



WCETR 2013

Determining Validity and Reliability of Data Gathering Instruments Used By Program Evaluation Studies in Turkey

Sefik Yasar ^a *, Asli Gundogan Cogenli ^b

^aAnadolu University, Faculty of Education, Eskisehir, Turkey

^bUsak University, Faculty of Education, Usak, Turkey

Abstract

The purpose of this study is to analyze the validity and reliability of inventories that were used in program evaluation studies realized in Turkey. In accordance with this purpose 12 articles and 38 theses have been analyzed within the scope of the research. According to the findings obtained in the research, it has been determined that regarding validity study in the inventories used in program evaluation studies in Turkey mostly content validity has been furnished; that regarding reliability study mostly the Cronbach Alpha coefficient has been calculated; and that in addition to them the inventories about which no validity and reliability study has been conducted are many in number, as well. Finally in the research a conclusion has been reached that validity and reliability studies of the inventories used in program evaluation researches in Turkey are not adequate.

© 2014 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

Selection and peer-review under responsibility of the Organizing Committee of WCETR 2013.

Keywords: Program evaluation, inventories, validity, reliability.

1. Introduction

The rapid changes in science and technology in the recent years have also affected the speed of information increase and facilitated formation of the knowledge society where new knowledge is generated every second. From these developments education programs have been affected as well and in the program development process particularly raising the individual type which 21st century needs has been intended for. Within this context, the

*Corresponding Author: Sefik Yasar. Tel.: +090-0533-312-9280

E-mail address: sefik.yasar@gmail.com

effort to adapt to the 21st century has brought along continuous development and taking up in detail of the programs with it.

Program evaluation constitutes the last phase of program development studies. Evaluation is to render judgments about the value of whatever is being evaluated (Fitzpatrick, Sanders and Worthen, 2004). Program evaluation is a process of data gathering through various measurement instruments about program effectiveness, comparing the obtained data with the criteria which are the indicators of program's effectiveness and making comments and decision making regarding effectiveness of the program (Erden, 1998). Validity and reliability of the decisions to be taken at the end of this process is directly related to validity and reliability of program evaluation process, as well (Saglam and Yuksel, 2007).

Reliability means that scores from an instrument are stable and consistent. Validity, however, means that the individual's scores from an instrument make sense, are meaningful, and enable you, as the researcher, to draw good conclusions from the sample you are studying to the population (Cresswell, 2005). When one thinks that validity and reliability of measurements to be made in a scientific research form the basic characteristics of the data gathering instrument (Ural and Kilic, 2006), it becomes clear that validity and reliability studies conducted have significant importance.

In program evaluation process, scientific studies and standards should be used as base and base of evaluation process should be formed by scientific methods (Yasar and etc., 2005). If no valid and reliable criterion is used in this process, the comments derived from research results become subjective and suspicious, too (Turgut, 1983). This situation causes questioning of being scientific of the research conducted. Within this context, using data gathering instruments that have been proven to be valid and reliable in program evaluation studies is an important factor from the perspective of enhancing the quality of the research and facilitating being scientific. It is known that after putting into effect of new educational programs in 2005 in Turkey, program evaluation studies have increased in numbers. But whether these research are adequate from the point of quality or not is a matter of discussion. Based on this, the fact that researches conducted related to program evaluation are many in number has given birth to the need for determining whether or not they are qualified. Likewise, because determining validity and reliability studies conducted on data gathering instruments used in research realized would provide information about being scientific of the research in question, they are considered important. Moving from these requirements, in this study it has been aimed to analyze the data gathering instruments used in program evaluation studies conducted at the primary education level in Turkey from the perspective of validity and reliability studies.

In the research answers for the following questions have been sought:

- In regards to inventories used in the analyzed researches which validity studies have been made ?
- In regards to inventories used in the analyzed researches which reliability studies have been made ?
- In regards to inventories used in the analyzed researches what are the validity and reliability studies conducted together?

It is considered that research findings will draw the attentions to the importance of securing validity and reliability of measurement tools in the program evaluation studies, shed light on to the researchers who will conduct program evaluation studies and that validity and reliability studies to be conducted meticulously will enhance the quality of studies made.

2. Method

In the research conducted by document review, 12 articles and 38 master thesis and PhD dissertations that were conducted on program evaluation between 2007-2012 in Turkey have been analyzed. For article review, reachable editions of various journals with referees have been analyzed and in the search engines searches have been performed with the key words "program evaluation". For master thesis and PhD dissertations the thesis in the National Thesis Center of Higher Education Council have been analyzed. The data gathering instruments analyzed within the scope of the research are limited to pre-developed inventories or inventories developed by researchers.

In order to provide internal validity for the research, the research report has been presented a field specialist and a qualitative research specialist. In line with opinions and suggestions of the specialists necessary corrections have been made in the research report. In order to enhance reliability of the research, in the data analysis process, two researchers have analyzed the documents together about the research questions and have created Table 1.

3. Findings

The validity and reliability studies realized in regards to inventories used in the research analyzed within the scope of the study have been presented in Table 1. As seen in Table 1, majority of the inventories used in the program evaluation studies conducted in Turkey have been developed by the researchers. Number of pre-developed inventories used in these studies on the other hand is quite a few.

3.1. Validity Studies Conducted Related to Inventories Used in the Researches Reviewed

The validity studies used in the reviewed researches have been determined as content validity, construct validity, criterion-related validity and face validity. Content validity is the assessment of whether or not the items are adequate in terms of quantity and quality to measure the characteristics that are wanted to be measured (Buyukozturk, 2011). Content validity requires cooperation with the field specialist and is determined by consulting the specialist (Tavsancil, 2006). When one looks at Table 1 it is seen that content validity has been achieved in total of 25 researches, seven of which are articles and 18 of which are thesis. It is seen that in order to attain content validity, draft scale items are presented to the evaluations and suggestions of academic staff and teachers working at Education Faculties of state universities in Turkey [A(3,4)-T(9,15,20)]. Besides, consultants in the Board of Education and Discipline have been consulted, as well [A(4,10)].

Table 1. Validity and reliability of inventories used by program evaluation studies

Data Gatherin g Instru ments	Develope r	Validity				Reliability				Unreport ed
		V ₁	V ₂	V ₃	V ₄	R ₁	R ₂	R ₃	R ₄	
	Previousl y develope d	A10	A10 T14					A10 T (13,14)		A(6,9) T(1,23)
Inventory		A(2,3,4,5 ,8, 11)	A(2,3,4 ,5,12)	A(2,1 1,12)		T(2 2,32)		A(1,2,3,4 ,5,8, 11)		A7
	Develope d by researche r	T(3,7,8,9 , 12,15,20, 24,25,26, 28, 29,32,33, 34,36,37, 38)	T(2,4,7 , 22,27,3 6, 37)	T 4	T15	T 4		T(2,4,5,1 5,18,21,2 2,24,26,2 7,28, 29,30,31, 34,35,36, 37, 38)	T18	T(6,10,1 6, 17,19)

V₁: Content Validity, V₂: Construct Validity, V₃: Criterion-related Validity, V₄: Face Validity, R₁: Test-retest Reliability, R₂: Split-half Reliability, R₃: Cronbach Alpha Reliability, R₄: Item-total Correlation, A.:Article, T.: Thesis.

Construct validity is related to validity of inferences made about the unobservable variables which cannot be observed by observed variables (Cokluk, Sekercioglu ve Buyukozturk, 2010). When the Table 1 is looked at it is seen that construct validity has been attained in total of thirteen researches, six of which are articles and seven of

which are thesis. Regarding the attainment of construct validity in all of the researches in question KMO coefficient and Barlett test value have been calculated and exploratory factor analysis has been made. The fact that confirmatory factor analysis is not used in any of the research is an attention grabbing point. Criterion-related validity is comparing the results obtained by the developed test or scale to the points belonging to a measurement criterion determined as standard and having a high correlation coefficient as a result of this comparison (Sencan, 2005). It is seen that in only one of the reviewed researches [T4] criterion validity has been used. In the research where primary education 8th grade Mathematics Teaching Program was evaluated in line with teachers' opinions, a pre-developed scale validity and reliability studies of which have been made and the scale which was developed within the scope of research have been applied to the same group and correlation between results taken from the two surveys has been checked. The correlation between the two scales has been calculated as .73 and it has been stated that criterion validity related to developed scale has been achieved. Finally the face validity which is related to name of the test, explanations, order or the questions and test etc. causes the answerer to develop an opinion regarding general validity of the test and it is evaluated by specialist's opinion (Buyukozturk, 2011). Face validity has been achieved in total of 4 researches, 3 of which are articles and 1 of which is thesis. In the researches in question it has been stated that about the achievement of face validity opinions of specialist academicians have been benefitted from.

3.2. Reliability Studies Conducted Related to Inventories Used in the Researches Reviewed

The reliability studies used in the reviewed researches have been determined as test-retest method, split-half method, Cronbach Alpha coefficient and item-total correlation. In the test-retest method the inventory is applied to the same group again after a while and relation between two applications is found (Balci, 2006). It is seen that in only one of the reviewed researches [T4] test-retest method has been used. The test-retest reliability of the inventory used in the research in which primary education 8th grade Mathematics Teaching Program was evaluated in line with teachers' opinions has been applied to 12 volunteer teachers with three weeks time interval and the correlation coefficient between the first and second application has been calculated to be .72. As a result of this operation it has been stated that inventory had determination power. Split-half method is used in the reliability calculations when the test measures the same and single variable (Erden, 1998) and it shows the consistency among the obtained test points (Buyukozturk, 2011). In the two of the reviewed researches [T(22,32)] it has been determined that the reliability was calculated with this method. Split-half reliability coefficient of the inventory used in a thesis where Religion Culture and Ethics Program was evaluated has been found to be .97. In the other thesis [T32] where Primary Education Mathematics Teaching Program was evaluated however, inventory has been applied to the groups excluding the research sample and it has been stated that by half way through test method reliability coefficient was calculated. Additionally, it has been stated that in order to calculate the reliability of the entire inventory Sperman-Brown formula was used and the value was approximately .93. Cronbach Alpha is used to test for internal consistency. If the items are scored as continuous variables (e.g., strongly agree to strongly disagree), the alpha provides a coefficient to estimate consistency of scores on an instrument (Creswell, 2005). As it is seen in Table 1 the Cronbach Alpha internal consistency coefficient has been used in total of 28 studies, 8 of which are articles and 20 of which are thesis. In some of the reviewed researches dimensions of the used scale and Cronbach Alpha value of the entire scale have been calculated separately, in some researches however a single value pertaining to entire scale has been calculated. Finally, the correlation analysis between total point and item points is for determining the reliabilities of items (Sencan, 2005). Only in one of the reviewed research [T18] it has been determined that reliability was calculated by item total point correlation but it has been seen that no detail regarding performed calculations was provided.

3.3. Validity-Reliability Studies That Are Conducted Together Related To Inventories Used In Reviewed Researches

When Table 1 is analyzed, in majority of the articles analyzed content and construct validity related to validity study have been achieved and Cronbach Alpha value related to reliability study has been calculated [A(3,4,5,10)]. In three of the researches in question however any of the validity and reliability studies have not been conducted [A(6,7,9)]. In the M6 coded article the Cronbach Alpha value of the original form of the scale has been reported to

be .98 .

When the thesis in the scope of research are analyzed, content validities related to validity studies of seven of them have been attained and Cronbach Alpha value related reliability study has been calculated. In the T37 coded thesis however both content and construct validities related to validity construct have been attained and Cronbach Alpha value related to reliability study has been calculated. Notwithstanding this, in the T4 coded thesis construct and criterion validities related to validity study have been attained and test-retest method related to reliability study and Cronbach Alpha coefficient have been used. In the T22 coded thesis construct validity related to validity study has been attained, split-half method related to reliability study and Cronbach Alpha coefficient have been used. Likewise, there are also studies in which only Cronbach Alpha value [T(5,11,13,21,30,31,36)] and only content validity [T(9,12,20,24,33)] have been attained. In the seven of the researches in question any of the validity and reliability studies have not been conducted [T(1,6,10,16,17,19,23)]. In the T1 coded research among them the researcher has reported the Cronbach Alpha value of the original form of the scale and in the other thesis however no explanation has been made related to validity and reliability study.

4. Conclusion and Recommendations

Program evaluation is a decision making process regarding whether the program has become effective or not. According to the decision made program can continue to be applied or it can be developed. In order for this decision to be objective and right the program evaluation process should be carried out in accordance with scientific standards. And this is realized through the use of valid and reliable measurement instruments in program evaluation process.

According to the findings obtained in the study, it has been determined that in the studies in Turkey mostly content validity and construct validity related to validity study are used. It has been identified that face validity and criterion validity on the other hand are not much preferred. It has been determined that Cronbach Alpha internal coefficient related to reliability study is quite frequently used. In addition to that, the conclusion has been reached that test-retest method, split-half method and item total point correlation are not widely preferred methods.

It has been determined that in some of the articles and dissertations analyzed within the scope of the research, content and construct validity related to validity study have been attained and Cronbach Alpha value related to reliability study has been calculated. Besides, the number of researches in which only Cronbach Alpha internal coefficient was calculated or only the content validity was attained is quite many. Besides, it has been determined that there are also articles and thesis for which none of the validity and reliability studies have been conducted.

As a result, it can be said that in the program evaluation studies conducted in Turkey, measurement tool development principles are not adequately adhered to, validity studies are usually limited to specialist's opinion, reliability studies however are mostly in the shape of analyzing the internal consistency and in the sense of determination reliability studies cannot be adequately made.

Recommendations are as follows: 1) Measurement – assessment specialists in Turkey should develop valid and reliable inventories that can be used in program evaluation studies. These type of studies can make significant contributions to program evaluation studies. 2) The researchers who will develop measurement tools should be competent in the field and know about measurement tool development principles and techniques. About this matter seminars, in-service trainings etc. can be provided by specialist people. 3) Measurement tool development studies should be handled with more rigorously. Within this context, researchers can be suggested to hold up the qualified studies in this field as an example.

References

- Balci, A. (2006). *Sosyal bilimlerde araştırma: Yöntem, teknik ve ilkeler* (6th ed.). Ankara: Pegema Yayıncılık.
- Buyukozturk, S. (2011). *Sosyal bilimler için veri analizi el kitabı* (13rd ed.). Ankara: Pegema Yayıncılık.
- Creswell, J. W. (2005). *Educational research: Planning, conducting and evaluating quantitative and qualitative research* (2nd ed.). New Jersey: Pearson Education.
- Cokluk, O., Sekercioglu, G. & Buyukozturk, S. (2010). *Sosyal bilimler için çok degiskenli istatistik SPSS ve LISREL uygulamaları* (1st ed.). Ankara: Pegem Akademi.
- Erden, M. (1998). *Egitimde program degerlendirme* (3rd ed.). Ankara: Anı Yayıncılık.

- Fitzpatrick, J. L., Sandres, J. R. & Worthen, B. R. (2004). *Program Evaluation: Alternative approaches and practical guidelines*. (3rd ed.). Boston: Pearson Education.
- Saglam, M. & Yuksel, İ. (2007). Program degerlendirmede meta analiz ve meta degerlendirme. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 18.
- Sencan, H. (2005). *Sosyal ve davranısal olcumlerde guvenilirlik ve gecerlilik* (1st ed.). Ankara: Seçkin Yayıncılık.
- Tavsancil, E. (2006). *Tutumların ölçülmesi ve SPSS ile veri analizi* (3rd ed.). Ankara: Nobel Yayın Dağıtım.
- Turgut, M. F. (1983). Program degerlendirme. *Cumhuriyet Döneminde Eğitim I*. İstanbul: Milli Eğitim Basımevi.
- Ural, A. & Kilic, İ. (2006). *Bilimsel araştırma süreci ve SPSS ile veri analizi* (2nd ed.). Ankara: Detay Yayıncılık.
- Yasar, Ş., Gultekin, M., Kose, N., Girmen, P. & Anagun, S. (2005, Temmuz). *The meta-evaluation of teacher training programs for elementary education in Turkey*. ATEA: Australian Teacher Education Association 33. Annual Confrance. 498-504, Avustralya.