensitivity level as; 0X = no sensitivity, 1X = very low sensitive,

Table 1: Distribution of the DVs based on the level of their sensitivity to IVs

Sensitivity level	Frequency	%	Cumulative %
None	76	48.4	48.4
Very low	41	26.11	74.51
Low	27	17.2	91.71
Moderate	4	2.54	94.25
High	4	2.54	96.8
Very high	5	3.2	100
Total	157	100	

0X = no sensitivity, 1X = very low sensitive, 2X = low sensitive, 3X = moderate sensitive, 4X = high sensitive to 5X and above = very high sensitivity

Table 2: Frequency of DV contamination by each IV

IV	DV frequency	%
Age	14	9.1
Sex	48	31.2
Race and Ethnic	8	5.2
Marital status	17	11.1
Religion	9	5.8
Higher education	29	19
Current occupation	6	3.8
Years of service (tenure)	12	7.8
Years of extension service	6	3.8
Organization	0	0
Expertise	5	3.2
Total	154	100

Table3: Reliability coefficient before and after variable refinery process

	Refinery stage			
Reliability				
Reliability		Partial		Overall
Before	0.87 (12 cases)	0.968 (36 cases)	0.96 (48 cases)	0.978 (96 cases)
After	0.87 (12 cases)	0.926 (32 cases)	0.956 (39 cases)	0.973 (83 cases)

[a]: Cronbach alpha

2X = low sensitive, 3X = moderate sensitive, 4X = high sensitive to 5X and above = very high sensitivity.

To refine the huge number of discrete variables involved, dependent variables with moderate sensitivity and above [in total 13 items out of 157 items (8.2%)] were excluded to avoid biasness in the inferential part of the study when generalization was intended in the main research.

Summing up the number of significant tests for DVs and IVs showed the number of each IV having a significant impact on the DVs as appeared in Table 2.

As indicated in this Table; sex, higher education and marital status were the most influential independent variables over dependent variables in this study, respectively.

Cronbach's Alpha Coefficient was calculated again after eliminating 13 contaminated (sensitive) items from the list of 96 items eligible in reliability testing. Results revealed that; α= 0.926 (for a set of 32 items), α= 0.956 (for a set of 39 items) and α=0.87, (same as before for a set of 12 items) and for the whole set of items revealed α= 0.973 (for 83 items),. Again, all of the variables showed high reliability. Results from both stages are presented in Table 3.

Comparing two sets of reliability coefficients (before and after variable refinery process) in Table 3, shows almost no statistically significant difference in the alpha scores. Although, mathematically speaking, the coefficients slightly decreased after the refinery process, while holding everything constant other than eliminating contaminated variables in the reliability calculations.

It can be concluded from what is presented so far, that:

- Although validity and reliability are interconnected concepts, increased notions validity does not necessarily mean higher reliability.
- Validity is really a qualitative issue in research and cannot be interpreted mathematically.
- Making decisions between reliability and validity to increase confidence in research findings is an interpretive and qualitative matter, rather than exclusively being based on a researcher's tact. Accordingly, it can be considered as a strategic issue indeed when we manipulate the number of variables, as the result of contamination by other variables (personal characteristics), therefore reliability may even decrease for the sake of validity. In a case like this a researcher may be confused by having two measures of reliability. The determinant for him in this situation to make the appropriate decision is validity i.e., higher validity should be preferred over higher reliability as was experienced in this research.
- For the researcher to be more certain about the validity and reliability of his/her attitudinal research instrument, he/she may hypothesize some notions about the impact or effects of the personal and professional characteristics of his respondents on their attitudes being investigated in that research. If this was the case, then he may phrase those notions as refinery hypotheses of the research rather than the research hypotheses percy. Herewith, the researcher can categorize two sets of hypothesis as; refinery hypotheses (to sustain validity and reliability) and research hypotheses to study the impact or effect of the independent variable/s on the dependent variable/s. Respecting this conclusion, in the current study, reliability coefficients in the last row of Table 3 were preferred for sake of the higher validity even

though they were a little bit lower than the coefficients in the second row of the same Table.

ACKNOWLEDGEMENT

Assistance in data collection and providing facilities during the study by the Center for Extension, Entrepreneurship and Professional Development (CEEPD), Department of Professional Development and Continuing Education and Faculty of Educational Studies, University Putra Malaysia (UPM) is gratefully appreciated.

REFERENCES

1. Garson, D., 2007. Validity. In: Quantitative methods in public administration. Accessed: www2.chass.ncsu.edu/garson/pa765/validity.htm. Jan. 2007.
2. Geolog, 2006. Validity, reliability and objectivity. In: Experimental research. Retrieved [August 11.2006] from; w.geolog.com/rm/rmstart.htm.
3. Golafshani, N., 2003. Understanding reliability and validity in qualitative research. The Qualitative Report: 8 (4): 597-607.
4. Heffner, C., 2004. Research methods. All Psych. Heffner Media Group Inc. Florida, USA.
5. Howell, J. *et al*., 2005. Reliability and validity. Writing@CSU. Colorado State University. Department of English. Colorado. USA.
6. Kelly, D., 1999. Measurement made accessible: A research approach using qualitative, quantitative and quality improvement methods. Sage Publication Inc. London, UK.
7. Kirk, J. and M. Miller, 1986. Reliability and validity. In: Qualitative research (Qualitative Research Methods). SAGE University Press, London. UK.
8. Megalink, 2004. Measurement, Reliability and Validity Syllabus for JUS 308. MegaLinks. In: Criminal Justice. North Carolina Wesleyan College. Rocky Mount. NC. USA.
9. Mitchell, M. and J. Jolley, 2004. Research design explained (with InfoTrac), 5th Edition. Thompson Higher Educ. Belmont, C.A. USA.
10. NN., 2005. Validity and reliability. In; Research methods and statistical resources. Department of Psychology, Georgetown University. Washington D.C. USA.
11. O'Connor, T., 2006. Lecture notes for research methods. MegaLinks. In: Criminal Justice. North Carolina Wesleyan College, Rocky Mount, NC, USA.
12. Organization for Economic Co-operation and Development (OECD), 2005. 'Statistical population. In: Glossary of Statistical Terms. OECD. Paris, France.
13. Oscher, J., 2006. Qualitative research, part 1. Longwood University Web Site. Virginia. USA.
14. Osterlind, J., 2004. Constructing test Items: Multiple-Choice, Constructed-Response, Performance and Other Formats (Evaluation in Education and Human Services). 4th Edn. Kluwer Academic Publisher Group. Dordrecht, Netherlands.
15. Salvucci, *et al*., 1997. Measurement error studies at the National Center for Educational Statistics. Department of Education. Washington D.C. USA.
16. Sommer, R. and B. Sommer, 2002. A Practical guide to behavioral research: tools and techniques, 5th Edn. Oxford University Press, NC. USA.
17. Trochim, W., 2002. 'Reliability and validity' In: Research Methods Knowledge Base. 2nd Edn. Atomic Dog Publishers. Cincinnati, Ohio. USA.
18. Woolf, L. (ND)., 2006. Research methods. Webster University. St. Louis, Missouri. USA Retrieved from<http://www.webster.edu/~woolflm/statmethods.html> on August 8, 2006.