Spor Bilimleri Araştırmaları Dergisi
Journal of Sport Sciences Research
Vol: 8, Issue: 3, October 2023
E-ISSN: 2548-0723
URL: http://www.dergipark.org.tr/jssr

Evaluation of Research Quality in Systematic Reviews Specific to Exercise and Sport Sciences: Adapting TESTEX Criteria to Turkish

Hüseyin Şahin UYSAL[1*], Ahmet Yavuz KARAFİL[1], Oğuzhan DALKIRAN[1], Seda Nur UYSAL[2], Sezgin KORKMAZ[1]

[1] Burdur Mehmet Akif Ersoy University, Faculty of Sports Sciences, Burdur.
[2] Burdur Mehmet Akif Ersoy University, Institute of Educational Sciences, Burdur.

Research Article
DOI: 10.25307/jssr.1301934
Received: 24.05.2023 Accepted: 17.09.2023 Online Published: 31.10.2023

Abstract

This study aims to adapt the TESTEX scale, which is preferred by researchers who make systematic reviews in the field of exercise and sports, into Turkish in a valid and reliable way. The study was carried out according to the observational research design, which is one of the quantitative research methods, and the study group consisted of five independent language experts and five researchers. The following procedures were carried out for the adaptation of the TESTEX scale to Turkish: (i) Consistency among independent language experts, (ii) Content validity of the TESTEX scale, (iii) Language validity of the TESTEX scale, (iv) Inter-rater reliability, (v) Inter-rater consistency. For statistical analysis of these procedures, Kendall-W agreement coefficient, content validity ratio, content validity index (CVI), Pearson correlation coefficient (r), Fisher'z effect size index (z), intra-class correlation coefficient (ICC), and Bland Altman analyses were carried out. Study results showed that the TESTEX scale was adapted to Turkish as valid and reliable. In conclusion, the Turkish version of the TESTEX scale may be preferred to evaluate the methodological quality of research articles in systematic reviews. In addition, researchers can benefit from the quality criteria in the scale to improve the methodological quality of their studies.
Keywords: Systematic review, Meta-analysis, Sport science, Exercise and sport, Training

Egzersiz ve Spor Bilimlerine Özgü Sistematik Derlemelerde Araştırma Kalitesinin Değerlendirilmesi: TESTEX Kriterlerinin Türkçeye Uyarlanması

Öz

Bu çalışma, egzersiz ve spor alanında sistematik derleme yapan araştırmacıların tercih ettiği TESTEX ölçeğini Türkçe' ye geçerli ve güvenilir olarak uyarlamayı amaçlamaktadır. Çalışma, nicel araştırma yöntemlerinden gözlemsel araştırma desenine göre gerçekleştirilmiş olup, çalışma grubunu beş bağımsız dil profesyoneli ve beş araştırmacı oluşturmuştur. TESTEX ölçeğinin Türkçe' ye uyarlanması için takip eden işlemler gerçekleştirilmiştir; (i) Bağımsız dil profesyonelleri arasındaki uyum, (ii) TESTEX ölçeğinin kapsam geçerliği, (iii) TESTEX ölçeğinin dil geçerliği, (iv) Değerlendiriciler arası güvenirlik, (v) Değerlendiriciler arası uyum. Bu işlemlerin istatistiksel analizi için Kendall-W uyuşum katsayısı, kapsam geçerlik oranı, kapsam geçerlik indeksi (CVI), pearson korelasyon katsayısı (r), Fisher'z etki büyüklüğü indeksi (z), sınıf-içi korelasyon katsayısı (ICC) ve Bland Altman analizleri gerçekleştirilmiştir. Çalışma sonuçları, TESTEX ölçeğinin Türkçe'ye geçerli ve güvenilir olarak uyarlandığını göstermiştir. Sonuç olarak, sistematik derlemelerde araştırma makalelerinin metodolojik kalitesini değerlendirmek için TESTEX ölçeğinin Türkçe formu tercih edilebilir. Ayrıca araştırmacılar, çalışmalarının metodolojik kalitesini artırmak için ölçekte yer alan kalite kriterlerinden faydalanabilir.
Anahtar kelimeler: Sistematik derleme, Meta analiz, Spor bilimi, Egzersiz ve spor, Antrenman

[*] **Corresponding Author:** Research Assistant Hüseyin Şahin UYSAL, **E-mail:** hsuysal@mehmetakif.edu.tr

INTRODUCTION

Exercise and sports science is a field of study that explores the connection between individuals and physical activities, considering physical, social, and cognitive aspects. Researchers have interpreted this field of science in various ways (Bishop, 2008; Haff et al., 2010; Hoffman, 2002). Hoffman (2002) argues that sports science draws on multiple disciplines, including physiology, biochemistry, biomechanics, nutrition, and endocrinology, to improve athletic performance. In another definition, sports science refers to the scientific principles that guide sports practices in order to increase athlete performance (Bishop, 2008). On the other hand, while it is stated that there are slight differences between exercise science and sports science, sports science is reported as a specialized component of exercise science (Haff et al., 2010). Since exercise and sports science are multidisciplinary, the experiences obtained from various disciplines constitute the research scope of exercise and sports science.

With the advancement of technology, knowledge in exercise and sports science has increased through social media and the internet. However, researchers have stated that these developments could have negative aspects, leading to biased or false knowledge on social media and the internet (Meel and Vishwakarma, 2020). While this situation limits researchers who aim to achieve valid and reliable knowledge, it highlights the significance of the scientific evidence pyramid. The pyramid of scientific evidence gives clues to researchers about the certainty of the evidence of a research method. Researchers stated that case reports and expert opinions have the lowest reliability and validity in scientific knowledge. They also claimed that the most reliable and valid scientific knowledge is found in systematic review and meta-analysis studies (Murad et al., 2016). Exercise and sports-specific systematic review studies are increasing rapidly, and researchers have conducted these studies to produce highly reliable and valid scientific knowledge (Juhl and Lund, 2018).

~~A~~ systematic review is a research methodology that involves the application of protocols to identify, select, and critically evaluate relevant research articles to address a specific research problem. Therefore, the quality of research articles can significantly affect the quality and validity of a systematic review (Acosta et al., 2020). Researchers have developed various methodological quality assessment tools for different scientific fields to increase the reliability and validity of systematic reviews (Maher et al., 2003; O'Brien et al., 2014). Methodological quality assessment tools are used to assess the reporting and overall quality of research. While the Physiotherapy Evidence Database (PEDro) scale is widely used to evaluate the methodological quality of the research articles in systematic reviews in the field of exercise and sports science, the researchers have realized that the quality criteria in the PEDro scale may not be exactly unique aspects of this field due to certain limitations (Smart et al., 2015). Therefore, researchers developed the Tool for the Assessment of Study Quality and Reporting in Exercise (TESTEX) scale to evaluate the methodological quality of research articles that is specific to the field of exercise and sports science (Smart et al., 2015).

The TESTEX scale consists of 12 quality criteria to evaluate the methodological quality of research articles and an evaluation is evaluated over a total of 15 points (Smart et al., 2015). In addition, the TESTEX scale includes 5 points for study quality and 10 points for reporting

quality (Smart et al., 2015). This scale includes various criteria on sample selection, intervention process, and training protocols, and researchers stated that the TESTEX scale has a very high level of reliability for assessing methodological quality (ICC = 0.96, 95% confidence interval [CI] = 0.89–0.98, $p < 0.001$) (Smart et al., 2015). While a reference score for the methodological quality level was not reported in the current study (Smart et al., 2015), the following reference scores were used to express the quality level in another study; < 4 points "poor quality," 4-7 points "modarate quality," 8-11 "good quality" and > 11 points "excellent quality" (Davies et al., 2021).

Checklists are not generally preferred to increase reporting quality in exercise and sports sciences studies (Abt et al., 2022). Therefore, there is a particular need to evaluate the methodological quality of research in systematic reviews (Abt et al., 2022; Bishop, 2008; Rico-González et al., 2021). Although many systematic reviews and meta-analysis studies have been carried out in the field of exercise and sports sciences in Turkey, it can be claimed that the methodological quality of studies is ignored.

This study aims to adapt the TESTEX scale into Turkish, which is used in exercise and sport-specific systematic reviews. This study is significant for researchers who make systematic reviews in the Turkish language to use methodological quality assessment tools. Moreover, this study may increase knowledge and interest in systematic review studies. To our knowledge, the TESTEX scale will be adapted to a language other than English for the first time. As a result of the analysis, we assumed that the TESTEX scale would be adapted to Turkish in a reliable and valid way.

METHODS

Study Design

In this study, the observational research design, one of the quantitative research methods, was preferred to adapt the TESTEX scale to Turkish (Büyüköztürk et al., 2022). The observational research design is used to determine the characteristics of a variable (Büyüköztürk et al., 2022). This study aimed to determine the validity and reliability of the TESTEX scale between its original form and its Turkish version. While the study was designed according to previous study protocols (Hür et al., 2022; Tekindal et al., 2021), the study protocol was pre-registered on the Open Science Framework (OSF) (DOI: https://doi.org/10.17605/OSF.IO/73F86). All documents used during the study are provided open-access via OSF (https://osf.io/e6kuh/, accessed date: 3.04.2023).

Data Collection Tool

In this study, the TESTEX scale was used as a data collection tool, and the scale was formed from a total of 12 quality criteria (Smart et al., 2015). While the scale includes two separate sections, namely the quality of study and the quality of reporting, nine items are evaluated over "1 point", two items "2 points," and one item "3 points" (Smart et al., 2015). In this study, the TESTEX scale was used to evaluate the methodological quality of individual studies. The researchers assessed the studies obtained through the Web of Science (WOS) database.

Ethical Approval

Ethics committee approval of the study was obtained by the Ethics Committee of Burdur Mehmet Akif Ersoy University Non-invasive at the meeting held on 01.03.2023 with meeting number 2023/23 with the decision number GO 2023/122.

Research Procedure

The data collection process of this study started with the permission of the researchers who developed the TESTEX scale (OSF). The quality criteria in the TESTEX scale were independently translated into Turkish by three researchers with high-level English reading skills. The translated TESTEX scale was checked by language experts who live abroad, use English as a daily conversational language, or prove that they had a good command of the Engilish language. After the feedback was received, the Turkish form of the scale was revised by three researchers. After the revision, the Turkish form of the TESTEX scale was translated into English by another language experts who is fluent in both languages, and the Turkish translation of the scale was evaluated by five independent language experts.

In the second stage of the data collection process, the methodological quality of studies was evaluated by two independent researchers with the TESTEX scale adapted to Turkish. It was decided to evaluate 20 studies with the TESTEX scale (Tekindal et al., 2021), and the studies were determined through the WOS database. The search was carried out with the following search criteria according to the expertise of the researchers; (i) containing at least one of the keywords plyometric training, complex training, contrast training, and HIIT training in the title of the article, (ii) being scanned in the category of sports science, (iii) written in the English language, (iv) double-blind peer-reviewed research articles. As a result of the search, 329 studies were found according to the specified criteria and exported in Excel format. The Research Randomizer software tool was used to avoid selection bias and ensure randomization throughout the screening (Urbaniak G. C. and Plous S. Research Randomizer Version 4.0; https://www.randomizer.org/, accessed date: 31.03.2023). Studies identified with WOS are included to be evaluated with the forms of the TESTEX scale in this study. The exported Excel file and detailed information on the randomization processes are provided access via OSF **(**https://osf.io/e6kuh/**,** accessed date: 3.04.2023).

The 20 studies included in the study were evaluated by two independent researchers with the Turkish and English forms of the TESTEX scale, and the evaluation results were used to analyse the study. Detailed information about the flow chart of the study is given in Figure 1.

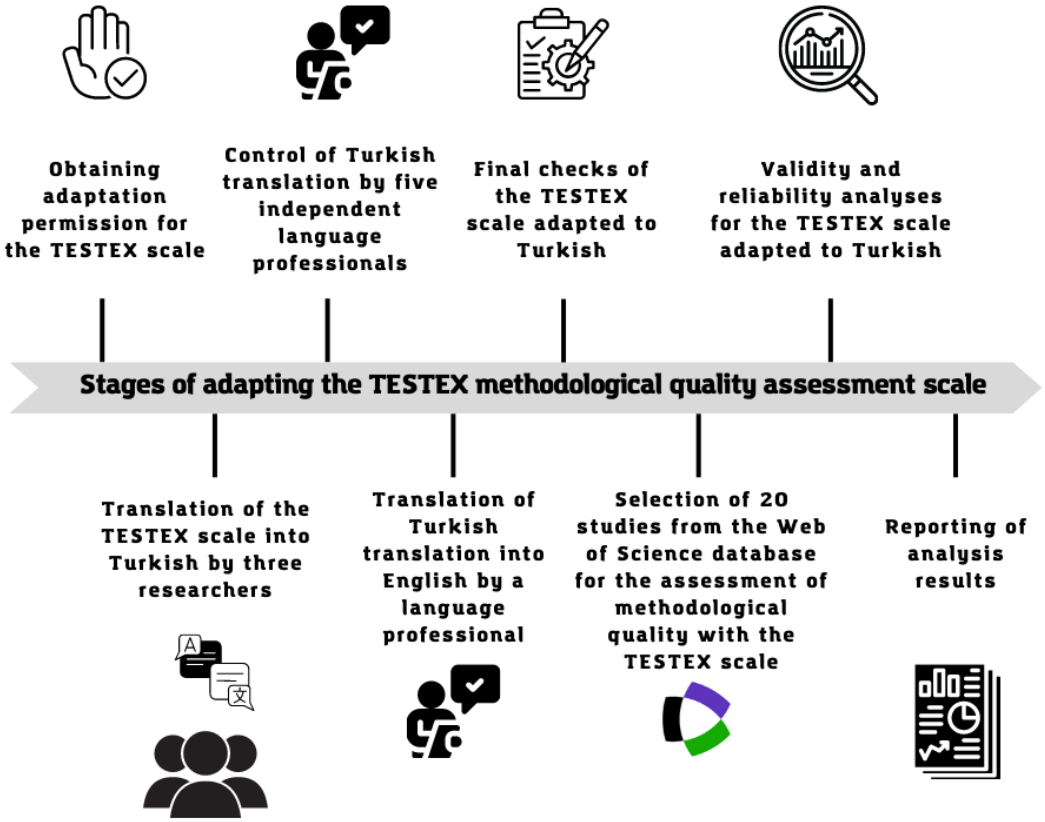

Figure 1. The stages of adapting the TESTEX methodological quality assessment scale into the Turkish

Statistical Analysis

The normality distributions of the analysed data were checked with skewness and kurtosis values, and the analysis results were reported with descriptive statistics (arithmetic mean ± standard deviation) or 95% confidence intervals.

The content validity ratio (CVR) and content validity index (CVI) were used to evaluate the content validity of the translated TESTEX scale (Yeşilyurt and Çapraz, 2018). The CVR and CVI values of the quality criteria were calculated according to the following formulas (Yeşilyurt and Çapraz, 2018);

$$CVR = \frac{Number\ of\ experts\ who\ said\ "suitable"\ to\ the\ translation\ of\ quality\ criteria}{Total\ number\ of\ expert\ \ /\ 2} - 1$$

$$CVI = \frac{\sum CVR\ sum\ of\ quality\ criteria}{Total\ number\ of\ quality\ criteria}$$

Language experts scored each quality criterion according to the following references, and CVR - CVI values were calculated; appropriate (3 points), suited but minor corrections required (2 points), the translation should be removed and rewritten (1 point) (Yeşilyurt and Çapraz, 2018). The obtained CVI value was compared with the content validity criterion reported by Lawshe (Lawshe, 1975). Quality criteria with a CVI value equal to or higher than 0.99 were assumed

to have content validity (Lawshe, 1975). On the other hand, the consistency between the scores of the language experts was analysed with the Kendall-W agreement coefficient.

Language validity was checked according to the scores given by two researchers who evaluated the TESTEX scale in Turkish and English, and validity was assessed with Pearson correlation coefficient analysis. The effect size of the validity was calculated by converting the obtained correlation coefficient to the Fisher z value. The Fisher z effect size was interpreted according to the following reference value: trivial (<0.10), small (0.10–0.29), moderate (0.30–0.49), large (0.50–0.69), very large (0.70–0.89), nearly perfect (>0.90) (Hopkins et al., 2009).

On the other hand, the inter-coder reliability between two researchers who evaluated 20 studies with the TESTEX scale was assessed according to the inter-class correlation coefficient (ICC two-way random absolute agreement), and the specified reference values and the level of reliability were reported (Koo and Li, 2016); poor (>0.50), moderate (>0.50-0.75), good (0.75-0.90), and excellent (0.90-1). Finally, the consistency between the two researchers who made the methodological quality assessment with the TESTEX scale was evaluated according to the Bland Altman graph, and the results were reported visually.

While the R version 4.1.0 (R Core Team) software was preferred for this study's statistical analysis, the statistical significance level was accepted as $\alpha = 0.05$ in all analyses. Detailed information about the statistical analysis made throughout the study is given in Table 1.

Table 1. Statistical analysis summary of the study

Assessment	Aim	Analysis
Checking translation evaluations by five independent language experts	Consistency between independent language experts	Kendall-W agreement coefficient analysis
Conformity of translated quality criteria according to independent language experts	Content validity	Content validity ratio Content validity index
Evaluation of 20 studies by two researchers with the Turkish and English forms of the TESTEX scale	Language validity	Pearson correlation coefficient (r) analysis Fisher'z effect size
Checking the similarity of the quality assessment scores given by the indepent two researchers	Inter-rater reliability	Inter-class correlation coefficient (ICC) analysis
Checking the consistency between the quality assessment scores given by the two researchers	Consistency between evaluators	Bland Altman graph

Not. TESTEX: Tool for the assessment of study quality and reporting in exercise; ICC: Inter-class correlation coefficient; r: Pearson correlation coefficient.

RESULTS

The content validity of the TESTEX scale, which was translated into Turkish according to the evaluations made by five independent language experts, was analysed according to the CVR and CVI values. After the initial assessments by independent language experts, it was understood that the Turkish translation of the first, tenth, and eleventh quality criteria needed revision. After the change, five independent language experts re-evaluated the Turkish translation of the TESTEX scale. As a result, it was determined that the Turkish translation of the TESTEX scale provided content validity (CVI = 1). On the other hand, the scores given by five independent language experts to Turkish translations for quality criteria were analysed with the Kendall-W coefficient of agreement and revealed moderate agreement between independent language experts. In addition, no statistically significant difference was found between the scores given by language experts to Turkish translations (Kendall's W = 0.26; p = 0.19).

Two independent researchers evaluated 20 studies on the subject with the Turkish and English versions of the TESTEX scale, and it was found that there was an excellent correlation between the assessments made in different languages (r = 0.99, 95% CI = 0.98 – 0.99, z = 2.64, p = 0.00). Detailed information on language validity is given in Figure 2.

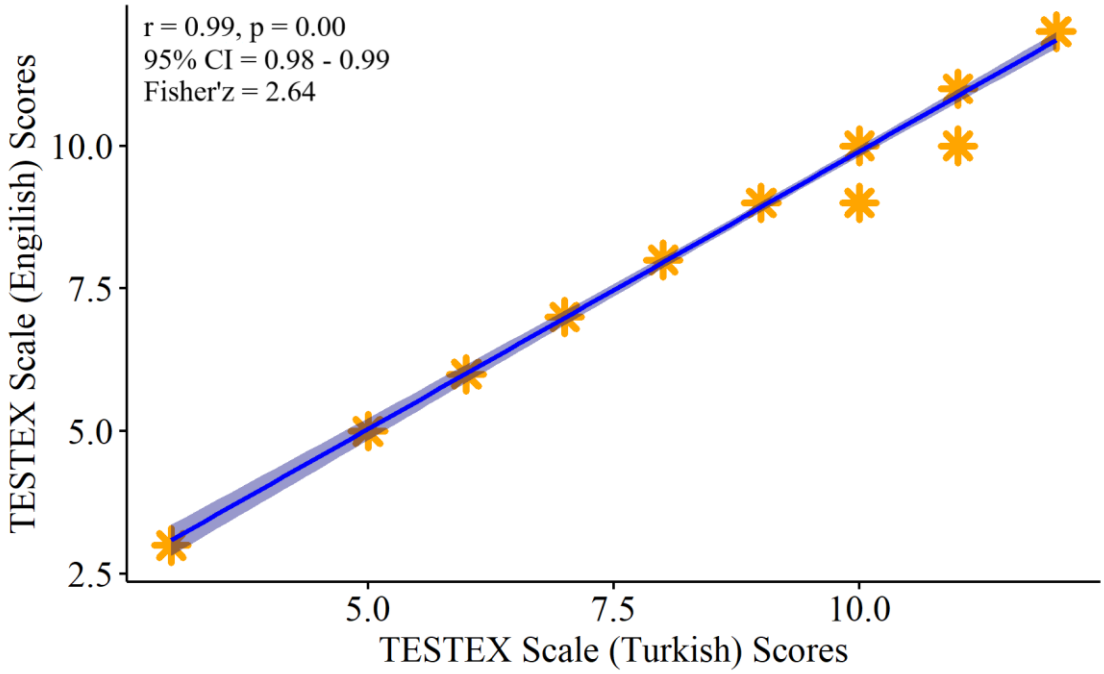

Figure 2. Language validity results of the TESTEX methodological quality assessment scale

In addition to the language validity, the consistency of the scores given by the two independent researchers to the research articles' quality was evaluated with the inter-class correlation coefficient, and the results showed an excellent level of reliability between the evaluators (ICC = 0.99, 95% CI = 0.98 – 1, p = 0.00). Finally, the mean bias between the scores given by the evaluators to the research articles' quality was examined with the Bland-Altman graph, and it was determined that there was a minimal bias between the two evaluators (Mean bias = -0.40 ± 1.17). The detailed analysis results for the consistency between evaluators are given in Figure 3.

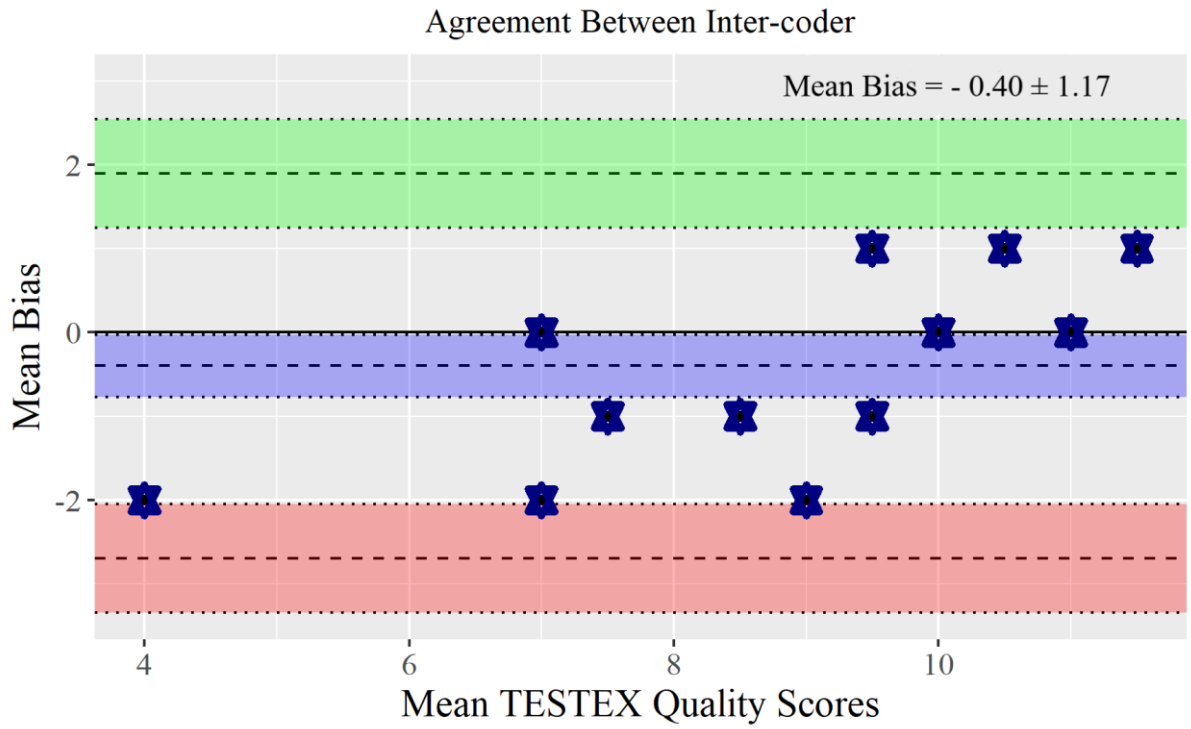

Figure 3. Results of consistency between evaluators using the TESTEX methodological quality assessment scale

DISCUSSION AND CONCLUSION

This study aimed to adapt the TESTEX methodological quality assessment scale into Turkish to evaluate the methodological quality of research articles in the field of exercise and sports. The study results showed that the TESTEX scale was adapted into Turkish as valid (CVI = 1; Kendall's W = 0.26; p = 0.19; r = 0.99, 95% CI = 0.98 – 0.99, z = 2.64, p = 0.00), and reliable (ICC = 0.99, 95% CI = 0.98 – 1, p = 0.00; Mean bias = -0.40 ± 1.17). Although there was a moderate level of agreement among independent language professionals, the results were statistically insignificant. The concept of statistical power can explain these results. When the research hypothesis cannot be tested with sufficient observations, statistically non-significant results may occur (Abt et al., 2020). The fact that only five independent language professionals participated in this study may not have provided a sufficient number of observations to test the hypothesis. Previous studies reported statistically insignificant results in the Kendal W agreement coefficient due to a similar reason (Çalışkan and Çınar, 2012; Hür et al., 2020). These results were attributed to the limited number of evaluators involved in the study. One of the results that needs to be explained is the standard deviation, which is observed to be high in mean bias values. Considering that evaluators assess out of 15 points, a standard deviation of 1.17 points may not significantly impact the methodological quality of research articles. Therefore, although the average bias and standard deviation obtained may appear high, they may be insignificant when evaluating the results in practice.

Many factors can influence the methodological and reporting quality of research articles in the field of exercise and sport (Abt et al., 2022; Bishop, 2008; Rico-González et al., 2021; Sainani and Chamari, 2022). Therefore, many researchers have recommended various methods to

improve the methodological quality of research articles (Abt et al., 2022; Bishop, 2008). In addition, researchers have claimed that studies designed according to various quality criteria can provide a better interpretation of sports performance (Bishop, 2008; Sainani and Chamari, 2022). While the TESTEX scale can be used as the methodological quality assessment tool in systematic reviews, it can also provide important clues to researchers who will conduct research in the field of exercise and sports to improve the quality of their studies. The TESTEX scale is divided into two sections, each containing 12 quality criteria for evaluating the quality of research articles (Smart et al., 2015). The Turkish version of the TESTEX methodological quality assessment scale is presented in Table 2.

While the TESTEX scale includes five quality criteria in the study quality section, these criteria offer researchers valuable insights into selecting sample groups and assessing their performance. Eligibility criteria express the requirement that participants should have similar characteristics. In addition, it has been stated that eligibility criteria are a prerequisite for the success of experimental studies (Su et al., 2023). Researchers realized that accurately determining eligibility criteria is crucial for predicting study results (Su et al., 2023). The second criterion in the "study quality" section highlights that sample groups should be distributed through randomization, and these randomization methods also need to be clearly stated. Randomization can be defined as the process of randomly assigning participants to either the experimental or control group in a research study (Alferes, 2012). Various randomization methods can be used for this purpose (Alferes, 2012). If a study fails to perform the necessary randomization, it may result in selection bias, harming its internal validity (Alferes, 2012; Berger et al., 2021). As the third criterion, concealing the group allocation from the participants may be necessary for improving the research articles' quality. Participants' learning of their group allocation may affect the reliability of the research intervention (Schulz, 2001; Sil et al., 2019). Therefore, researchers stated that allocation concealment helps to minimize selection bias in randomized controlled experiments (Kahan et al., 2015; Schulz, 2001; Sil et al., 2019). Similarly, independent researchers can provide support in evaluating of primary outcome measures. If the researchers who conducted the study evaluate the intervention results or are aware of the group allocations, it may lead to selection bias. Therefore, the reliability of the study may be negatively affected (Sil et al., 2019; Smart et al., 2015). Finally, the fact that the experimental and control groups had similar performance criteria in the pre-test is significant in terms of demonstrating the effectiveness of the intervention in the post-test. If the randomization of the groups is successful and the sample groups are similar in the pretest, researchers may argue that the observed outcomes between groups are due to the intervention rather than differences in baseline characteristics (de Boer et al., 2015; Elkins, 2015).

The TESTEX scale comprises seven distinct reporting criteria, in addition to the study quality criteria. Reporting criteria focus on the significance details presented in a study’s method and conclusion sections. These criteria cover participation rates in research interventions and testing, application, and writing of statistical analyses, monitoring of experimental and control groups, and providing details about these groups. A long interval of time may be necessary to observe the effects of interventions on various motor skills (Smart et al., 2015). However, participants may withdraw from research or cannot attend intervention sessions for various

reasons (Smart et al., 2015). Therefore, researchers stated that the percentage of participants who attend all intervention sessions should be reported, and their dependence on intervention sessions should aim for at least 85% (Smart et al., 2015).

On the other hand, researchers have argued that if the participants leave of a study for various reasons, their data should still be included in the analysis (Williams et al., 1999). This analysis method referred to as "intent-to-treat" in the literature. Intent-to-treat (ITT) analysis is a statistical method in which all randomized participants are included in the statistical analysis (Williams et al., 1999). According to this method, participants are analysed based on the group they were initially assigned, regardless of which treatment or intervention (McCoy, 2017; Williams et al., 1999). The ITT method emphasizes that even if a participants left of the research for various reasons, their data should still be included in the analysis. Howewer, researchers have also stated that this method has advantages and disadvantages (McCoy, 2017). Inter-group comparisons are necessary to reveal the intervention effect in a study (Marusteri and Bacarea, 2010). Reporting criteria should include performing hypothesis tests and presenting these tests with point estimation and variability measures such as arithmetic mean ± standard deviation, as a methodological quality criterion (Smart et al., 2015). Finally, researchers have recommended that the method section includes information about the specific activities carried out by experimental and control groups performed during the intervention to be reported in the method section (Smart et al., 2015).

Researchers conducting systematic reviews in exercise and sports may prefer to use the TESTEX scale to evaluate the methodological quality of research articles. Furthermore, researchers who want to improve the methodological quality of their studies can use the quality criteria of the TESTEX scale to design their studies. To assign points to the eligibility criteria, which is the first criterion in the evaluation phase, evaluators can explore the participants (or universe/sample) paragraph of the method section. In this section, authors typically provide details about the participants. If the authors have clearly stated the characteristics of the participants (in tabular form if possible), 1 point can be given to this criterion. For group randomization, the participants (or universe/sample) paragraph of the method section can be reviewed. The authors may include details about the randomization process in the research model (or Experimental Approach to the Problem) paragraph of the method section. If the authors clarify that group allocation was randomized, evaluators may assign a score of 1 to this criterion. The third and fifth criteria, which involve concealing the group allocation from participants and researchers, cannot be applied in our field for various reasons, or the authors may not have clearly expressed the implementation procedures. Therefore, evaluators assign 0 points to these criteria in numerous systematic reviews (Cuthbert et al., 2020; Vasconcelos et al., 2020). Suppose the authors clearly state that the group allocation is concealed from participants or evaluators. In this scenario, evaluators may allocate 1 point to the third criterion (concealment to participants) and 1 point to the fifth criterion (concealment to authors). The fourth criterion, the similarity of the research groups in the pre-test, can be evaluated by analysing the first paragraph of the results section. Authors may provide details of the pre-test results of the groups in the first paragraph of the results section to demonstrate the statistical differences between the groups as a result of the intervention. In this section, the authors typically indicate that the groups were similar in the pre-test by stating, "No significant

differences were found in the baseline...". If the authors have not given details about the pre-test in this section, the tables in the results section can be used as a secondary option for the evaluation. If the evaluators can determine that the groups are similar in the pre-test, in written or tabular form, 1 point can be given for the fourth criterion. The sixth criterion is the quality criterion and the TESTEX scale gets the highest score for it. In this criterion, the authors should provide the exercise participation rates for both the experimental and control groups. Additionally, the authors should indicate whether any adverse events, such as injuries or illnesses, occurred during the intervention and provide information on the source of these events. Finally, the authors should indicate the percentage or number of participants who completed the intervention process. When scoring this criterion, evaluators can explore the method section's participants (or universe/sample) paragraph to obtain knowledge about adverse events.

On the other hand, the first paragraph of the results section can be examined to check the intervention dependency ratio. In some cases, the authors can also present the number of participants included in the pre-test and post-test with the tables in the results section. Evaluators may award 1 point to this criterion if an adverse event is reported. If information about the adherence rates to the exercise and intervention process can be obtained in this criterion, assessment can be made for the sixth criterion out of 3 points. To evaluate the seventh criterion, the number of participants included in the pre-test and in the post-test can be examined. If the number of participants included in the pre-test and post-test is the same, or if the data of the participants who left the research are included in the analysis, 1 point can be given for the seventh criterion. The results section for the eighth and ninth criteria should be explored. If inter-group comparisons are reported with more than one outcome criterion in a study, 2 points can be given to the eighth criterion. On the other hand, if the research results are reported together with the arithmetic mean and standard deviation values, then it can be evaluated as 1 point for the ninth criterion. The training program (or research procedure) paragraph of the method section can be examined for other quality criteria. If the authors reported the activity of the control group during the study, 1 point could be award to the tenth criterion. Evaluators may assign 1 point to the eleventh criterion if there has been a change in training load (scope, intensity, intensity, frequency) at least once during the intervention period. Finally, if the authors clearly stated the intervention group's training program (table as possible), evaluators can make a 1-point assessment for the twelfth criterion.

As a result, researchers can use the Turkish version of the TESTEX scale as valid and reliable in their systematic reviews. Expressed suggestions can be considered when evaluating the TESTEX scale. In addition, researchers who want to improve the methodological quality of studies can examine the quality criteria of the TESTEX scale.

Acknowledgments
We would like to thank Mehmet Ulaş, Onursal Aksakallı, Yüksel Uysal, Taisia Konduyeva and Alper Delikanlı, who contributed to our study as independent language expert in the language translation phase of this study.

Conflict of Interest: There is no personal or financial conflict of interest between the authors of this study and the independent experts who contributed to the research.

Author Contributions: Research Design - HŞU and AYK.; Data Collection - AYK, OD, SNU, SK; Statistical Analysis - HŞU; Preparation of the article - HŞU, AYK, OD, SNU, SK.

Ethical Approval
Ethics Committee: Burdur Mehmet Akif Ersoy University Non-invasive Clinical Research Ethics Committee
Date: 01.03.2023
Protocol Number: 2023/03

REFERENCES

Abt, G., Jobson, S., Morin, J. B., Passfield, L., Sampaio, J., Sunderland, C., & Twist, C. (2022). Raising the bar in sports performance research. *Journal of Sports Sciences*, *40*(2), 125–129. https://doi.org/10.1080/02640414.2021.2024334

Acosta, S., Garza, T., Hsu, H. Y., & Goodson, P. (2020). Assessing quality in systematic literature reviews: A study of novice rater training. *SAGE Open*, *10*(3), 1–11. https://doi.org/10.1177/2158244020939530

Alferes, V. R. (2012). *Methods of randomization in experimental design*. SAGE Publications.

Berger, V. W., Bour, L. J., Carter, K., Chipman, J. J., Everett, C. C., Heussen, N., Hewitt, C., Hilgers, R. D., Luo, Y. A., Renteria, J., Ryeznik, Y., Sverdlov, O., Uschner, D., & Beckman, R. A. (2021). A roadmap to using randomization in clinical trials. *BMC Medical Research Methodology*, *21*(1), 1–24. https://doi.org/10.1186/S12874-021-01303-Z/TABLES/3

Bishop, D. (2008). An applied research model for the sport sciences. *Sports Medicine*, *38*(3), 253–263. https://doi.org/10.2165/00007256-200838030-00005

Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş., & Demirel, F. (2022). *Eğitimde Bilimsel Araştırma Yöntemleri* (33. baskı). Pegem Akademi.

Cuthbert, M., Ripley, N., Mcmahon, J. J., Evans, M., Haff, · G Gregory, & Comfort, P. (2020). The effect of nordic hamstring exercise intervention volume on eccentric strength and muscle architecture adaptations: A systematic review and meta-analyses. *Sports Medicine*, *50*, 83–99. https://doi.org/10.1007/s40279-019-01178-7

Çalışkan, T., & Çınar, S. (2012). Akran desteği: Geçerlik güvenirlik çalışması. *Marmara Üniversitesi Sağlık Bilimleri Enstitüsü Dergisi*, *2*(1), 1-7.

Davies, T. B., Tran, D. L., Hogan, C. M., Gregory Haff, G., & Latella, C. (2021). Chronic effects of altering resistance training set configurations Using Cluster Sets: A systematic review and meta-analysis. *Sports Medicine*, *51*, 707–736. https://doi.org/10.1007/s40279-020-01408-3

de Boer, M. R., Waterlander, W. E., Kuijper, L. D. J., Steenhuis, I. H. M., & Twisk, J. W. R. (2015). Testing for baseline differences in randomized controlled trials: An unhealthy research behavior that is hard to eradicate. *International Journal of Behavioral Nutrition and Physical Activity*, *12*(1), 1–8. https://doi.org/10.1186/S12966-015-0162-Z/TABLES/1

Elkins, M. R. (2015). Assessing baseline comparability in randomised trials. *Journal of Physiotherapy*, *61*(4), 228–230. https://doi.org/10.1016/J.JPHYS.2015.07.005

Haff, G. G., Bishop, D., Hoffman, J., Kawamori, N., Newton, R. U., Sands, B., & Stone, M. (2010). Sport Science. *Strength and Conditioning Journal*, *32*(2), 33–45. https://doi.org/10.1519/SSC.0B013E3181D59C74

Hoffman, J. R. (2002). *Physiological aspects of sport training and performance.* Human Kinetics Publishers Inc.

Hür, G., Zengin, H., Karakaya Suzan, Ö., Kolukisa, T., Eroğlu, A., & Çinar, N. (2022). Joanna Brıggs Enstitüsü (JBE) randomize kontrollü çalışmalar için kalite değerlendirme aracının Türkçe'ye uyarlaması. *Sağlık Bilimlerinde İleri Araştırmalar Dergisi,5*(2), 112–117. https://doi.org/10.26650/JARHS2022-1100929

Hopkins, W.G., Marshall, S.W., Batterham, A.M., & Hanin, J. (2009). Progressive statistics for studies in sports medicine and exercise science. *Medicine Science Sports*, *41*(1), 3–12. https://doi.org/10.1249/MSS.0b013e31818cb278

Juhl, C. B., & Lund, H. (2018). Do we really need another systematic review?. *British Journal of Sports Medicine*, *52*(22), 1408–1409. https://doi.org/10.1136/BJSPORTS-2018-099832

Kahan, B. C., Rehal, S., & Cro, S. (2015). Risk of selection bias in randomised trials. *Trials*, *16*(1), 1–7. https://doi.org/10.1186/S13063-015-0920-X/TABLES/3

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/J.JCM.2016.02.012

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, *28*(4), 563–575. https://doi.org/10.1111/J.1744-6570.1975.TB01393.X

Maher, C. G., Sherrington, C., Herbert, R. D., Moseley, A. M., & Elkins, M. (2003). Reliability of the PEDro scale for rating quality of randomized controlled trials. *Physical Therapy*, *83*(8), 713-721. https://doi.org/10.1111/j.1365-2753.2005.00574.x

Marusteri, M., & Bacarea, V. (2010). Comparing groups for statistical differences: How to choose the right statistical test? *Biochemia Medica*, *20*(1), 15–32. https://doi.org/10.11613/BM.2010.004

McCoy, C. E. (2017). Understanding the Intention-to-treat Principle in Randomized Controlled Trials. *Western Journal of Emergency Medicine*, *18*(6), 1075. https://doi.org/10.5811/WESTJEM.2017.8.35985

Meel, P., & Vishwakarma, D. K. (2020). Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. *Expert Systems with Applications*, *153*, 112986. https://doi.org/10.1016/J.ESWA.2019.112986

Murad, M. H., Asi, N., Alsawas, M., & Alahdab, F. (2016). New evidence pyramid. *BMJ Evidence-Based Medicine*, *21*(4), 125–127. https://doi.org/10.1136/EBMED-2016-110401

O'Brien, B. C., Harris, I. B., Beckman, T. J., Reed, D. A., & Cook, D. A. (2014). Standards for reporting qualitative research: a synthesis of recommendations. *Academic Medicine*, *89*(9), 1245-1251. https://doi.org/10.1097/ACM.0000000000000388

Rico-González, M., Pino-Ortega, J., Clemente, F. M., & Arcos, A. L. (2021). Guidelines for performing systematic reviews in sports science. *Biology of Sport*, *39*(2), 463–471. https://doi.org/10.5114/BIOLSPORT.2022.106386

Sainani, K., & Chamari, K. (2022). Wish List for improving the quality of statistics in sport science. *International Journal of Sports Physiology and Performance*, *17*(5), 673–674. https://doi.org/10.1123/IJSPP.2022-0023

Schulz, K. F. (2001). Assessing allocation concealment and blinding in randomised controlled trials: why bother? *Evidence-Based Nursing*, *4*(1), 4–6. https://doi.org/10.1136/EBN.4.1.4

Sil, A., Kumar, P., Kumar, R., & Das, N. K. (2019). Selection of control, randomization, blinding, and allocation concealment. *Indian Dermatology Online Journal*, *10*(5), 605. https://doi.org/10.4103/IDOJ.IDOJ_149_19

Smart, N. A., Waldron, M., Ismail, H., Giallauria, F., Vigorito, C., Cornelissen, V., & Dieberg, G. (2015). Validation of a new tool for the assessment of study quality and reporting in exercise training studies: TESTEX. *International Journal of Evidence-Based Healthcare*, *13*(1), 9–18. https://doi.org/10.1097/XEB.0000000000000020

Su, Q., Cheng, G., & Huang, J. (2023). A review of research on eligibility criteria for clinical trials. *Clinical and Experimental Medicine*, 13. https://doi.org/10.1007/S10238-022-00975-1

Tekindal, M., Özden, S. A., Gedik, T. Ege, A., Erim, F., & Tekindal, M. A. (2021). Nitel araştırmaların raporlanmasında standartlar: SRQR kontrol listesinin Türkçe uyarlaması. *OPUS Uluslararası Toplum Araştırmaları Dergisi*, *18*(42), 5425–5443. https://doi.org/10.26466/OPUS.882177

Vasconcelos, B. B., Protzen, G. V., Galliano, L. M., Kirk, C., & Del Vecchio, F. B. (2020). Effects of high-intensity interval training in combat sports: A systematic review with meta-analysis. *Journal of Strength and Conditioning Research*, *34*(3), 888–900. https://doi.org/10.1519/JSC.0000000000003255

Yeşilyurt, S., & Çapraz, C. (2018). Ölçek geliştirme çalışmalarında kullanılan kapsam geçerliği için bir yol haritası. *Erzincan University Journal of Education Faculty*, *20*(1), 251–264. https://doi.org/10.17556/ERZIEFD.297741

Williams, G. D., Stinson, F. S., Lane, J. D., Tunson, S. L., & Dufour, M. C. (1999). What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ*, *319*(7211), 670–674. https://doi.org/10.1136/BMJ.319.7211.670

Table 2. Turkish version of the TESTEX methodological quality assessment scale

Kriter	Açıklama	Puanlama
Çalışma kalitesi		
1 – Belirtilen uygunluk kriterleri	Uygunluk kriterleri belirlenmeli, yerine getirilmeli ve tüm katılımcılar için spesifik tanısal test değerleri sağlanmalıdır.	1 Puan – uygunluk kriterleri açıkça belirtilmiş ve yerine getirilmişse
2 – Belirtilen randomizasyon	Denekleri tedavi gruplarına ayırmak için kullanılan yöntemin tanımlanması sağlanmalıdır.	1 Puan – yöntemler açıklanırsa ve bunlar gerçekten randomize ise (örneğin bozuk para atma, rastgele üretilen sayıların sırası)
3 – Ayırmayı gizleme	Grup ayırımının gizlenip gizlenmediği belirtilmelidir; yani bir deneğin araştırmaya dahil edilmeye uygun olup olmadığı belirtilmeli ve denek hangi gruba ayrıldığından haberdar olmamalıdır (bu karar verildiğinde).	1 Puan – grup ayırımı çalışmaya dahil edilmeye uygun deneklerden gizlenmişse (örn. randomizasyondan önce onay verilmelidir)
4 – Başlangıçta benzer gruplar	Randomize edilen tüm katılımcıların başlangıç verileri sunulmalıdır. Tedavi grupları arasında tedavi edilen durumun ciddiyetinin ölçülmesinde anlamlı bir fark olmamalıdır.	1 Puan – temel veriler grup ayrımına göre ayrılır, sunulur ve hiçbir fark görünmezse
Tüm katılımcıların körlenmesi Tüm terapistlerin (eğiticilerin) kör edilmesi	Bu madde puanlanmadı. Bu madde puanlanmadı.	Puan yok Puan yok
5 – Değerlendiricinin körleştirilmesi (en az bir önemli çıktı için)	Denekleri ve/veya terapistleri körlemek her zaman mümkün değildir; Bununla birlikte, değerlendiricilerin körlenmesi uygundur. Birincil çıktı ölçütlerinin değerlendiricileri deneklerin müdahale tahsisine körlenmiş ise, bu açıkça belirtilmelidir.	1 Puan - en az bir değerlendiricinin olduğu ifade edilmişsse 1 birincil çıktı ölçüsü grup tahsisine körlenmişse
Çalışmanın raporlaması		
6 – Deneklerin %85 'inde değerlendirilen çıktı ölçümleri	Her iki grupta da çalışmayı tamamlayan deneklerin yüzdesi bildirilmelidir. Her müdahale grubu için herhangi bir olumsuz olay (ciddi tıbbi olaylar, ölümler, hastaneye yatışlar vb.) rapor edilmelidir. Çalışmadan çekilmeyen egzersiz deneklerinin tamamladıkları egzersiz seanslarının yüzdesi bildirilmelidir.	Puan yok – çalışmadan çekilmeler >%15 ise 1 Puan – çalışmaya bağlılık >%85 ise 1 Puan – olumsuz olaylar bildirilirse 1 Puan – egzersize katılım bildirilirse Toplam Olası Skor- 3 puan
7 – Müdahele amacına uygunluk analizi	Bir denek çalışmadan geri çekildiğinde, bu analiz, çıktı ölçütlerinin her biri için elde edilen son değerin müdahale sonrası değer olarak kullanılması veya temel değerinin son değer olarak kullanılmasıyla gerçekleştirilir. Bu analiz, çalışmayı tamamlayanların verilerine ve yapılan analizlere eklenmelidir.	1 Puan – çıktılar üzerinde müdahele amacına uygunluk analizi yapıldıysa
8 – Rapor edilen gruplar arası istatistiksel karşılaştırmalar	Egzersiz ve karşılaştırma (kontrol) grubunun birincil ve en az bir ikincil çıktı için karşılaştırılması yapılmalıdır.	1 Puan – ilgili birincil çıktı ölçüsü için gruplar arası istatistiksel karşılaştırmalar rapor edilirse 1 Puan – en az bir ikincil çıktı ölçümü için gruplar arası istatistiksel karşılaştırmalar rapor edilirse Toplam Olası Skor - 2 puan
9 – Rapor edilen tüm çıktı ölçümleri için nokta tahminleri ve değişkenlik ölçümleri	Nokta tahminleri tüm çıktılar için sağlanmalıdır, aksi takdirde bu yanlı çıktı raporlaması olarak kabul edilebilir.	1 Puan – tüm sonuçlar nokta tahminleriyle raporlanırsa
10 – Kontrol gruplarında aktivite izleme	Kontrol grubundaki denekler müdahale grubuna geçerse gruplar arasındaki farklılıklar azalabilir. Deneklerin üçte biri bunu yaparken, bu etkinin ölçülebilmesi ve sayısallaştırılması için egzersiz günlüğü veya aktivite izleme gibi bazı önlemlerin sağlanması gerekir.	1 Puan – kontrol grubundaki deneklerden fiziksel aktivite seviyelerini bildirmeleri istenirse ve veriler sunulursa
11 – Sabit sürdürülen bağıl egzersiz şiddeti	Egzersiz şiddeti birçok kişi tarafından adaptasyon için en iyi uyarıcı olarak kabul edilir. Denekler belirli bir şiddet ile bir egzersiz programına başladıktan sonra o egzersiz programına uyum sağlamaya başlarlar. Çalışma süresi boyunca bağıl şiddet, egzersize uyum sağlayanlarda düşecektir. Bu yüzden egzersiz kapasitesinin periyodik olarak değerlendirilmesi ve egzersiz şiddetinin sabit kalması için egzersiz şiddetin güncellenmesi (veya uyum sağlayanlarda güncellenmesi) gerekir.	1 Puan – Bağıl şiddeti sabit tutmak için egzersiz yükü güncellenirse
12 – Egzersiz yoğunluğu ve enerji tüketimi	Egzersiz parametreleri; seans, program süresi, seans sıklığı, egzersiz antrenman şiddeti ve yöntemi olarak açık bir şekilde raporlanmalıdır.	1 Puan – egzersiz yoğunluğu ve enerji tüketimi hesaplanabilirse
	Olası toplam	**15 puan**