Cypriot Journal of Educational Sciences

Volume 16, Issue 2, (2021) 859-868

www.cjes.eu

Development of a cello performance measurement tool: Validity and reliability study

Mehmet Can Çiftçibaşı [a] [1], Burdur Mehmet Akif Ersoy University, Faculty of Education, Burdur, 15030, Turkey, https://orcid.org/0000-0002-8825-1001

Suggested Citation:

Çiftçibaşı, M, C. (2021). Development of a cello performance measurement tool: Validity and reliability study. *Cypriot Journal of Educational Science*. *16*(2), 859-868. https://doi.org/10.18844/cjes.v16i2.5688

Received from January 05, 2021; revised from February 02, 2021; accepted from April 10, 2021.

Selection and peer review under responsibility of Prof. Dr. Huseyin Uzunboylu, Higher Education Planning, Supervision, Accreditation and Coordination Board, Cyprus.

Abstract

The aim of this study was to develop a new performance measurement tool with proven validity and reliability to be used in the semester final exams of the individual instrument (cello) courses in institutions in which music teachers are trained. The study has a descriptive nature and a documentary scan was made during the preparation of the items to be included in the measurement tool. The participants of the study comprised 17 instructors who teach cello in 17 different universities in Turkey and a total of 50 students from 4 different music teaching programmes. In the study, two different data collection tools were used, namely the opinion form applied to the instructors of the cello lesson and the measurement tool developed by the researcher, in order to determine the importance of the criterion skills in the tool. Internal consistency was examined for the reliability of the tool and Cronbach's alpha and Cohen's kappa formulas were used. The Cronbach's alpha reliability coefficient was determined to be .94 and the statistical significance level was determined to be .05. As a result of the study, a valid and reliable measurement tool for measuring cello performance was developed.

Keywords: Cello education, Measuring cello performance, Performance measurement tool

* ADDRESS FOR CORRESPONDENCE: Mehmet Can, Çiftçibaşı, Burdur Mehmet Akif Ersoy University, Faculty of Education, Music Education Department, Burdur, Turkey
E-mail address: mcciftcibasi@gmail.com / Tel.: +90-248-213-1000/4266

1. Introduction

Among the important dimensions of string instrument training is the measurement of musical performance. A healthy measurement and evaluation are of great importance in determining to what extent the behaviours that are aimed to be acquired are realised and how an individual develops during the instrument training process. In addition to the technical infrastructure, it is possible to achieve a good performance with the combination of many correct components. Variables such as the period characteristics of the works, technical difficulties, and the musical expressions stated by the composer in his works as well as the expressiveness, intellectual accumulation, and musical perception of each player are the factors that affect the measurement of musical performance in a healthy way. As a result of the data obtained from the measurement, reorganisation of the training process, making future-oriented decisions, implementing a different training programme, or determining and procuring the necessary materials can be achieved while the process is ongoing; in other words, before it is too late.

For musicians and educators, determining the factors that affect the evaluation of performance quality enables conscious analysis of musical performance quality (Russell, 2010). One of the major difficulties in assessing complex behaviour such as music performance is that the measurements used are typically subjective judgments based on erratic and uncontrolled observations. It is difficult to achieve consensus among raters about the adequacy of a performance. One method for improving the assessment of musical performance is to replace the scorings raters give with their overall impressions via rating scales obtained through more systematic procedures (Abeles, 1973).

“In terms of musical education, which is a process, it is necessary to base musical performance-based teaching and testing this teaching process on an objective basis” (Schleuter, 1996). Environments to test musical performance without using a written measurement tool with validity and reliability will not be free of subjective judgments (Özdemir, 2012). If a written measurement tool is not used, the student who wants to get feedback about his/her performance will be incompletely informed related to it and this will result in his/her not getting healthy feedback regarding his/her development. However, when considered in the long term, it will not be possible to determine the developing/not developing and functioning/not functioning aspects of the teaching process (Dalkıran, 2006). According to Fiske, many evaluators do not know how they determine their performance grades (as cited in Ciorba & Smith, 2009). Music teachers generally do not use a measurement tool to keep track of the progress of their students in their instruments. In fact, it is seen that some teachers make evaluations by using criteria that are unrelated to instrument playing skills such as behaviour, attitude, and attendance (Wesolowski, 2012). According to MENC, music performance evaluations are reliable to the extent that the same skill or knowledge measures give the same results, each performance is evaluated with the same procedures, and raters share the same expectations (MENC, 1996 cited in Latimer, Bergee, & Cohen, 2015).

It is obvious that performance tests provide an undeniable benefit in music education. However, the tests should be continuously developed in order to take advantage of their expected benefit. The scientific method to be followed in development is the experimental (empirical) approach consisting of test preparation, application, evaluation, and correction phases (Uçan, 2005).

The aim of the present study was to develop a new performance measurement tool with proven validity and reliability to be used in the semester final exams of the individual instrument (cello) course in institutions in which music teachers are trained. The problem sentence of the study was determined to be "Is the measurement tool developed for the performance dimension of the cello lesson valid and reliable in institutions that train music teachers?" The sub-problems examined regarding the aim of the study and the specified problem are given below.

1. Is the measurement tool developed within the scope of the study a reliable measurement tool?
2. Is the measurement tool developed within the scope of the study a valid measurement tool?

2. Method

2.1. Study model and pattern

This study is of descriptive nature in terms of ensuring the development of a valid and reliable measurement tool in order to evaluate student performances in the semester final exams of cello lessons in institutions in which music teachers are trained. Documentary screening was performed at the stage of the grounding of the study and the preparation of the items to be included in the measurement tool, the relevant literature was scanned, the relevant curriculum was analysed, and the criteria skills in the measurement tool were determined. In the second stage of the study, the importance coefficients of the items were created in line with the opinions of 17 field experts who were cello instructors in 17 different universities in Turkey. In the third stage of the study, testing of the validity and reliability of the measuring tool was carried out by passing on to the hypothesis process.

2.2. Participants

The participants of the study are 17 instructors who teach cello in 17 different universities in Turkey and a total of 50 students from 4 different music teaching programmes. In this context, demographic information about the 17 cello trainers whose opinions were sought during the process of determining the items and creating the importance coefficients of these items is presented in Table 1 and Table 2.

Table 1. Demographic information of the faculty members

		f
Title	Professor	1
	Assoc. Prof.	4
	Dr. Faculty Member	6
	Lecturer	4
	Research Associate	2
	Total	17
Educational Level	Doctorate	11
	Postgraduate	6
	Total	17
Experience	6-10 years	4
	11-15 years	3
	16-20 years	7
	21-25	3
	Total	17

According to Table 1, the cello teachers participating in the study consist of 1 Professor, 4 Associate Professors, 6 Dr Faculty Members, 4 Lecturers, and 2 Research Associates. Considering their years of experience, it is seen that there are 4 faculty members with between 6 and 10 years, 3 with between 11 and 15 years, 7 with between 16 and 20 years, and 3 with between 21 and 25 years. Accordingly, based on the seniority and education levels of the participants who gave their opinions in the study, they can be considered sufficiently experienced.

Table 2. Institutions where the instructors work

Institutions	f
Gazi University	2
Atatürk University	2
Pamukkale University	1
Aksaray University	1
Nevşehir Hacı Bektaş Veli University	1
Kafkas University	1
Sinop University	1
Dokuz Eylül University	1

Süleyman Demirel University	1
Van Yüzüncü Yıl University	1
Tokat Gaziosmanpaşa University	1
Kırıkkale University	1
Harran University	1
Cumhuriyet University	1
Ondokuz Mayıs University	1
Total	17

According to Table 2, 2 instructors each from Gazi University and Atatürk University and 1 instructor each from other universities gave their opinions.

Table 3. Distribution of students according to the university they study at

University	f
Mehmet Akif Ersoy University	16
Ondokuz Mayıs University	12
Pamukkale University	12
Gazi University	10
Total	50

As seen in Table 3, 50 students participated in the study, comprising 16 students from Mehmet Akif Ersoy University, 12 from Ondokuz Mayıs University, 12 from Pamukkale University, and 10 from Gazi University.

2.3. Data collection tools

Two different data collection tools were used, namely the opinion form applied to the instructors of the cello lesson and the measurement tool developed by the researcher, in order to determine the importance of the criterion skills in the developed measurement tool. The criterion skills in the measurement tool were created by obtaining the opinions of four instructors who were conducting the cello lesson in the departments of music education, together with an examination of the relevant literature and the cello lesson curriculum. In order to be used in the semester final exams of the cello lesson, it was decided to include 12 criterion skills considered to be important in the performance measurement tool. The criterion skills in the measurement tool are given below.

1. The work played being suitable for the grade level of the student
2. Being able to hold the cello and the string correctly
3. Being able to play the notes correctly and according to their values
4. Being able to apply the string techniques in the work correctly
5. Being able to play with good intonation
6. Being able to play with good articulation
7. Being able to play the work without any observable contraction or pain problems.
8. Being able to play in accordance with the specified speed terms
9. Being able to play in accordance with the specified musical dynamics and expressions
10. Being able to do vibrato effectively
11. Being able to play the work in accordance with the characteristics of the period
12. Impressing the audience musically

The importance level of the determined criteria skills was determined by applying the opinion form prepared as a five-point Likert scale to the instructors of the cello lesson.

In the applied questionnaire, the instructors were asked to give points from one to five for each criterion skill. Considering the mean rank of the given scores, the coefficients of the criterion skills were determined. According to the questionnaire results obtained, the mean rank values of each criterion skill are given in Table 4.

Table 4. Mean rank values of criterion skills

No	Criterion Skills	Mean Rank
1	The work played being suitable for the grade level of the student	4.11
2	Being able to hold the cello and the string correctly	4.17
3	Being able to play the notes correctly and according to their values	4.88
4	Being able to apply the string techniques in the work correctly	4.05
5	Being able to play with good intonation	4.94
6	Being able to play with good articulation	3.76
7	Being able to play the work without any observable contraction or pain problems.	3.23
8	Being able to play in accordance with the specified speed terms	3.64
9	Being able to play in accordance with the specified musical dynamics and expressions	3.17
10	Being able to do vibrato effectively	3.29
11	Being able to play the work in accordance with the characteristics of the period	3.17
12	Impressing the audience musically	3.23

Table 4 shows the mean rank values of the criterion skills. In order for the criterion skills to be grouped according to their importance, the starting mean rank was accepted as 3 points. After that score, when equal intervals were created with a difference of 0.40 points, the criterion skills were collected in 4 groups. When coefficients from 1 to 3 were given consecutively for each group, the lowest score was determined as 20 and the highest score as 100 in the measurement tool created. The coefficient values given according to the mean rank of the criterion skills are shown in Table 5.

Table 5. Coefficient values of criterion skills according to their mean rank

Criterion Skill No	Mean Rank Range	Coefficient	Total Score
7,9,10,11,12	3.00 – 3.40	1	25
6,8	3.41 – 3.80	1.5	15
1,2,4	3.81 – 4.20	2	30
3,5	4.61 – 5.00	3	30
Total Score of Criterion Skills in the Measurement Tool			100

2.4. Data analysis

The data obtained within the framework of the sub-problems of the study designed in line with the determined aim and problem were analysed with IBM SPSS Statistics 22. A significance level of .05 was chosen for the measurement of statistical significance. The validity and reliability of the scoring key created were tested by applying it to a group of 50 students. At that stage, the data were obtained by giving scores separately by the researcher and a field expert simultaneously.

2.4.1. Reliability

For testing the reliability of the measurement tool, the internal consistency was first examined. In the examination of the internal consistency of the obtained test scores, Cronbach's alpha formula (Büyüköztürk, 2010) was applied to the means of the scores given by the instructors. Cohen's kappa formula, which calculates the level of harmony between two raters (Kutlu et al., 2010), was used to determine the statistical appropriateness of the mean score.

Another element of scoring reliability sought in the study is the consistency between the total scores obtained from the measuring tool. In that case, the scores in the subcategories were ignored and the correlation between the total scores (Kutlu et al., 2010) was examined in order to calculate the consistency between the raters. In order to determine the sensitivity of the measurement tool developed, the relationship between the scores obtained by the traditional scoring method and those obtained from the measurement tool was examined with the Pearson moment correlation coefficient.

2.4.2. Validity

To test the validity of the measurement tool, the content validity was first examined. Principal component analysis was performed for construct validity, and item analysis was performed to examine item validity.

3. Findings

3.1. Findings regarding the reliability of the measurement tool

In order to be used in the individual instrument (cello) course of the Faculty of Education Department of Music Education, a Graded Scoring Key which includes different criteria, skills, and coefficients was developed in line with the curriculum and expert opinions. Findings regarding the validity and reliability of the measurement tool are given below.

3.1.1. Determining internal consistency

The Cronbach's alpha reliability coefficient of the measuring tool was calculated by taking the mean of the markings recorded by 2 cello instructors for each item, who evaluated the students using the measuring tool. In order to test whether the means of the markings recorded by the 2 instructors can be taken, the consistency of the scores given by these instructors with the measurement tool was examined with the kappa coefficient, which indicates the consistency between raters. In order to obtain consistency between the scores given by the raters, kappa values calculated for each criterion skill of the graded scoring key were found to be statistically significant ($p < 0.001$). The kappa values of the two raters are between 0.41 and 0.73. The mean of the kappa coefficients is .53. According to Landis and Koch (1977), this shows that there is "intermediate level" harmony between raters.

Table 6. Value ranges for the interpretation of kappa statistics

(κ) Value	Interpretation
< 0.00	Weak
0.00 – 0.20	Unimportant
0.21 – 0.40	Low
0.41 – 0.60	Medium
0.61 – 0.80	Important
0.81 – 1.00	Very High

Accordingly, the averaging process mentioned above was continued and the Cronbach's alpha reliability coefficient was determined as .94, which was obtained from the application of 12 items to 50 students and based on the mean item scores. This value shows that the items in the measurement tool measure with very high consistency with each other. Considering that the widely accepted lower limit of reliability coefficient for performance evaluation-based measurements is .70, this internal consistency coefficient obtained indicates that the measurement tool can be regarded as having a high level of reliability.

In order to determine the sensitivity of the measurement tool, its relationship with the traditional scoring method was examined and the findings obtained are presented below. The relationship between the scores obtained by the traditional scoring method and the scores obtained from the measuring tool was examined with the Pearson moment correlation coefficient and the results are presented in Table 7.

Table 7. Correlation between scores given using the measuring tool and scores given without using the measuring tool

Measurement Method	n	$\bar{x}$	sd	correlation	p
Traditional Measurement	50	59.40	21.60	0.970	0.000*
Measuring with Measurement Tool	50	61.42	19.43		

*p<0.05

As seen in Table 7, there is a significant relationship (.97) between the two scoring methods. This result can be considered to constitute important evidence of the sensitivity of the measuring tool, and it can be seen as important evidence in terms of showing the similarity of the measuring tool with an existing method.

3.2. Findings regarding the validity of the measurement tool

In order to examine the content validity, construct validity, and validity of each item regarding the validity of the measurement tool, lower–upper group item discrimination was examined. The evidence obtained for validity is reported below.

3.2.1 Content validity

In order to determine the content validity of the measurement tool, 4 instructors who teach cello in relevant departments of universities and educators who are experts in the fields of measurement and evaluation and statistics were interviewed. During these interviews, the opinions of the educators related to the content, expression, and scoring aspects of the measurement tool were obtained and corrections deemed necessary in line with their recommendations were made. The educators stated that the measuring tool was suitable for the purpose of measuring cello performance in terms of scope.

3.2.2. Construct validity

Main component analysis was applied to determine the construct validity of the measurement tool. In Table 8, measurement tool items and factor loads of these items are given.

Table 8. Statistical results showing the construct validity results of the measurement tool

	Criterion Skills	
1	The work played being suitable for the grade level of the student	.970
2	Being able to hold the cello and the string correctly	.969
3	Being able to play the notes correctly and according to their values	.967
4	Being able to apply the string techniques in the work correctly with a nice tone	.967
5	Being able to play with good intonation	.966
6	Being able to play with good articulation	.966
7	Being able to play the work without any observable contraction or pain problems.	.970
8	Being able to play in accordance with the specified speed terms	.967
9	Being able to play in accordance with the specified musical dynamics and expressions	.968
10	Being able to do vibrato effectively	.967
11	Being able to play the work in accordance with the characteristics of the period	.965
12	Impressing the audience musically	.965
	Cronbach's Alpha Coefficient	.97

When the component loads of each item in Table 8, which show the degree of serving the basic purpose of the measuring tool, are examined, it is seen that they are between .97 and .96. If the component loads and all items measure positively, the acceptable lower value limit is .30 (Dalkıran, 2006). It was observed that all factor loads in the developed measurement tool were well above the

specified limit. In that case, it can be concluded that all the items in the measurement tool are suitable for the purpose of measuring the performance of the cello in terms of construct.

3.3.3. Item validity review

Determining whether each item can statistically differentiate between students who are known to be in the upper group in terms of cello performance and those who are known to be in the lower group can be considered an indicator of whether that item is measuring in accordance with the purpose of the measurement tool. In other words, determining it means revealing the validity of the item. For this purpose, 50 students were divided into lower and upper groups of 27% in terms of test scores. In order to determine whether these two groups of students can be distinguished from each other in terms of item scores, all items were compared using the independent groups t-test. Findings regarding the analyses performed are shown in Table 9 below.

Table 9. Independent sample t-test results for each criterion skill between the lower and upper groups of 27% determined according to the performance measurement tool

Criterion Skill	Group	n	$\overline{x}$	sd	t	df	p
M1	upper27	14	4.2857	.72627	6.735	26	.000
	lower27	14	2.2143	.89258			
M2	upper27	14	4.4286	.75593	8.189	26	.000
	lower27	14	2.4286	.51355			
M3	upper27	14	4.5714	.51355	11.667	26	.000
	lower27	14	2.0714	.61573			
M4	upper27	14	4.2857	.72627	8.046	26	.000
	lower27	14	2.0714	.73005			
M5	upper27	14	4.2857	.46881	12.781	26	.000
	lower27	14	1.8571	.53452			
M6	upper27	14	3.7143	.82542	10.071	26	.000
	lower27	14	1.2143	.42582			
M7	upper27	14	4.0714	1.07161	5.430	26	.000
	lower27	14	2.3571	.49725			
M8	upper27	14	4.2143	.80178	8.477	26	.000
	lower27	14	1.8571	.66299			
M9	upper27	14	4.2143	.80178	9.417	26	.000
	lower27	14	1.6429	.63332			
M10	upper27	14	4.1429	.66299	12.449	26	.000
	lower27	14	1.2143	.57893			
M11	upper27	14	4.0000	.67937	11.746	26	.000
	lower27	14	1.3571	.49725			
M12	upper27	14	3.8571	.66299	13.435	26	.000
	lower27	14	1.1429	.36314			

*p<0.05

As can be seen in Table 9, all items differentiate the lower and upper groups sufficiently. Considering the obtained p values, it was determined that the developed measurement tool has item validity.

4. Discussion and conclusion

In order to determine the reliability of the developed cello performance measurement tool, its internal consistency was checked. First of all, the kappa test was applied to determine the usability of the mean scores given by the raters, and it was observed that there was consistency between raters in line with the results obtained. Then the Cronbach's alpha test was applied over the averages of the

raters' scores, and it was found that the reliability coefficient and the measuring tool had high internal consistency. In order to determine the sensitivity of the measuring tool, its relationship with the traditional scoring method was examined with the Pearson moment product correlation coefficient. Based on the findings, the existence of a significant relationship between the two scoring methods shows that the measuring tool has sensitivity.

In order to determine the validity of the developed measurement tool, the content validity was examined first. Field experts, whose opinions were obtained, stated that the measurement tool was suitable for the purpose of measuring cello performance in terms of content, expression, and scoring. Main component analysis was applied in order to determine the construct validity of the measuring tool and it was found that all factor loads were well above the acceptable limit. Thus, it can be surmised that all the items in the measurement tool are suitable for the purpose of measuring the performance of the cello in terms of construct.

The measurement tool developed was applied to the students in order to be used during the evaluation to determine item validity. These students were divided into lower and upper groups of 27% in terms of their achievement levels, and a t-test was applied to both groups for each criterion skill. It was determined that the measurement tool has item validity due to the significant difference found by the t-test.

The tool for measuring cello performance developed consists of 3 dimensions in total: compliance with the requirements of the program, technical skills, and musical skills, and 12 criteria of skills. The "Cello Rubric" developed by Birel and Albuz (2014) consists of 3 main dimensions, i.e., physical skills, technical skills, and musical skills, and 18 items. Developed by Öztürk and Güdek (2016), the "Cello Graded Scoring Key" consists of 3 dimensions, namely physical skills, musical skills, and presentation skills, and 10 target skills. In this case, it can be concluded that the measurement tool developed within the scope of the present study is consistent with the existing studies in the literature.

5. Recommendations

The tool for measuring cello performance developed by the researcher by testing its validity and reliability can be used in the semester final exams of the cello lesson in the institutions in which music teachers are trained in Turkey. The cello performance measurement tool developed can be used in these institutions after examination of the individual instrument cello lesson curriculum in Fine Arts high schools and making the necessary arrangements. Based on the cello performance measurement tool, a performance measurement tool can be developed for other string instruments such as the violin, viola, and contrabass.

References

Abeles, H. F. (1973). Development and validation of a clarinet performance adjudication scale, *The National Association for Music Education*, 21(3), 246-255. https://doi.org/10.2307/3345094.

Bilgen, Ö. B., & Doğan, N. (2017). Puanlayıcılar arası güvenirlik belirleme tekniklerinin karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 8(1), 63-78. https://doi.org/10.21031/epod.294847

Birel A.S., & Albuz A. (2014). Viyolonsel öğretiminde performansı değerlendirmeye yönelik hazırlanan dereceli puanlama anahtarının (rubrik) sınanması ve değerlendirilmesi. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 18(3), 281-207. https://dergipark.org.tr/tr/pub/ataunisosbil/issue/2837/38704

Büyüköztürk, Ş. (2010). *Sosyal Bilimler için Veri Analizi El Kitabı.* Pegem Akademi.

Ciorba, C. R., & Smith, N. Y. (2009). Measurement of instrumental and vocal undergraduate performance juries using a multidimensional assessment rubric. *Journal of Research in Music Education*, 57(1), 5-15. https://doi.org/10.1177/0022429409333405

Dalkıran, E. (2006). *Keman eğitiminde performansın ölçülmesi* [Unpublished doctoral dissertation]. Institute of Educational Sciences, Gazi University.

Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 33(2), 363-374. https://doi.org/10.2307/2529786

Latimer M. E., Bergee M. J., & Cohen M. (2015) Reliability and perceived pedagogical utility of a weighted music performance assessment rubric. *Journal of Research in Music Education*, 58(2), 168–183. https://doi.org/10.1177/0022429410369836

Özdemir, G. (2012). *Müziksel okuma (solfej) performans testi tasarımı* [Unpublished doctoral dissertation]. Institute of Educational Sciences, Burdur Mehmet Akif Ersoy University

Öztürk, D., & Güdek, B. (2016). Viyolonsel performans değerlendirmesine yönelik dereceli puanlama anahtarının (rubrik) geliştirilmesi*. Afyon Kocatepe Üniversitesi Akademik Müzik Araştırmaları Dergisi, 2(3), 31-50.* https://doi.org/10.5578/amrj.10447

Russell, B. E. (2010). *The empirical testing of musical performance assessment paradigm* [Unpublished doctoral dissertation]. University of Miami.

Schleuter, S. L. (1996). *A Sound Approach to Teaching Instrumentalist***.** Schirmer Books.

Uçan, A. (2005). *Müzik Eğitimi, Temel Kavramlar, İlkeler ve Yaklaşımlar ve Türkiye'deki Durum*. Evrensel Müzikevi.

Wesolowski, B. C. (2012). Understanding and developing rubrics for music performance assessment. *Music Educators Journal*, 98(3), 36-42, https://doi.org/10.1177/0027432111432524