ORİJİNAL ARAŞTIRMA ORIGINAL RESEARCH

DOI: 10.5336/nurses.2018-61334

The Turkish Adaptation of a Quality Assessment Tool for Quantitative Studies: Validity and Reliability Analyses

Nicel Çalışmalar İçin Kalite Değerlendirme Aracı'nın Türkçe Uyarlaması: Geçerlik ve Güvenirlik Analizleri

Emine ERGİN,[a]
Belgin AKIN[a]

[a]Department of Public Health Nursing, Selçuk University Faculty of Health Sciences, Konya, TÜRKİYE

Received: 04.05.2018
Received in revised form: 07.07.2018
Accepted: 12.07.2018
Available online: 26.11.2018

Correspondence:
Emine ERGİN
Selcuk University
Faculty of Health Sciences,
Department of Public Health Nursing,
Konya,
TURKEY/TÜRKİYE
eminesariselcuk@gmail.com

In study was presented as a verbal statement at the X[th] International Statistics Days, 7-9 October 2016, Giresun, Turkey.

ABSTRACT Objective: One of the main problems for the validity of meta-analytical studies is quality assessment of studies to be included in meta-analysis. This study aimed to examine the quality of quantitative studies and to conduct validity and reliability studies of the Turkish translation of the Quality Assessment Tool for Quantitative Studies. **Material and Methods:** For this tool, language equivalence was examined using translation-back translation method, content validity was evaluated by consulting expert opinion, and reliability was determined depending on inter-rater reliability. The researchers used a content validity index to evaluate the expert opinion and also using Cohen's Kappa. **Results:** The expert evaluation showed a content validity index was 0.99. The opinions of eight experts were evaluated using Kendall W analysis, which revealed that there was no statistical difference (Kendall W=0.13) among their opinions and that their scores were consistent with each other. The present researchers also observed that the Kappa values were between 0.668 and 1 in different studies. **Conclusion:** This study translated the Quality Assessment Tool for Quantitative Studies into Turkish, and determined that it is a reliable tool that can be used to assess the quality of quantitative studies.

Keywords: Quality assessment; quantitative studies; reliability; Turkish version; validity

ÖZET Amaç: Meta analiz çalışmalarının geçerliğindeki temel sorunlardan birisi: Meta analize dahil edilecek çalışmaların kalitesinin değerlendirilmesidir. Bu çalışmada nicel çalışmaların kalitesini değerlendirmek için geliştirilen Nicel Çalışmalar için Kalite Değerlendirme Aracı'nın Türkçe formunun geçerlik ve güvenirlik analizlerinin yapılması amaçlanmıştır. **Gereç ve Yöntemler:** Nicel Çalışmalar için Kalite Değerlendirme Aracı'nın dil eşdeğerliği geri-çeviri yöntemi; kapsam geçerliği uzman görüşüne başvurularak; güvenirliği gözlemciler arası güvenirlik ile incelenmiştir. Uzman görüşlerinin değerlendirilmesi için kapsam geçerlik indeksi (KGİ) kullanılmıştır. Güvenilirlik yönünden gözlemciler arası Kappa analizi ile değerlendirilmiştir. **Bulgular:** Uzman değerlendirmelerine göre KGİ=0,99 bulunmuştur. Sekiz uzmanın görüşleri Kendall W analizi ile de değerlendirilmiş, aralarında istatistiksel olarak farkın olmadığı (Kendall W=0,13) saptanarak, uzman puanlarının uyumlu olduğu görülmüştür. Kappa değerleri farklı çalışma türlerinde 0.668-1 arasında bulunmuştur. **Sonuç:** Türkçe'ye uyarlanan "Nicel Çalışmaların Kalitesini Değerlendirme Aracı"nın nicel çalışmaların kalitesini değerlendirmede kullanılabilecek güvenli bir araç olduğu belirlenmiştir.

Anahtar Kelimeler: Kalite değerlendirme; nicel çalışma; güvenilirlik; Türkçe versiyon; geçerlilik

In recent years, there has been an increase in the number of journals published in a variety of study areas, and this has led to an increase in the number of studies and papers as well. To support this statement, a study stated that between the years 1997 and 2014, the

number of scientific journals published in Turkey increased from 643 to 1679.[1] The authors of that study state that referring to only one study would be insufficient to solve a problem, and recommend synthesizing the results from multiple independent studies on the same subject.[2,3]

As of early twentieth century, researchers began to use modern analytical methods to synthesize the results of empirical studies on the same subject published by different researchers. In time, new methods were developed to produce these syntheses: for instance, systematic review and meta-analysis include the systematic presentation and synthesis of the data provided by any study that they analyze.[4,5] These two methods, which are now accepted as the way to access current literature, are becoming more important and necessary each day. They are important not only for the overall structure of science, but also for the makers and implementers of policy.[3,6] These methods assist the reader to evaluate the inconsistencies in scientific literature and examine the causes of inconsistency. That increases the predictive power of studies, provides cost-effective results, and creates new approaches that can be used in studies.[5,7,8] For this reason, researchers need high-quality studies that produce the high-level evidence needed to judge effective use of time and money.

One of the fundamental problems related to the validity of systematic reviews and meta-analyses is the quality assessment of the studies that should be included.[4] This is of critical importance for researchers, clinicians, and policy-makers.[9] Assessing the quality of the primary studies is essential when conducting systematic reviews and meta-analyses to prevent bias.[10,11] There is no open process providing information about the aspects that add quality to studies, or how the assessment should be made.[12]

The quality assessment of studies is not an easy process in any way. There are different tools to specific to different study designs in the relevant literature to be used to assess the quality of quantitative, qualitative and mixed-design studies when synthesizing studies. While some of these tools, which make a significant contribution to obtaining evidence-based information, are commonly used and suggested, some have been subjected to criticism. Selection bias, performance bias, assignment bias, reporting bias and other bias types affect internal validity. Therefore, Cochrane stated that all methodological quality assessment tools should focus on the risk of bias.[13]

It is important to accurately assess the applicability of tools in systematic reviews and meta-analyses in methodological quality assessment. The literature includes a large number of methodological quality assessment tools such as the tool for randomized controlled studies (Cochrane Collaboration's tool, the Physiotherapy Evidence Database (PEDro) scale, the Modified Jadad Scale, the Delphi List, CASP checklist for RCT ve the NICE methodology checklist for RCT); non-randomized studies The Methodological Index for Non-Randomized Studies (MINORS) and Reisch's tool; analytical studies, especially for cohort and case control studies (The CASP checklist, the SIGN methodology tools, and the Newcastle-Ottawa Scale (NOS).

First, study type should be decided and the most appropriate tool for that study should be selected. In addition, external validity is also an important, but often ignored fact that should be involved in methodological quality assessments carried out to generate evidence.[13-15]

There are specific tools addressed to the assessment of studies with different aspects. Quality assessment of the studies included in the reviews that address quantitative studies with different designs poses a problem. Using different tools for the assessment of primary studies leads to different results.[16,17] There are also some studies in the relevant literature that assess these tools, validity and reliability of most of which are being discussed.[18-20]

There is no consensus on the best method or tool for the assessment of the risk of bias today. Large numbers of tools with different content and features may pose a problem in the quality assessment of the reviews.[21] Some tools specific to study design (e.g., the 5-point Oxford Quality Rating Scale) are considered to be inappropriate for non-pharmacological studies since they are intended for pharmacological studies.[22]

The Quality Assessment Tool for Quantitative Studies (QATQS), in Turkish, Nicel Çalışmalar için Kalite Değerlendirme Aracı (NÇKDA) (Appendix 1), which was created in Canada by the Effective Public Health Project to assess the initiatives addressed to public health as well as the initiatives for health protection and improvement and recommended by the Cochrane Review Group (CRG), has advantages over other tools since it allows for the quality assessment of quantitative studies with different designs. The fact that the QATQS questions the generalizability to the target population can be regarded as a superiority of this tool in terms of partially involving external validity. The QATQS has been indicated to be appropriate for systematic reviews assessing the effectiveness of public health nursing.[18] It has been affirmed that the QATQS study can be used for the assessment of the quality of public health studies focusing on family health, sexual health, prevention of chronic diseases, injuries, and substance use. A study that included 20 randomized controlled studies compared the Cochrane Collaboration Risk of Bias Tool (CCRBT) and the QATQS; it found that there was lower consistency among the observers in the CCRBT than in the QATQS.[19]

In particular, the literature published in Turkish requires assessment tools to evaluate the quality of studies in methodological terms. The present study was examines the validity and reliability of "Turkish Version of the Quality Assessment Tool for Quantitative Studies (TQATQS)". The researchers also aimed to provide a new tool to the relevant literature that can be used to assess the quality of the quantitative studies conducted in Turkey.

MATERIAL AND METHODS

STUDY DESIGN

This is a methodological study conducted between June 2015 and August 2016 with the purpose of translating the QATQS into Turkish, and to assess its validity and reliability.

In this process all the implementations are given with workflow diagram (Figure 1).

DATA COLLECTION TOOL

Quality Assessment Tool for Quantitative Studies (QATQS)

This tool was created in scope of Effective Public Health Practice Project (EPHPP) to be used in studies concerning public health, and it is still being used by some countries (Australian, Kanada). The QATQS, a standardized tool that is used to determine and evaluate the evidence supporting the practice in public health, also includes a comprehensive glossary on the practice and assessment steps.[18] It consists of eight areas: bias of selection, study design, confounder, blinding, data collection method, exclusion and withdrawal from the study integrity of intervention and analysis. Each area, except for integrity of intervention and analysis are scored as 1=Strong, 2=Moderate, and 3=Poor. After each area is scored, the study is given a general score based on the glossary. At this point, having no Poor scores indicated a methodologically strong study, one Poor score indicates a study of moderate reliability study, and two or more Poor scores indicates a methodologically unreliable study. Based on the assessment and scoring, the final decision of each assessor is expressed as 1=Strong, 2= Moderate, and 3= Poor. After scoring, any inconsistencies between the assessors are examined along with the reasons for any inconsistency. There are no scores given for intervention integrity and analysis. These areas act as a guide for assessors when there is hesitation about the quality of the study, and they also contribute to the Discussion section of this study.

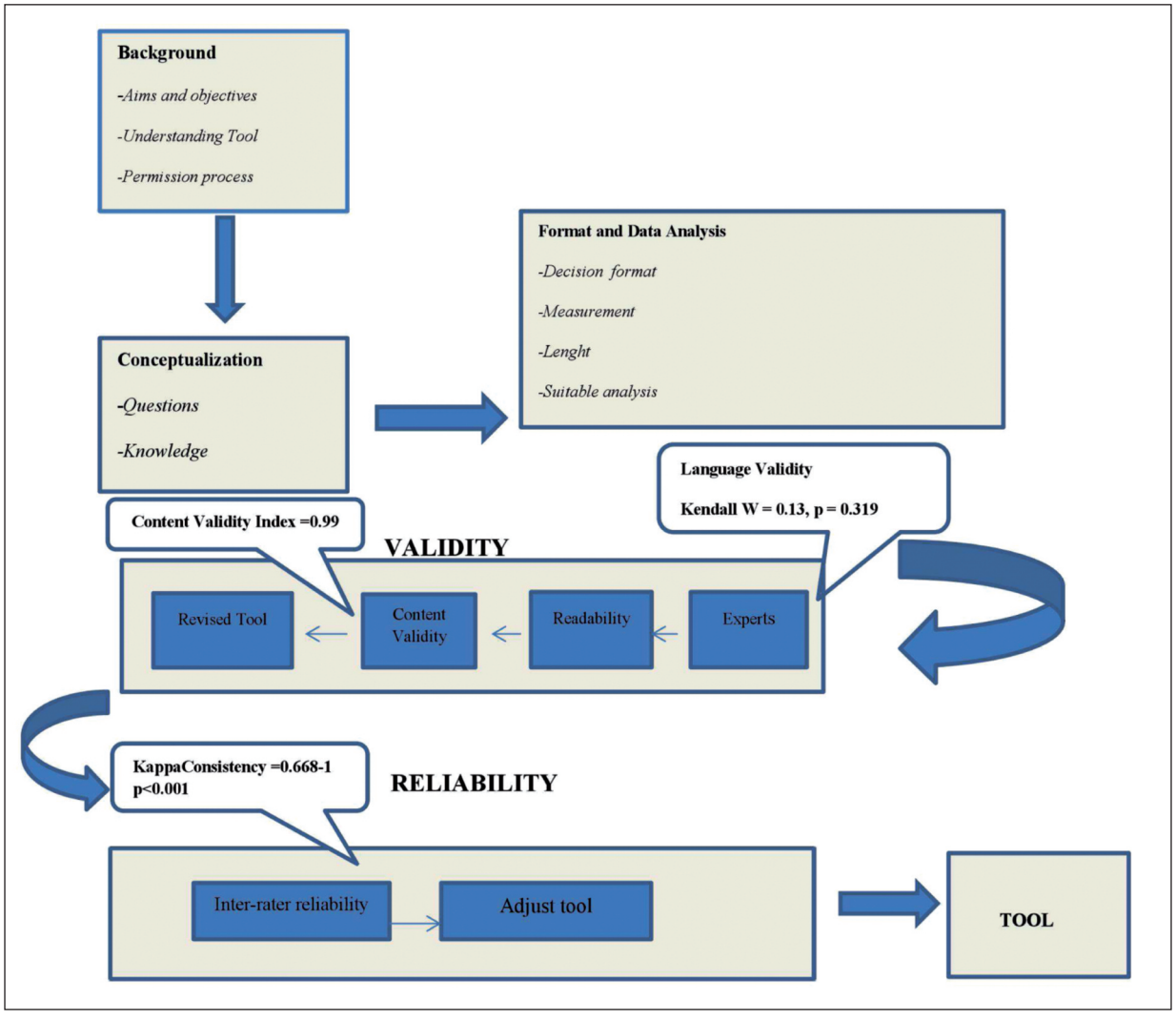

FIGURE 1: Workflow diagram.

The validity and reliability studies of the original scale were conducted by Thomas et al. (2004). The content validity of the tool was evaluated based on the opinions of six experts, and preliminary practice evaluation was done by testing the quality of ten studies together by four experts who specialized in critical assessment and public health. The consistency among interviewers regarding this tool was evaluated in collaboration with two interviewers through a random selection of primary studies and was found to be Kappa 0.74 and Kappa 0.61.[18]

THE VALIDITY AND RELIABIITY OF THE TURKISH VERSION OF THE QUALITY ASSESSMENT TOOL FOR QUANTITATIVE STUDIES (T-QATQS)

Validity

This section will focus on the language and content validity of the T-QATQS regarding the general validity of the tool. The researchers used group translation and translation-back-translation methods to determine the language validity of the study. When the language validity study was completed, the researchers consulted expert opinion regarding content validity using the content validity index

created by Burns and Groves (2009).[23] The researchers also e-mailed the tool and its attachment to eight faculty members in different universities. These faculty members had at least a doctoral degree and had specialized in Statistics [3], Public Health Nursing [4], and Obstetrics Nursing [1], and had experience in research, nursing, public health, scales, systematic assessment, and meta-analysis. Their opinions, and responses were used as a basis for revising the tool.

Reliability

Language and content validity of the study was completed with the practice described previously, and the present researchers evaluated the interrater reliability considering the reliability of the study in general. In this context, five articles with different designs (randomized controlled, controlled clinical trial, cohort, case control, and descriptive-correlational) were selected randomly, sent to two independent researchers who were asked to make another evaluation. Two expert researcher specialists in research, nursing, validity and reliability, public health, and meta-analysis were provided with detailed information about the use of the T-QATQS; they also assessed the studies independently. The reliability of this practice was evaluated using inter-rater Kappa analysis.

Data Analysis

The data collected by this study were analyzed using SPSS 20.00 software in the digital environment. The descriptive data were analyzed using numbers, percentages, means and standard deviation, and the significance level of the study was set at $p<0.05$. Content validity was determined using the Content Validity Index; Kendall analysis was also conducted. Considering the reliability of the study, the researchers evaluated the kappa index in inter-rater reliability (Table 1).

TABLE 1: The aspects evaluated and the statistical methods used in the study.

Evaluated aspect	Statistical methods used
Language and content validity	Content validity index
	Kendall analysis
Inter-rater reliability	Kappa coefficient

Ethical Consideration

The researchers received permission from McMasters University, as well as from the the professors there who were creators of the assessment tool, to translate it into Turkish.

Study Limitations

This study included five studies with five different designs to provide the validity and reliability of the tool, and two experts made their contributions as well. It will strengthen the practice if a larger number of studies are evaluated by more researchers. However, the researchers of this study decided to select one study from each design, considering high workloads and busy schedules of the faculty members in Turkey to make this assessment. This situation is a limitation of this study.

RESULTS

VALIDITY OF THE TURKISH VERSION OF THE QUALITY ASSESSMENT TOOL FOR QUANTITATIVE STUDIES (T-QATQS)

Validity is the degree of an assessment tool to which it is capable of assessing a variable. Although reliability is the first condition for the validity of any study, validity is not always guaranteed by the provision of reliability.[24,25] The studies included in this research were evaluated considering language and content validity of the T- QATQS. The researchers used group translation and the translation-back-translation method to determine the language and content validity of this tool. The tool was translated into Turkish by five experts who were fluent in both Turkish and English (a faculty member nurse, two nurse instructors and two professional translators), and a common text was created based on an evaluation of these translated texts. Afterwards, another expert made the backtranslation of the tool into English, which was the language of the original tool. Two experts who were fluent in both English and Turkish compared the English expressions and the trans-

lated expressions, and evaluated the understandability of these expressions by checking their suitability.*

When the language validity was ensured, the researchers consulted eight faculty members about content validity and suitability for culture. The researchers used the Content Validity Index to evaluate the experts' opinions concerning ways of expression, suitability for the study area, and the content.[23] The researchers also asked the experts to score the items they presented them from 1 to 4 (1- Not suitable, 2- Somewhat suitable (the items need to be made suitable), 3- Fairly suitable (suitable but needs small modifications), and 4- Very suitable). In this assessment, the Content Validity Index is 0.80 when the experts score 80% of the items either 3 or 4.[24,26] Content Validity Index was found 0.99. For language validity scores given by five experts were evaluated by Kendall W analysis, and no statistically significant difference was found among them (Kendall W=0.13, p=0.319), which showed that their scores were consistent with each other.

** During the translation of the tool into Turkish, its glossary was also translated by experts (Appendix 2).*

RELIABILITY OF THE TURKISH VERSION OF THE QUALITY ASSESSMENT TOOL FOR QUANTITATIVE STUDIES (TQATQS)

The researchers examined inter-rater reliability regarding the reliability of this tool. The Kappa values were found between 0.668 and 1 by different types of studies used in the reliability analysis. The weakest reliability (kappa=0.668, p<0.001) was found for the descriptive study, and the strongest reliability (kappa=1, p<0.001) was found for the randomized controlled study. Between the results derived by the two observers, there was a acceptable consistency for descriptive (0.668) and case control studies (0.768), a very good consistency for cohort study (0.928), and a perfect consistency for controlled clinical trial (0.937) and randomized controlled (1) studies (Table 2).

TABLE 2: The consistency between the observers in the studies which administered the quality assessment tool for quantitative studies.

Type of Study	Kappa value	P value
Descriptive	0.668	0.000
Case Control	0.768	0.000
Cohort	0.928	0.000
Controlled Clinical trial	0.937	0.000
Randomized Controlled Trial	1	0.000

DISCUSSION

Because there the number of publications is increasing, it has become more important to go from evidence to suggestion, and make a critical assessment of these evidences. Tools have been developed to evaluate the methodological quality of different types of studies included in reviews according to their features. Some of these tools are recommended for use, whereas the others are unnecessary.[27,28]

The assessment of the methodological quality of any study is very important. There is a range of tools intended to assess methodological quality for different study areas and different designs. However, more than half of these tools lack the characteristics that are needed to make a collective assessment of certain study types.13 Methodological quality usually refers to internal validity, which is open to many types of bias (e.g., selection bias, bias in performance, bias in reporting) during the research procedure.[29,30] For this reason, Cochrane recommends that the tools assessing methodological quality first focus on the risk of bias.[30]

Although the one of the dimensions of the TQATQS also assesses external validity, different tools are needed to assess external validity. External validity means the generalizability of the findings obtained by one or a number of studies. It is expressed as the possibility of obtaining the same results for the studies conducted with a population and at a place and time similar to that of the original study.[15] One tool that assesses external validity

is the External Validity Assessment Tool (EVAT). Better understanding of the external validity of interventions increases the importance of studies and can increase evidence enabling effective interventions to become widespread.[14,15]

The researchers used Content Validity Index to ensure the language and content validity of this tool. This index is suggested to be 0.80 or above to realize content validity for any tool.[24,26] The experts assessment revealed a Content Validity Index of 0.99, and there were no items below the value of 0.80.

The finding of CVI: 0.99 shows that the TQATQS has high content validity. Reliability means the consistency of the questions in a test or questionnaire with each other and how accurately the assessment tool reflects the desired results.[25] The common deficiency of quality determination tools is subjectivity.[13] For this reason, it is necessary that users have research epidemiological knowledge, and have a professional academic attitude. In this case, the best way to avoid bias from evaluators is to have two assessors carry out independent assessments and use cross-checking as well.[31] In the context of the re-liability analyses of the T-QATQS, this study con-sidered the consistency among interviewers, which evaluates the consistency between two or more interviewers regarding the consistency degree of the Kappa coefficient.[32] In Kappa consistency analysis, the Kappa coefficient lies between 0 and 1. Accordingly, the values between 0.93 and 1.00 indicated perfect consistency; 0.81-0.92 indicates very good consistency; 0.61-0.80 indicates good consistency; 0.41-0.60 indicates moderate consistency; 0.21-0.40 indicates a consistency below moderate level; and 0.01-0.20 indicates weak consistency.[33,34] In this study, the consistency among independent observers ranged between 0.66 and 1, which indicates good consistency. Similarly, in the original study by Thomas et al. (2004), the consistency among independent observers ranged from 0.61 to 0.74.[18]

The study type is the primary condition that determines methodological quality. Therefore, the selection of the relevant tool is important. Comprehensive knowledge and lots of practice are the requirements for the accurate evaluation of methodological quality.[13] In this study, the independent assessors rated each study individually as strong, moderate, or weak. The studies that were commonly accepted to have strong quality included the lowest level of bias, and their results were also valid. Of the five types of studies included in this study (randomized controlled trial, controlled clinical trial, cohort, case control, and descriptive-correlational), the randomized controlled trial, which is at the highest step of the evidence pyramid, scored 1 in the consistency among interviewers. The study findings revealed that the lowest consistency was in the descriptive study, and the highest was in the randomized controlled trial.

To make certain biased results invalid and produce a fairer and more accurate assessment of studies, the assessors evaluated these studies from a broader perspective considering their strengths and weaknesses in themselves. Although the present researchers thought that this situation increased subjectivity, which was the limitation of the study, this risk was reduced by the fact that the assessors had similar formal education levels, and their levels of knowledge, background and experience level were close to each other as well. The assessment tool in this study scores the studies considering certain criteria. However, it leaves the final decision to the assessor thanks to the items that are not scored but included in the assessment.

CONCLUSION

This study is concerned with the Turkish adaptation of QATQS, which had been created to assess the methodological quality of quantitative studies, and concluded that it is a valid and reliable tool. This tool was created with the aim of conducting high quality studies, and to contribute to the need for evidence in making decisions about public health practices. It can be applied to any research article with quantitative content. Using a glossary containing detailed explanations of the

items will help derive standardized results from the assessors. The limitations of this study are the inadequate number of experts in this field and that only five studies with different designs were assessed by two experts due to heavy workloads.

Acknowledgements

The researcher has consulted Professor Doctor Emel Ege, Professor Doctor Nihal Esin, Professor Doctor Behice Erci, Professor Doctor İlker Ercan, Associate Professor Doctor Filiz Hisar, and Assistant Professor Doctor Ali Kış in the period of language and content validity and reliability analysis, specially thanks dear statistics expert Assistant Professor Doctor Aydın Karakoca for analyses and thanks them for their valuable time and effort.

Source of Finance

During this study, no financial or spiritual support was received neither from any pharmaceutical company that has a direct connection with the research subject, nor from a company that provides or produces medical instruments and materials which may negatively affect the evaluation process of this study.

Conflict of Interest

No conflicts of interest between the authors and / or family members of the scientific and medical committee members or members of the potential conflicts of interest, counseling, expertise, working conditions, share holding and similar situations in any firm.

Authorship Contributions

Study conception and design: *Belgin Akın, Emine Ergin;* ***Consultancy:*** *Belgin Akın, Emine Ergin;* ***Acquisition of data:*** *Emine Ergin, Belgin Akın;* ***Analysis and interpretation of data:*** *Emine Ergin, Belgin Akın;* ***Literature review:*** *Emine Ergin, Belgin Akın;* ***Writing of article:*** *Emine Ergin, Belgin Akın;* ***Critical review:*** *Belgin Akın;* ***Resources, funding:*** *Emine Ergin, Belgin Akın.*

Appendix 1. Turkish Version of the Quality Assessment Tool for Quantitative Studies (T-QATQS)

QUALITY ASSESSMENT TOOL FOR QUANTITATIVE STUDIES

COMPONENT RATINGS

A) SELECTION BIAS

(Q1) Are the individuals selected to participate in the study likely to be representative of the target population?

1 Very likely
2 Somewhat likely
3 Not likely
4 Can't tell

(Q2) What percentage of selected individuals agreed to participate?

1 80 - 100% agreement
2 60 – 79% agreement
3 less than 60% agreement
4 Not applicable
5 Can't tell

RATE THIS SECTION	STRONG	MODERATE	WEAK
See dictionary	1	2	3

B) STUDY DESIGN

Indicate the study design

1 Randomized controlled trial
2 Controlled clinical trial
3 Cohort analytic (two group pre + post)
4 Case-control
5 Cohort (one group pre + post (before and after))
6 Interrupted time series
7 Other specify _____
8 Can't tell

Was the study described as randomized? If NO, go to Component C.

No Yes

If Yes, was the method of randomization described? (See dictionary)

No Yes

If Yes, was the method appropriate? (See dictionary)

No Yes

RATE THIS SECTION	STRONG	MODERATE	WEAK
See dictionary	1	2	3

C) CONFOUNDERS

(Q1) Were there important differences between groups prior to the intervention?

1 Yes
2 No
3 Can't tell

The following are examples of confounders:

1 Race
2 Sex
3 Marital status/family
4 Age
5 SES (income or class)
6 Education
7 Health status
8 Pre-intervention score on outcome measure

(Q2) If yes, indicate the percentage of relevant confounders that were controlled (either in the design (e.g. stratification, matching) or analysis)?

1 80 – 100% (most)
2 60 – 79% (some)
3 Less than 60% (few or none)
4 Can't Tell

RATE THIS SECTION	STRONG	MODERATE	WEAK
See dictionary	1	2	3

D) BLINDING

(Q1) Was (were) the outcome assessor(s) aware of the intervention or exposure status of participants?

1 Yes
2 No
3 Can't tell

(Q2) Were the study participants aware of the research question?

1 Yes
2 No
3 Can't tell

RATE THIS SECTION	STRONG	MODERATE	WEAK
See dictionary	1	2	3

E) DATA COLLECTION METHODS

(Q1) Were data collection tools shown to be valid?

1 Yes
2 No
3 Can't tell

(Q2) Were data collection tools shown to be reliable?

1 Yes
2 No
3 Can't tell

RATE THIS SECTION	STRONG	MODERATE	WEAK
See dictionary	1	2	3

F) WITHDRAWALS AND DROP-OUTS

(Q1) Were withdrawals and drop-outs reported in terms of numbers and/or reasons per group?

1 Yes
2 No
3 Can't tell
4 Not Applicable (i.e. one time surveys or interviews)

(Q2) Indicate the percentage of participants completing the study. (If the percentage differs by groups, record the lowest).

1 80 -100%
2 60 - 79%
3 less than 60%
4 Can't tell
5 Not Applicable (i.e. Retrospective case-control)

RATE THIS SECTION	STRONG	MODERATE	WEAK	
See dictionary	1	2	3	Not Applicable

G) INTERVENTION INTEGRITY

(Q1) What percentage of participants received the allocated intervention or exposure of interest?

1 80 -100%
2 60 - 79%
3 less than 60%
4 Can't tell

(Q2) Was the consistency of the intervention measured?

1 Yes
2 No
3 Can't tell

(Q3) Is it likely that subjects received an unintended intervention (contamination or co-intervention) that may influence the results?

4 Yes
5 No
6 Can't tell

H) ANALYSES

(Q1) Indicate the unit of allocation (circle one)

community organization/institution practice/office individual

(Q2) Indicate the unit of analysis (circle one)

community organization/institution practice/office individual

(Q3) Are the statistical methods appropriate for the study design?

1 Yes
2 No
3 Can't tell

(Q4) Is the analysis performed by intervention allocation status (i.e. intention to treat) rather than the actual intervention received?

1 Yes
2 No
3 Can't tell

GLOBAL RATING

COMPONENT RATINGS

Please transcribe the information from the gray boxes on pages 1-4 onto this page. See dictionary on how to rate this section.

A	**SELECTION BIAS**	**STRONG**	**MODERATE**	**WEAK**	
		1	2	3	
B	**STUDY DESIGN**	**STRONG**	**MODERATE**	**WEAK**	
		1	2	3	
C	**CONFOUNDERS**	**STRONG**	**MODERATE**	**WEAK**	
		1	2	3	
D	**BLINDING**	**STRONG**	**MODERATE**	**WEAK**	
		1	2	3	
E	**DATA COLLECTION METHOD**	**STRONG**	**MODERATE**	**WEAK**	
		1	2	3	
F	**WITHDRAWALS AND DROPOUTS**	**STRONG**	**MODERATE**	**WEAK**	
		1	2	3	Not Applicable

GLOBAL RATING FOR THIS PAPER (circle one):

1	STRONG	(no WEAK ratings)
2	MODERATE	(one WEAK rating)
3	WEAK	(two or more WEAK ratings)

With both reviewers discussing the ratings:

Is there a discrepancy between the two reviewers with respect to the component (A-F) ratings?

No Yes

If yes, indicate the reason for the discrepancy

1	Oversight
2	Differences in interpretation of criteria
3	Differences in interpretation of study

Final decision of both reviewers (circle one):

1	**STRONG**
2	**MODERATE**
3	**WEAK**

Appendix 2. Glossary of Turkish Version of the Quality Assessment Tool for Quantitative Studies (T-QATOS)

Quality Assessment Tool for Quantitative Studies Dictionary

The purpose of this dictionary is to describe items in the tool thereby assisting raters to score study quality. Due to under-reporting or lack of clarity in the primary study, raters will need to make judgements about the extent that bias may be present. When making judgements about each component, raters should form their opinion based upon information contained in the study rather than making inferences about what the authors intended.

A) SELECTION BIAS

(Q1) Participants are more likely to be representative of the target population if they are randomly selected from a comprehensive list of individuals in the target population (score very likely). They may not be representative if they are referred from a source (e.g. clinic) in a systematic manner (score somewhat likely) or self-referred (score not likely).

(Q2) Refers to the % of subjects in the control and intervention groups that agreed to participate in the study before they were assigned to intervention or control groups.

B) STUDY DESIGN

In this section, raters assess the likelihood of bias due to the allocation process in an experimental study. For observational studies, raters assess the extent that assessments of exposure and outcome are likely to be independent. Generally, the type of design is a good indicator of the extent of bias. In stronger designs, an equivalent control group is present and the allocation process is such that the investigators are unable to predict the sequence.

Randomized Controlled Trial (RCT)

An experimental design where investigators randomly allocate eligible people to an intervention or control group. A rater should describe a study as an RCT if the randomization sequence allows each study participant to have the same chance of receiving each intervention and the investigators could not predict which intervention was next. If the investigators do not describe the allocation process and only use the words 'random' or 'randomly', the study is described as a controlled clinical trial.

See below for more details.

Was the study described as randomized?

Score YES, if the authors used words such as random allocation, randomly assigned, and random assignment.

Score NO, if no mention of randomization is made.

Was the method of randomization described?

Score YES, if the authors describe any method used to generate a random allocation sequence.

Score NO, if the authors do not describe the allocation method or describe methods of allocation such as alternation, case record numbers, dates of birth, day of the week, and any allocation procedure that is entirely transparent before assignment, such as an open list of random numbers of assignments.
If NO is scored, then the study is a controlled clinical trial.

Was the method appropriate?

Score YES, if the randomization sequence allowed each study participant to have the same chance of receiving each intervention and the investigators could not predict which intervention was next. Examples of appropriate approaches include assignment of subjects by a central office unaware of subject characteristics, or sequentially numbered, sealed, opaque envelopes.

Score NO, if the randomization sequence is open to the individuals responsible for recruiting and allocating participants or providing the intervention, since those individuals can influence the allocation process, either knowingly or unknowingly.

If NO is scored, then the study is a controlled clinical trial.

Controlled Clinical Trial (CCT)
An experimental study design where the method of allocating study subjects to intervention or control groups is open to individuals responsible for recruiting subjects or providing the intervention. The method of allocation is transparent before assignment, e.g. an open list of random numbers or allocation by date of birth, etc.

Cohort analytic (two group pre and post)
An observational study design where groups are assembled according to whether or not exposure to the intervention has occurred. Exposure to the intervention is not under the control of the investigators. Study groups might be non-equivalent or not comparable on some feature that affects outcome.

Case control study
A retrospective study design where the investigators gather 'cases' of people who already have the outcome of interest and 'controls' who do not. Both groups are then questioned or their records examined about whether they received the intervention exposure of interest.

Cohort (one group pre + post (before and after)
The same group is pretested, given an intervention, and tested immediately after the intervention. The intervention group, by means of the pretest, act as their own control group.

Interrupted time series
A time series consists of multiple observations over time. Observations can be on the same units (e.g. individuals over time) or on different but similar units (e.g. student achievement scores for particular grade and school). Interrupted time series analysis requires knowing the specific point in the series when an intervention occurred.

C) CONFOUNDERS

By definition, a confounder is a variable that is associated with the intervention or exposure and causally related to the outcome of interest. Even in a robust study design, groups may not be balanced with respect to important variables prior to the intervention. The authors should indicate if confounders were controlled in the design (by stratification or matching) or in the analysis. If the allocation to intervention and control groups is randomized, the authors must report that the groups were balanced at baseline with respect to confounders (either in the text or a table).

D) BLINDING

(Q1) Assessors should be described as blinded to which participants were in the control and intervention groups. The purpose of blinding the outcome assessors (who might also be the care providers) is to protect against detection bias.

(Q2) Study participants should not be aware of (i.e. blinded to) the research question. The purpose of blinding the participants is to protect against reporting bias.

E) DATA COLLECTION METHODS

Tools for primary outcome measures must be described as reliable and valid. If 'face' validity or 'content' validity has been demonstrated, this is acceptable. Some sources from which data may be collected are described below:

Self reported data includes data that is collected from participants in the study (e.g. completing a questionnaire, survey, answering questions during an interview, etc.).

Assessment/Screening includes objective data that is retrieved by the researchers. (e.g. observations by investigators).

Medical Records/Vital Statistics refers to the types of formal records used for the extraction of the data.

Reliability and validity can be reported in the study or in a separate study. For example, some standard assessment tools have known reliability and validity.

F) WITHDRAWALS AND DROP-OUTS

Score **YES** if the authors describe BOTH the numbers and reasons for withdrawals and drop-outs.

Score **NO** if either the numbers or reasons for withdrawals and drop-outs are not reported.

The percentage of participants completing the study refers to the % of subjects remaining in the study at the final data collection period in all groups (i.e. control and intervention groups).

G) INTERVENTION INTEGRITY

The number of participants receiving the intended intervention should be noted (consider both frequency and intensity). For example, the authors may have reported that at least 80 percent of the participants received the complete intervention. The authors should describe a method of measuring if the intervention was provided to all participants the same way. As well, the authors should indicate if subjects received an unintended intervention that may have influenced the outcomes. For example, co-intervention occurs when the study group receives an additional intervention (other than that intended). In this case, it is possible that the effect of the intervention may be over-estimated. Contamination refers to situations where the control group accidentally receives the study intervention. This could result in an under-estimation of the impact of the intervention.

H) ANALYSIS APPROPRIATE TO QUESTION

Was the quantitative analysis appropriate to the research question being asked?

An intention-to-treat analysis is one in which all the participants in a trial are analyzed according to the intervention to which they were allocated, whether they received it or not. Intention-to-treat analyses are favoured in assessments of effectiveness as they mirror the noncompliance and treatment changes that are likely to occur when the intervention is used in practice, and because of the risk of attrition bias when participants are excluded from the analysis.

Component Ratings of Study:

For each of the six components A – F, use the following descriptions as a roadmap.

A) SELECTION BIAS

Strong: The selected individuals are very likely to be representative of the target population (Q1 is 1) **and** there is greater than 80% participation (Q2 is 1).

Moderate: The selected individuals are at least somewhat likely to be representative of the target population (Q1 is 1 or 2); **and** there is 60 - 79% participation (Q2 is 2). 'Moderate' may also be assigned if Q1 is 1 or 2 and Q2 is 5 (can't tell).

Weak: The selected individuals are not likely to be representative of the target population (Q1 is 3); **or** there is less than 60% participation (Q2 is 3) **or** selection is not described (Q1 is 4); and the level of participation is not described (Q2 is 5).

B) DESIGN

Strong: will be assigned to those articles that described RCTs and CCTs.

Moderate: will be assigned to those that described a cohort analytic study, a case control study, a cohort design, or an interrupted time series.

Weak: will be assigned to those that used any other method or did not state the method used.

C) CONFOUNDERS

Strong: will be assigned to those articles that controlled for at least 80% of relevant confounders (Q1 is 2); **or** (Q2 is 1).

Moderate: will be given to those studies that controlled for 60 – 79% of relevant confounders (Q1 is 1) **and** (Q2 is 2).

Weak: will be assigned when less than 60% of relevant confounders were controlled (Q1 is 1) **and** (Q2 is 3) **or** control of confounders was not described (Q1 is 3) **and** (Q2 is 4).

D) BLINDING

Strong: The outcome assessor is not aware of the intervention status of participants (Q1 is 2); **and t**he study participants are not aware of the research question (Q2 is 2).

Moderate: The outcome assessor is not aware of the intervention status of participants (Q1 is 2); **or** the study participants are not aware of the research question (Q2 is 2); **or b**linding is not described (Q1 is 3 and Q2 is 3).

Weak: The outcome assessor is aware of the intervention status of participants (Q1 is 1); **and** the study participants are aware of the research question (Q2 is 1).

E) DATA COLLECTION METHODS

Strong: The data collection tools have been shown to be valid (Q1 is 1); **and** the data collection tools have been shown to be reliable (Q2 is 1).

Moderate: The data collection tools have been shown to be valid (Q1 is 1); **and** the data collection tools have not been shown to be reliable (Q2 is 2) **or** reliability is not described (Q2 is 3).

Weak: The data collection tools have not been shown to be valid (Q1 is 2) **or** both reliability and validity are not described (Q1 is 3 and Q2 is 3).

F) WITHDRAWALS AND DROP-OUTS - a rating of:

Strong: will be assigned when the follow-up rate is 80% or greater (Q2 is 1).

Moderate: will be assigned when the follow-up rate is 60 – 79% (Q2 is 2) **OR** Q2 is 5 (N/A).

Weak: will be assigned when a follow-up rate is less than 60% (Q2 is 3) or if the withdrawals and drop-outs were not described (Q2 is 4).

REFERENCES

1. Kozak N. [Health sciences journals]. Türkiye Akademik Dergiler Rehberi. 1. Baskı. Ankara: Detay Yayıncılık; 2014. p.2-10.
2. Borenstein M, Hedges LV, Higgins J, Rothstein H. How a meta analysis work? Introduction to Meta-Analysis. 1st ed. Hoboken, NJ, ABD: John Wiley & Sons Inc; 2009. p.1-7.
3. Üstün U, Eryılmaz A. [A research methodology to conduct effective research syntheses: meta-analysis]. Education and Science 2014;39(174): 1-32.
4. Stangl DK, Berry DA. Meta-analysis: past and present challenges. Meta-Analysis in Medicine and Health Policy. 1st ed. New York, NY: Marcel Dekker; 2000. p.1-22.
5. Haidich AB. Meta-analysis in medical research. Hippokratia 2010;14(Suppl 1):29-37.
6. Chan ME, Arvey RD. Meta-analysis and the development of knowledge. Perspect Psychol Sci 2012;7(1):79-92.
7. Naylor CD. Meta-analysis and the meta-epidemiology of clinical research. BMJ 1997;315 (7109):617-9.
8. Deliktaş A, Kabukcuoğlu K, Kış A. [Meta-analysis application process in nurs-ing: a guide intended for methodology]. Journal of Human Sciences 2016;13(1):1906-25.
9. Baker R, Jackson D. A new approach to outliers in meta-analysis. Health Care Manag Sci 2008;11(2):121-31.
10. Sanderson S, Tatt ID, Higgins JP. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and anno-tated bibliography. Int J Epidemiol 2007;36(3) 666-76.
11. Mhaskar R, Djulbegovic B, Magazin A, Soares HP, Kumar A. Published meth-odological quality of randomized controlled trials does not reflect the actual quality assessed in protocols. J Clin Epidemiol 2012;65(6):602-9.
12. Cooper H, Hedges LV, Valentine JC. Coding the literature. The Handbook of Research Synthesis and Meta-Analysis. 2nd ed. NewYork, USA: Russell Sage Foundation; 2009. p.129-35.
13. Zeng X, Zhang Y, Kwong JS, Zhang C, Li S, Sun F, et al. The methodological quality assessment tools for preclinical and clinical studies, systematic review and meta-analysis, and clinical practice guideline: a systematic review. J Evid Based Med 2015;8(1):2-10.
14. Khorsan R, Crawford C. How to assess the external validity and model validity of therapeutic trials: a conceptual approach to systematic review methodology. Evid Based Complement Alternat Med 2014;2014:694804.
15. Klesges LM, Dzewaltowski DA, Glasgow RE. Review of external validity re-porting in childhood obesity prevention research. Am J Prev Med 2008;34(3):216-23.
16. Herbison P, Hay-Smith J, Gillespie WJ. Adjustment of meta-analyses on the basis of quality scores should be abandoned. J Clin Epidemiol 2006;59(12):1249-56.
17. Jüni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clin-ical trials for meta-analysis. JAMA 1999;282(11):1054-60.
18. Thomas BH, Ciliska D, Dobbins M, Micucci S. A process for systematically re-viewing the literature: providing the research evidence for public health nursing interventions. Worldviews Evid Based Nurs 2004;1(3):176-84.
19. Armijo-Olivo S, Stiles CR, Hagen NA, Biondo PD, Cummings GC. Assessment of study quality for systematic reviews: a comparison of the cochrane collaboration risk of bias tool and the effective public health practice project quality assessment tool: methodological research. J Eval Clin Pract 2012;18(1):12-8.
20. Östlund U, Kidd L, Wengström Y, Rowa-Dewar N. Combining qualitative and quantitative research within mixed method research designs: a methodological review. Int J Nurs Stud 2011;48(3):369-83.
21. Viswanathan M, Ansari MT, Berkman ND, Chang S, Hartling L, McPheeters LM, et al. Assessing the risk of bias of individual studies in systematic reviews of health care interventions. AHRQ Methods for Effective Health Care. March 2012. AHRQ Publication No. 12-EHC047-EF.
22. Uman LS. Systematic reviews and meta-analyses. J Can Acad Child Adolesc Psychiatry 2011;20(1):57-9.
23. Burn N, Grove SK. The research process. The Practice of Nursing Research: Appraisal, Synthesis and Generation of Evidence. 8th ed. Maryland Heights, Missouri: Saunders Elsevier 2009. p.400-20.
24. Gözüm S, Aksayan S. [Guide II: the psychometric properties and cultural com-parison for intercultural scale adaptation]. Nursing Research and Development Magazine 2003;5(1): 3-14.
25. Kalaycı Ş. [Reliability analysis]. SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri. 5. Baskı. Ankara: Asil Yayınları; 2010. p.404-6.
26. Burns N, Grove S. Evolution of evidence-based practice in nursing. The Practice of Nursing Research: Appraisal, Synthesis, and Generation of Evidence. 6th ed. St. Louis: W.B. Saunders Company; 2009. p.15-20.
27. Guyatt GH, Oxman AD, Kunz R, Falck-Ytter Y, Vist GE, Liberati A, et al. Go-ing from evidence to recommendations. BMJ 2008;336 (7652):1049-51.
28. Andrews J, Guyatt G, Oxman AD, Alderson P, Dahm P, Falck-Ytter Y, et al. GRADE guidelines. 14. going from evidence to recommendations: the signifi-cance and presentation of recommendations. J Clin Epidemiol 2013;66 (7):719-25.
29. Campbell DT. Factors relevant to the validity of experiments in social settings. Psychol Bull 1957;54(4):297-312.
30. Higgins J, Altman DG. Chapter 8: Assessing risk of bias in included studies. In: Higgins J, Green S, eds. Cochrane Handbook for Systematic Reviews of Inter-ventions Version 5.0. The Cochrane Collaboration; 2011.
31. Gold C, Erkkilä J, Crawford MJ. Shifting effects in randomised controlled trials of complex interventions: a new kind of performance bias? Acta Psychiatr Scand 2012;126(5):307-14.
32. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statis-tic. Fam Med 2005;37(5):360-3.
33. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3): 276-82.
34. Hallgren KA. Computing inter-rater reliability for observational data: an over-view and tutorial. Tutor Quant Methods Psychol 2012;8(1): 23-34.