

Investigation of validity and reliability works in postgraduate mathematics theses that adopt qualitative research in Turkey: Functions of data collection tool

Özkan Ergene^{1a}, Esin Zelal Yazıcı², and Ali Delice²

¹Sakarya University, Faculty of Education, Hendek, 54300, Turkey

²Marmara University, Ataturk Faculty of Education, Istanbul, 34744, Turkey

Abstract. In this research, master and doctorate thesis which adopts qualitative method in mathematics education studies in Turkey are thoroughly inspected in terms of their validity and reliability, thus the research has a case study as a research model. From Thesis Web site by Council of Higher Education, 119 master thesis which are published and contains both primary and secondary mathematical studies and 27 doctorate thesis are analysed by using document analysis technique and data is resolved by means of content analysis. It was seen in almost more than half of these theses in which reliability and validity are done, that in addition to the credibility, transmissibility, consistency and confirmability strategies that is used for qualitative paradigm, Statistical calculations are also used in regards of quantitative paradigm. Also, in most of the scale development and adaptation works, language validity technique was not used and in almost half of them, it was observed that strategies such as test works were not employed. In this case, it shows that Reliability and Credibility works remains unfulfilled and awareness of analysis strategies that are caused by scale differences remains low.

1 Introduction

It is required that results and verisimilitudinous of the scientific research that opens a new path to new movements in times of paradigm changes, are questioned. This is identified as systematic data gathering process in an environment which is open to the development of knowledge. In contemplation of the fact that it is impossible to reach the ultimate reality, it is an acceptable fact that every single scientific research would have a certain amount of error. Generally, error is described as the difference between the observed value and

^a Corresponding author: ozkanergene@sakarya.edu.tr

measured value and it should also be addressed in researches of education field. In research period, error can be originated by individual's personal characteristics [2 and 5]. Errors in assessment are examined as constant, systematic and accidental. In assessment, constant error in which values in assessment remains same regardless of how many times it is done, and it always has the same impact. Systematic error on the other hand changes depending on the properties that are assessed, people that are involved and environment where the assessment is performed and these two errors effect the reliability of assessment [7]. Accidental error whose source, direction, size are random and intervenes with assessment results in a random way, poses a different situation compared to the other two and it affects not only reliability of assessment process but also the validity which is one of the qualifications that needs to be present in assessment tools.

The more an assessment is away from accidental errors the more reliable it is [2]. Also, reliability is more related to the data itself that is collected through certain tests, assessments and data collection tools than their content, shape and structure. Crocker and Algina [4] and Grounlund and Linn [8] expresses that reliability belongs to the data that is collected. From this perspective, reliability can be described as getting the similar results from the assessments in which same methods are used in different groups that belongs to the same environment [4 and 16]. In application of a reliable assessment tool, it can be observed that results overlaps even it is applied in different times. After the assessment procedure, the coefficient that is calculated stays between 0 and 1 and the closer this value to 1, the more reliable the assessment process becomes [1]. Through the whole education year, it is expected from reliability values to be as close to 1 as possible so that the accuracy of the results can be considered as secure. The inspection of reliability also helps studying of the validity. An assessment tool's degree of assessing intended feature is called validity of these tools in assessing [9]. In another words, validity is in what degree an assessment tool assessed its intended variable. In education, variable that is intended to be assessed is the success of the students thus attainment of the objectives. After the assessment process, validity value is valued between -1 and 1 and in the cases that number is over 0.30 and more, reliability of the assessments can be discussed [1].

In order to locate error's amount, its direction and to remove possible sources of error, it is required that validity and reliability of variables such as method of research, study group, data collection tools, data analysis and also the research itself are performed. Although its descriptions are varied among different researchers, validity, in its general view, can be described as how much it serves to the purpose of its measurement tools [17 and 18]. Reliability, on the other hand, can be described as getting the similar results from the different measurements in which same method is used in different groups that belongs to the same population [4]. Validity and reliability works, which shows differences according to the research being qualitative or quantitative, are important factors that should be emphasized meticulously [3]. In quantitative researches, while performing validity and reliability works, through strategies that is based on statistical calculations, results are generalized. However, certain strategies such as "credibility" for internal validity, "transferability" for external validity/generalizability, "dependability" for reliability and "confirmability" for objectivity are suggested to be used for qualitative researches on educational field, which focuses presenting a fact or a situation instead of focusing exact and explicit results [10]. In this context, in studies of educational field, different ways are followed that consist validity and reliability evidences. In this research, it is aimed to analyze studies on validity and reliability of the master theses and doctoral dissertations on the field of mathematics education which adopt qualitative approach in Turkey. In accordance with this purpose, answers are looked for to the following research questions:

- How are the validity works of post graduate theses on the field of mathematics education in Turkey which are done between 2005 and 2014 and present in the website of the Council of Higher Education (YÖK)?
- How are the reliability works of post graduate theses on the field of mathematics education in Turkey which are done between 2005 and 2014 and present in the website of the Council of Higher Education (YÖK)?

2 Methodology

To examine reliability and validity studies in theses that are carried out by using qualitative approaches in mathematics education in Turkey, theses, which are the main source itself and “printed document” [15], are searched by adopting case study model [3]. As qualitative and quantitative theses are effected by different paradigms, this research only contains the theses that are based on qualitative paradigm. Research data is gathered through document analysis which is one of the methods of qualitative data collection.

2.1 The aim of the research

In this research, it is aimed to examine psychometric features of master theses and doctoral dissertations on mathematics education in Turkey, which has qualitative paradigms. In correlation with this purpose, validity and reliability of master theses and doctoral dissertations on mathematics education in Turkey that were done between 2005 and 2014 by employing qualitative method are analyzed.

2.2 Study Group

The study group is consisted of 119 master theses and 27 doctoral dissertations which are published between 2005 and 2014 on primary school mathematics teaching, secondary school mathematics teaching and mathematics education studies are available upon request on the website of the Higher Education Council's website. Therefore, to create this study group, it can be stated that purposeful sampling technique is used [14].

2.3 Data Analysis

Validity and reliability of the theses that are used in this research are analyzed through content analysis method. First it was observed if these works are actually done. It was seen that in most (%73) of the theses, validity and reliability studies are actually performed. Later on, data gathered through these studies are coded, grouped under certain themes, presented with descriptive statistics and interpreted.

During the coding phase, if a study uses more than single strategy in terms of ensuring validity and reliability, it is covered in all aspects. For example, in one of the studies to ensure validity, both peer debriefing and data triangulation is employed and it is coded accordingly in both categories. To ensure reliability between interraters of this research, data is coded independently by researchers. Among these independent coding that is done, interrater reliability is found as %92.

3 Findings

As a result of analyzing master theses doctoral dissertations on mathematics education in Turkey, it was founded that more than half (%54,1) of these theses and dissertations are

done by researchers from Gazi University, Karadeniz Technical University and Marmara University. In %73 of these, it was found that at least one of the validity or reliability studies are performed. In the rest of them though, it was found that such studies are not performed. When method chapters of the former theses and dissertations are examined, it was found that %58 of the theses and dissertations mentioned this topic not as a separate part but within this chapter. In %27researches' validity and reliability, in %4 only researches' validity, in %2 only researches' reliability; in %6 tools' validity and reliability and in %3both researches' and tools' validity and reliability headlines were founded.

In %78 of the researches that has validity and reliability studies, credibility, transferability, dependability and confirmability strategies are used. Lincoln and Guba (1985), to establish credibility, suggest certain methods such as prolonged engagements, persistent observation, triangulation, peer debriefing/review and member checking; to ensure transferability, suggest thick description and purposeful sampling; to ensure dependability, suggest inquiry audit; to ensure confirmability, suggest confirmability audit, audit trail, triangulation and reflexivity. In the theses and dissertations studied, it was found that credibility, transferability, dependability and confirmability strategies were found under the headline of validity and reliability but evaluations were found that did not use any of these terminologies (%22).

When, methods that are used to ensure the validity of the researches in theses are analyzed, it was founded that data triangulation (% 45), which means using more than one type of data collection device, peer debriefing/review (%17) are found to be the mostly used methods. As seen at Table 1, these methods are followed by member checking (% 8), prolonged engagement (% 6).

Table 1. Methods to ensure the validity of the researches

Validity of research	Frequency	Percentage
Triangulation	49	45
Peer debriefing/review	19	17
Member checking	9	8
Prolonged engagement	7	6
Sampling strategies	5	5
Deep focused data collection	5	5
Thick description	3	3
Detailed report	3	3
Pilot scheme	3	3
Raw data	1	1
Direct citation	1	1
Data from natural environment	1	1
Using multiple case study	1	1
Confirmation audit	1	1
Chain of evidence	1	1

When, methods that are used to ensure reliability of the research were analyzed, it was seen that most methods follows a similar pattern as validity. Here at Table 2, mostly used method is peer debriefing/review (%20), this method is followed by data triangulation which is in the first place in validity (% 14) and thick description (%10), direct citation (%9).

Table 2. Methods to ensure the reliability of the researches

Reliability of research	Frequency	Percentage
Peer debriefing/review	32	20
Triangulation	23	14
Thick description	17	10
Direct citation	15	9
Prolonged engagement	11	7
Interrater reliability	11	7
Detailed report	9	6
Member checking	8	5
Audio/video record	8	5
Sampling strategies	4	3
Pilot scheme	3	2
Descriptive analyzes	3	2
Raw data	3	2
Comfirmation audit	3	2
Deep focused data collection	3	2
Reliability coefficient	3	2
Chain of evidence	2	1
Cohen's kappa coefficient	2	1

When theses on data collection tools' validity and reliability are studied, again same methods are found with the theses that have validity and reliability works of research. In more than half of these researches, peer debriefing/review get the top spot as it is in reliability of the research (%53) is used to ensure the validity of the research tools. This method is followed by pilot scheme (%22). Data triangulation, although used quite a lot in the validity and reliability of the research itself, falls down in terms of its usage on the validity of the tools with a number of only %3 (Table 3).

It is seen that different approaches are employed when the methods to ensure the reliability of tools are examined. On the other side item analysis method, generally used in the validity and reliability studies of the researches that are run through quantitative paradigm, is seen (%38). Strategies such Spearman Brown two half test correlation, KR-20 coefficient, Cronbach alpha coefficient, Pearson correlation coefficient, Spearman correlation coefficient are also seen to be employed (each %1).

Table 3. Methods to ensure the reliability of the research tools

Tools' Reliability	Frequency	Percentage
Item analysis	13	38
Pilot scheme	7	21
Interrater reliability	6	18
Consistence between researchers	4	12
Audio/video record	2	5
Member checking	1	3
Triangulation	1	3

4 Discussion and Results

It is expected that validity and reliability studies are performed meticulously and step by step to test the accuracy of the scientific researches. It must not be ruled out that, the interrogable of the data and calculations collected through scientific research should be minimal. Data and the reason to get them should follow the same pattern. Postgraduate theses are considered to be the first step into the science world and paradigms, experience and wisdom gained through the journey will determine the quality of the works and the path to carve. It shows that there is a problem in regards of the usage of the strategies and methods as required or theses with this much importance not consisting validity and reliability in them. Also, in researches on validity and reliability, peer debriefing/review, data triangulation, and direct citation are one of the mostly used methods and they are not been attentive as one should. It can also be seen that these techniques are used more because they require less effort [6].

The fact that strategies which is used in researches that runs through quantitative paradigm, such as Kaiser-Maier-Olkin, KR-20 are found in the theses, shows that there are problems in regards of differences and awareness. This situation, inherently, affects validity and reliability [11]. In addition to this, strategies such as credibility, transferability, dependability and confirmability [10] and their constant usage in most of the researches, can be shown as a solid and required evidence that qualitative paradigm based researches actually include validity and reliability works. Techniques that are used for validity and reliability are often employed in post graduate theses without their other steps are correctly used.

Consequently, even though, most thesis include validity and reliability studies are performed, it can be forecasted that they are done only for the sake of being done [12 and 13]. This can be seen by their strategy usage, independence of their usage and their predictive analysis. This situation shows that awareness of different analyzing strategies is low and this is caused by assessment differences which is the reason of the differences between qualitative and quantitative paradigms. It is quite striking that validity and reliability which shows if researching techniques (and even researches themselves) are eligible to be used scientifically only remains as words rather than their context and paradigm basis. It is also striking in quarter of the studies they are not even encountered at all.

References

1. Başol, G., Eğitimde Ölçme ve Değerlendirme. Pegem Akademi. Ankara (2015)
2. Baykul, Y., Eğitimde ve psikolojide ölçme klasik test teorisi ve uygulaması. Ankara: ÖSYM Yayınları (2000)
3. Cohen, L., Manion, L. & Morrison, K., Research Methods in Education (5th ed.). London: Routledge/Falmer (2000)
4. Crocker, L. M. & Algina, L., Introduction to classical and modern test theory. New York: Holt, Rinehart and Winston (1986)
5. Erkuş, A., Psikometri üzerine yazılar. Ankara: Türk Psikologlar Derneği (2003)
6. Erkuş, A., Ölçek geliştirme ve uyarlama çalışmalarında karşılaşılan sorunlar, Türk Psikoloji Bülteni, 13 (40), 17-25 (2007)
7. Gelbal, S., Ölçme ve Değerlendirme. Eskişehir: Anadolu Üniversitesi Yayınları (2013)
8. Gronlund, N. E., & Linn, R. L., Measurement and evaluation in teaching (6th ed.). New York: Macmillan (1990)

9. Haladyna, T.M., The Conditions of Assessment of Student Learning in Arizona: In The Conditions of Pre-K–12 Education in Arizona: 2004, ed. A. Molnar. Tempe, Ariz. :Education Policy Studies Laboratory at Arizona State University (2004)
10. Lincoln, Y. S., & Guba, E. G., Naturalistic Inquiry. Beverly Hills, CA: Sage (1985)
11. Karagöz, Y. & Ekici, S. Sosyal Bilimlerde Yapılan Uygulamalı Araştırmalarda Kullanılan İstatistiksel Teknikler Ve Ölçekler. C.Ü. İktisadi ve İdari Bilimler Dergisi, Cilt 5, Sayı 1 (2004)
12. Nunnally, J.C., Psychometric theory, NewYork: McGraw Hill (1978)
13. Nunnally, J. C., & Bernstein, I. H., Psychometric theory (3rd ed.). New York: McGrawHill (1994)
14. Patton, M. Q., Qualitative Evaluation and Research Methods (2nd ed.). Newbury Park, Calif: Sage Publication (1990)
15. Robson, C., Real World Research: A resource for social scientists and practitioner researchers (2nd Ed.). Oxford: Blackwell. (2002)
16. Tan Ş., Öğretimde Ölçme ve Değerlendirme. Pegem Akademi. 8. Baskı. Ankara (2013)
17. Tan,Ş. ve Erdoğan, A., Öğretimi Planlama ve Değerlendirme, Ankara, Pegem A Yayıncılık (2004)
18. Tekin, H., Eğitimde Ölçme ve Değerlendirme, 17. Baskı, Ankara, Yargı Yayınevi (2004)