CREATION OF A TEST BATTERY FOR THE EVALUATION OF RHYTHMIC FEELINGS IN UNIVERSITY STUDENTS IN THE FIELD OF PHYSICAL EDUCATION AND SPORT

Alena Kašparová, Kateřina Doležalová & Viléma Novotná

Faculty of Physical Education and Sport, Charles University, Prague, Czech

Original article, DOI:10.52165/sgj.13.2.231-241

Abstract
Optimal movement rhythmisation is considered one of the basic prerequisites for improvements in the quality of movement performance using a particular technique. Well-developed rhythm-movement patterns play a role in successful learning of various physical activities as well as in athletic performance. University students – future PE and sports teachers – should improve their rhythmic feel skills during their studies so that they can use them later in their work and develop them in their future students. This requires the creation of a test battery for the evaluation of rhythmic feel skills through a series of music tests. This paper presents the results of tests taken by 121 university students at UK FTVS in Prague, the Czech Republic, and AWFIS in Gdańsk, Poland. The test battery focused on three types of music-motor skills: perception skills and activities (items 1-18), reproduction skills and activities (items 19-27) and production skills and activities (item 28). The data were statistically processed using the classical test theory (factor analysis) and the item response theory (two-parameter model). Statistical methods also included reliability calculation and test validity. The expected rejection of the proposed hypothesis was confirmed both for the classical test theory and for the item response theory. The only exception was model 4 where, however, fit indices (especially TLI = 0.537) pointed more at a lack of evidence for hypothesis rejection than a perfect conformity of the model and data. The intention was to create and test models with the best data compliance. The best data compliance was found in models no. 1 and 5. Model 1 [CFI = 0.927, TLI=0.916, SRMR = 0.09, RMSEA (5 %) = 0.03, RMSEA (95 %) = 0.059] had a structure that corresponded to the proposed test battery and showed a relatively good compliance with data although IRT identified several problematic items. Model 5 [CFI = 0.956, TLI=0,942, SRMR = 0.073, RMSEA (5 %) = 0.03, RMSEA (95 %) = 0.111] was unidimensional (reproduction factor feeding items 19 through 27) and its fit indices showed better compliance of model and data. An optimised test battery should be developed based on these models followed by another validation of the test battery using statistical analyses.

Keywords: *musical-motor skills, test battery, rhythm, factor analysis, item response theory.*

INTRODUCTION

The concept of rhythmic feelings refers to both rhythm and feelings, things that we encounter in many areas of our lives. They play a key role in correcting both the physical and psychological development including, for example, breathing, verbal and memory skills. They are also the cornerstone of any physical activity technique, movement or athletic performance. This is especially true for gymnastics where movements and skills are learned in rhythmic structures with or without an apparatus and require the mastery of movement with optimal rhythm or, in other words, the execution of movement with the best time and space.

Gymnastic skills are also used in many other sports. A wide range of gymnastic coordination and fitness exercises are used in the training of young athletes in many sports as well as in the technical training for other sports that require a high degree of coordination (Strešková, 2005). Gymnastic exercises are often performed to music. Rhythmic gymnastics has the most sophisticated form of music-motor education. It is the basis of training for gymnastics as well as rhythmic gymnastics (Mihule & Šťastná, 1993).

As part of their study, future teachers and coaches learn how to develop and assess rhythmic aptitudes. In their future work, they should be able to apply their skills when training their students and athletes in order to improve their motor skills and athletic performance (Novotná et al., 2012).

Měkota and Novosad (2005) define rhythm as the dynamic and temporal division of movement. They describe the acoustic and visual aspects of rhythmic skills and divide these skills into rhythmic perception and rhythmic realisation. High-level rhythmic skills are believed to help with learning processes.

Váňová together with Skopal and Sedlák (2007, 2013) and Holas (1985) studied musicality in children. Their study inspired us to explore and develop tests for the evaluation of rhythmic feeling abilities with future physical education teachers and sports coaches, taking the definition of "rhythmic feeling" as our starting point.

Rhythmic feeling is considered a psychological category. In other words, it is one of the musical abilities, or internal structures, that is not directly manifested and as such cannot be directly measured. It is manifested externally through activities and perception, reproduction and production skills that, in turn can be measured. These activities and skills must be tested separately, and tests must eliminate their mutual interferences (Sedlák & Váňová, 2013).

Perception skills and activities are based on the tested subject's perception and internal processing and the evaluation of a situation. Reproduction skills and activities are based on the repetition of a reproduced (demonstrated) task. Production skills and activities are based on the production of rhythm or movement based on a specific task with a strong emphasis on creativity.

Music is an integral part of rhythm issues. Music creates ergogenic effects that improve work performance beyond expected levels of endurance, energy, strength, and productivity (Edworthy J. and H. Waring, 2006).

The presented research partially follows up on the results of a dissertation thesis of Brtníková (2008). Brtníková (2008) created a test battery for the diagnosis of musicality and motor skills in secondary school female students as part of their physical education classes. Her work is based on the previous research on music and movement practice by music experts Kos (1975), Mihule and Appelt (1963), and also on musicality tests developed by music experts Bentley (1966), Seashore (1915, 1936) and others. Similar to Brtníková (2008), Shmulevich and Povel (2000), Moseley (2004), Grühn, Galley and Kluth (2006) and others also used the tests of music experts in their original or modified version in music and movement practice.

More recent research in this area tends to focus primarily on the evaluation of already acquired specific motor skills, rather than on the actual level of rhythmic feeling itself.

The aim of this study is to present a design of a new test battery for assessing rhythmic feelings and relevant statistical analysis to assess whether the proposed test items can be part of a test battery and whether this test battery can be used in practice and can form the basis for possible future standardization. Tests that assess the rhythmic feelings of university students are still lacking in sports and physical education.

METHODS

The newly created test battery is based on findings reported in Czech and international literature and in collaboration with experts on sports and music. From Holas (1985), we took and applied the test structuring into several subtests. In the first section "Perception skills and activities", a new test content was created while keeping the example and the number of tests. The test structure and design builds on the work in music psychology and diagnostics by Váňová, Skopal and Sedlák (2007, 2013), which allowed for the interpretation of rhythmic feelings and its components.

Test description:

1) Perception skills and activities:

Each test section contains 1 sample and 6 test tasks.

A) Melodic memory – 3x listening to a familiar melody. The subject is asked whether or not there are any differences in the third recording.

B) Tonal feel – 2x listening to a tone sequence. The subject is asked whether another tone, played separately occurred in the recording.

C) Rhythm memory – each task contains three rhythmic sections played on percussion instruments. The subject is asked whether or not there are any differences in the third recording.

These three test partitions (A, B, C) are processed on a test CD. The subjects are provided with recordings including the sound marks and instructions. The testing takes place in a group setting. In addition to the rhythmic tasks, the tests also include tasks focused on other musical abilities. We were already aware of this inaccuracy when the test battery was created, so it is clear that the statistical analysis will reveal this fact and allow the appropriate changes to be proposed.

2) Reproduction skills and activities:

Each test section contains 3 test tasks.

A) Static rhythmisation (no movement) – three rhythmic variations reproduced by clapping.

B) Dynamic rhythmisation (movement in space) – subjects are asked to repeat a rhythmic movement sequence demonstrated by the examiner to a metronome beat of 114MM.

C) Rhythmisation with added elements (music and apparatus) – Subjects are asked to repeat a short rhythmic movement etude using wooden sticks.

The testing is done individually. Before each test, subjects can try out the etude and demonstrate the etude with verbal rhythmisation.

3) Production skills and activities:

A) Production with music and verbal rhythmisation

Subjects are asked to create a movement etude of eight 4/4 bars including a jump of his/her choice. The test section "Production skills and activities" contains only one item for time reasons. Therefore, this item is worked from the point of view of the regression analysis in relation to the previous two test sections.

The point scores in the individual tests of the test battery are recorded on paper sheets and then transcribed into a computer. 1 point = the subject performed the rhythmic test flawlessly, 0 points = the subject failed to perform the rhythmic task or performed it with mistakes. The skill diagnostic does not focus on the quality of the movement but purely on the rhythmic component of movement. The aim of the proposed diagnosis and evaluation of the rhythmic feelings is to find out which activities pose the biggest problems to students and where the biggest gaps are in their theoretical and practical education. The findings may indicate a lack of stimuli in the teaching of gymnastics and music-movement education. The results of the research should help steer teaching towards the learning of skills and activities that promote and develop rhythmic feelings.

Hypotheses

H1) The proposed test battery includes three dominant factors – perception, reproduction and production.

If H1) is disproved, we expect:

H2) Items 1 through 18 explained through the perception factor.

H3) Items 19 through 27 explained through the reproduction factor.

The testing sample consisted of university students, both female and male, over the age of 19, studying Physical Education and Sports. The test included 121 students – 48 women and 73 men. The students were recruited at the Faculty of Physical Education and Sport, Charles University (UK FTVS; n = 76) and the Akademia Wychowania Fizycnego i Sportu, Gdańsk (AWFIS; n = 45). The tests were translated into Polish for the Polish subjects.

Testing took place during 4 to 5 lessons (1 lesson took 75-90 minutes) in the gymnastics gym and the surrounding area. During the testing, in addition to the examiner, another teacher was present and delivered the teaching.

The research tested the level of rhythmic feelings in 121 subjects using 28 items in a newly created test battery. Rhythmic feelings were not demonstrated directly but through the perception (items 1-18), reproduction (items 19-27) and production (item 28) skills and activities. The statistical analysis of the results was performed using the R software.

At first an exploratory factor analysis (EFA) was performed, one of the methods of classical test theory (CTT). The goal of EFA was to find the true number of existing factors and assesses the factor load on test items, thus getting an insight into the structure of the proposed test in order to assign different factors to the relevant groups of test items. Further, a confirmatory factor analysis (CFA) was performed, which is, again a part of the CTT. Based on the EFA results, a model was created that assigned factors to different item groups according to the test structure. The conformity of the models with the measured data was determined on the basis of the so-called fit indexes. Specifically, in this case, a model was created with 3 factors - perception (items 1 to 18), reproduction (items 19 to 27) and production (item 28). The production factor consisted of a single item which was included for the purposes of a regressive analysis. Classical test theory considers the test battery as a whole, which means that if only some items or groups of items from the test battery were analysed, the results would not be valid. This is because the CTT considers the items to be interdependent and any item that would be excluded from the test would remove essential test information. This consideration is particularly satisfactory for those tests that have already undergone initial optimization and are used in a stable form. A completely new test was created for this research, which is why the item response theory (IRT) was applied to a portion of the statistical processing in order to allow individual evaluation of each item and to treat every item as a separate test tool (Urbánek & Šimeček, 2001).

RESULTS

Exploratory factor analysis

In the R statistical software, an exploratory factor analysis was performed using the *psych library* and the measured data, the number of extracted factors and the rotation type as input for the specific commands. The rotation types used in this analysis are the "varimax" argument, which is part of the orthogonal rotation1 type, and the "oblimin" argument, which is part of the oblique rotation2 type. Extracted factors were counted for values 3 and 4 using the minimal residue extraction method.

Three factor extraction. The application of the three extracted factors did not yield a good compliance between the model and data, as documented by the TLI (Tucker-Lewis Index) with a value of 0.863, although the (Root Mean Square Error of Approximation) index was rather favourable with a value of 0.032. More information about the factors was provided by a percentage value of spread, explained by the relevant factor. The biggest spread value was detected for factor 1 (14%), followed by factor 2 (6%) and factor 3 (4%). The total value of the spread explained by the three extracted factors was 24%, which is a very small value.

Four factor extraction. The results yielded by four factor extraction were more favourable in terms of the fit indexes (RMSEA = 0.025, TLI = 0.928), which was to be expected because the additional factor contributed to the explanation of a larger part of the total spread (a total of 28%). A factor load analysis also showed a trend similar to the three-factor extraction in that it was observed that factor 1 fed primarily items 19-28 (with some exceptions). Other load factors did not feed items in a way that would indicate a clear trend in line with the test design.

Another output included in the EFA was a test of the internal consistency of the test battery using the Cronbach alpha numeric parameter. This value was calculated to be 0.8, which is generally considered as acceptable (the threshold typically being 0.7) although some authors (Lance, 2006) accept this test as reliable only at values higher than 0.9. It should be mentioned that the reliability is calculated for the entire test battery and is considered constant for the entire interval of subjects' skills.

EFA conclusion. In the light of the above results, it can be stated that in the proposed test battery concept (three factors corresponding to specific items), there was only one factor feeding items 19 – 28. The question, however, is whether it was the reproduction factor or the production factor or a factor combining these two manifestations of rhythmic feelings. More testing with more than one production item would be needed to answer this question. Another finding stemming from the EFA was that at least four factors would need to be extracted in order to ensure sufficient model-data compliance.

Confirmatory factor analysis

The confirmatory factor analysis worked with a specific model structure design. Two models (Model 1 and Model 2) were analysed – both assigned the relevant three factors to all items as per the proposed test battery. In one case (Model 2), we calculated a regression analysis of the production-perception and production-reproduction relationships.

Figures 1 and 2 show the structure of the two models.

The final fit indexes show model compliance with the measured data. It is apparent that the model-data compliance was not acceptable. The only favourable index was the RMSEA index. However, literature (Hu & Bentler, 1999) indicates that a proof of model-data compliance requires more indexes with results under or above the calculated value. In this particular case, CFI and TLI indexes yielded very low values. The chi-square test rejected the hypothesis of the proposed models.

CFA conclusion. Three dominant factors were assumed in this analysis. Based on the fit index values, both models failed to show good compliance with the measured data. A detailed analysis of the co-variation values (standard deviations, spreads) would be possible, but these values were not interpreted due to the fit index results.

Item response theory

As part of the item response theory (IRT), the relationship between the latent variable (θ – theta) and the item response was studied. In this case, the latent variable was the factor and the item response characterised by the probability distribution above the latent variable. This distribution was defined by three parameters – the difficulty parameter, the discrimination parameter and the guessing parameter.

Factor analysis using IRT

In this section, a factor analysis was performed in the R software using the *mirt library*. Several models were designed and tested. Model 1 had the same structure as model 1 in the confirmatory factor analysis section and the factor correlation was permitted. Model 2 was based on model 1 but the perception factor included only items 13 to 18 because the first twelve items focused on tonal feelings and melodic memory. Model 3 assumed that the first 18 items were defined by the same factor and similarly Model 4 considered items 13 through 18. Model 5 included items 19 through 27 in a single factor using a two-parameter model.

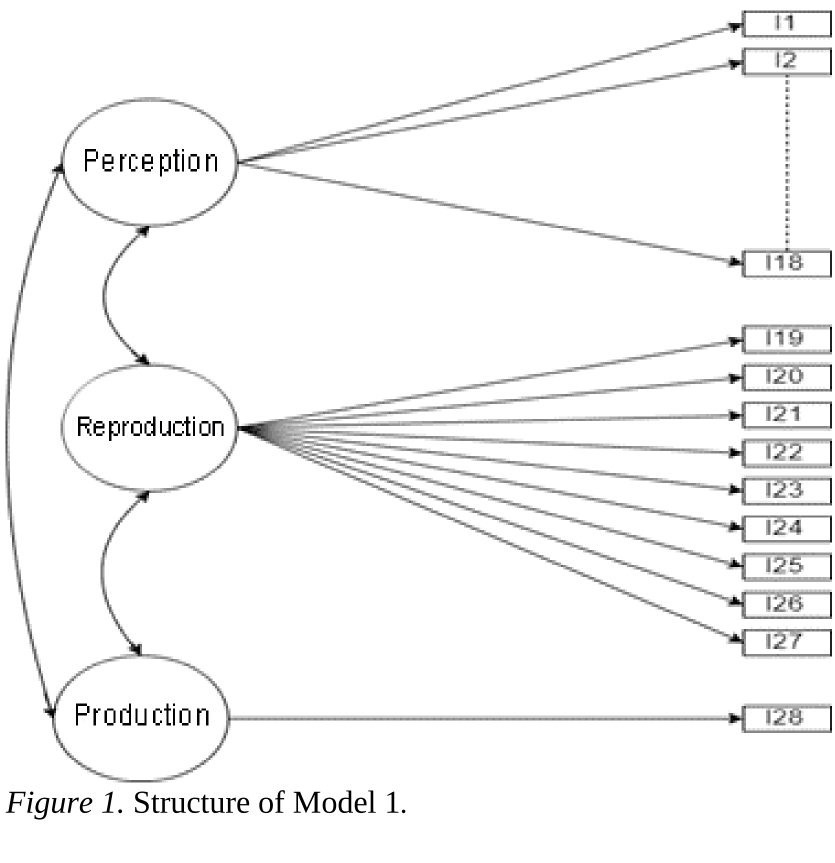

Figure 1. Structure of Model 1.

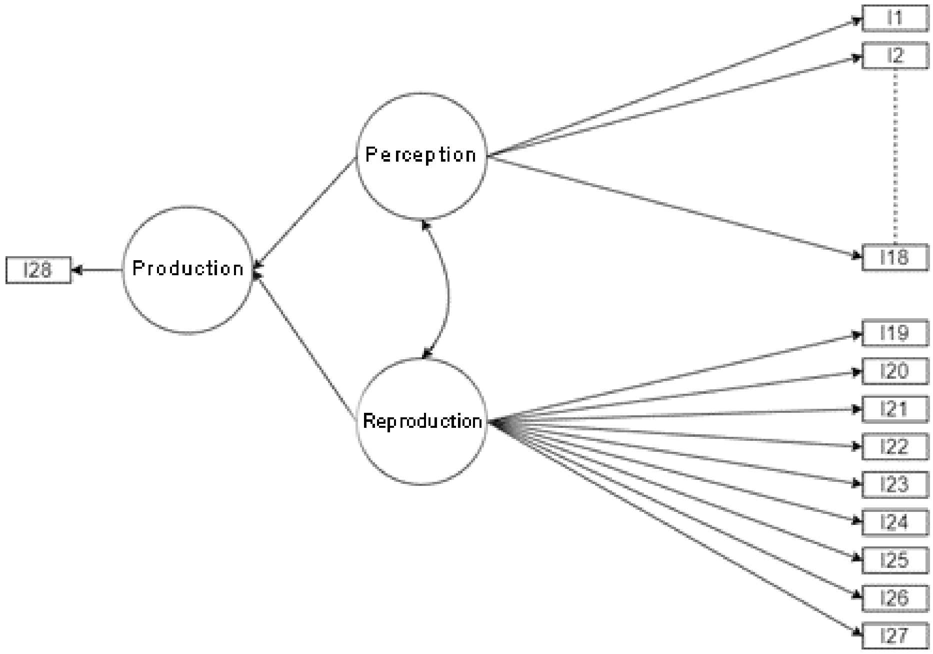

Figure 2. Structure of Model 2.

Table 1
Final fit indexes of the proposed models 1 to 5.

	CFI	TLI	RMSEA	RMSEA (5 %)	RMSEA (95 %)	SRMR	p (χ)
Cutoff	> 0.9	> 0.95	< 0.06	< 0.05	< 0.08	< 0.08	> 0.05
Model 1	0.927	0.916	0.046	0.03	0.059	0.09	0.001
Model 2	0.926	0.907	0.067	0.044	0.087	0.088	0
Model 3	0.866	0.825	0.045	0.012	0.067	0.089	0.036
Model 4	0.907	0.537	0.075	0	0.18	0.11	0.168
Model 5	0.956	0.942	0.074	0.03	0.111	0.073	0.016

Table 2
Correlation matrices of model 1.

	F1	F2	F3
F1	1		
F2	0.795	1	
F3	0.696	0.974	1

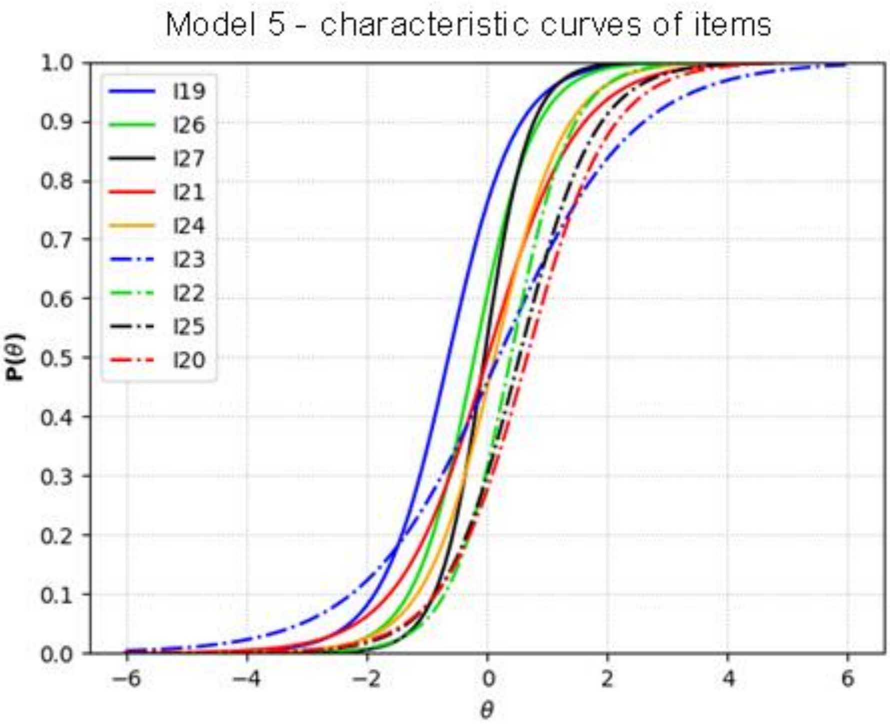

Figure 3. Characteristic curves of items 19 to 27 of model 5.

The above table shows that the worst fit indexes were found for Model 4 and Model 3 where TLI values reach especially low values. Rather surprisingly, the chi-square test forced a rejection of all models with the exception of Model 4, which showed favourable results despite the fact that TLI, which was based directly on the chi-square test, gave very unfavourable results. For this reason, Model 4 is no longer considered. Model 2, once again, showed unfavourable results with only two of the six indexes being acceptable. The most favourable results were yielded by Model 1 and Model 5 with relatively satisfactory values for the initial test battery analysis.

Model 1

It turned out that for items 1, 8, 15 and 18 focused on perceptual skills and activities, the difficulty parameter is up to a hundred times higher than for all other items (I1 = 34.3; I8 = 78.17; I15 = 53.39; I18 = 32.53). This means that for a proband, in order to have a 50% chance of answering an item correctly, it must have a latency value several times higher than for other items. On the latency scale from -6 to 6, which includes all other items, the characteristic curves of items 1, 8, 15, and 18 then appear as horizontal lines — that is, they do not discriminate at all on this scale. The empirical reliability of the perception factor was 0.77, the reproduction factor 0.85 and the production factor 0.85. From this it can be concluded that the largest measurement error occurs in the perception factor.

The correlation coefficients between factors are also important variables (Table 2). Significant values of correlation with positive signs can be observed, which the design concept implicitly assumed. The most interesting is the correlation coefficient between factor 2 (reproductive ability) and factor 3 (production ability), which is based on 0.974 and suggests that these two factors correlate very significantly. The question is how this correlation coefficient would change if factor 3 contained more than one item.

Model 5

Regarding the values of the parameters of individual items, all values of the discrimination parameter came out greater than 1, i.e. the characteristic curves of the items (Figure 3) have the trend direction. It can be seen from the figure that item 23 has the lowest sensitivity, the slope of which is the least steep, while item 27 distinguishes the most. Another monitored parameter is the difficulty parameter, which distinguishes between difficult items (small difficulty parameter) and easy items (large difficulty parameter). In the legend, it is then possible to follow from top to bottom the most difficult items (items 19 and 26) to the simplest (items 25 and 20). The value of empirical reliability for model 5 was 0.785.

Conclusion of statistical results

All proposed hypotheses had to be rejected because the chi-square determined perfect model-data compliance. It was clearly impossible to not reject these hypotheses for the newly created test although the aim was to detect a good compliance rather than a perfect one.

Based on the results of the factor analysis using the classic test theory, the model-data compliance was not acceptable. In a factor analysis, part of the item response theory, two models were found with an acceptable data compliance considering that this was only a pilot test. One of these models included three factors and the other model included only one factor, specifically reproduction skills.

DISCUSSION AND CONCLUSIONS

This paper introduced research using a test battery for the evaluation of three aspects of rhythmic feelings in tested subjects. The purpose of the statistical analysis was to verify that the test structure was really a three-factor structure and, if possible, to differentiate each factor. Both the classical test theory and the item response theory led to a rejection of the proposed hypothesis (models) on a 95% level of significance, the only exception being Model 4 in the item response theory. Considering other fit indexes (especially TLI, which is based on the chi-square test), the results may be interpreted as failing to identify sufficient evidence for model rejection. In the current form, therefore, the test cannot be used to assess rhythmic feelings but requires optimisation, followed by more testing and statistical analysis.

Hypothesis rejection had been expected because very few newly designed test batteries have passed the chi-square test. However, this research was focused more on finding good model-data compliance according to different fit indexes, rather than a perfect compliance in a chi-square test.

Based on the results of the statistical analysis, the following recommendations can be proposed for the next testing:

- if possible, increase the number of test subjects by at least twofold (the more, the better),
- reduce the number of items falling into perceptual skills and activities, or, in consultation with a music expert, review the difficulty of existing items (1 to 18),
- in the new test include only those items from the section of perceptual skills and activities that focus on rhythmic memory,
- include more items in the section of production skills and activities and observe a correlation coefficient between the factor of reproductive abilities and production abilities. Only then will it be revealed whether these two factors are really so significantly correlated,
- reduce the number of items in the reproductive skills and activities section, for time reasons. The statistical analysis performed will help you decide which items to exclude.

Based on previous recommendations, the new structure of the test could look like this. The perception skills and activities section could include only 6 items focusing on rhythmic memory. The reproduction skills and activities section could include 4-6 items considering that this part of the test is the most time consuming. The production skills and activities section could include 2-3 items. The new test battery would thus include a total of 12-15 items. This would significantly reduce the time needed for conducting the test and would also probably increase the informative value of the test. However, newly gathered results must be assessed by statistical analysis before any decision regarding further optimisation or standardisation of the test battery.

ACKNOWLEDGEMENT

This research was part of the grant project No. 1162218 (Measurement of rhythmical feelings with university students in physical education and sports) confirmed by the Grant Agency of the Charles University and the PROGRES Q19 project of the UK FTVS. The research was approved by the Ethics Committee of the UK FTVS under ref. no. 097/2019.

REFERENCES

Bentley, A. (1966). *Musical ability in children and its measurement*. London: Harrap.

Brtníková, M. (2008). *Modernizace hudebně pohybové výchovy.* (Disertační práce). Praha: UK FTVS.

Edworthy, J. & H. Waring. (2006). The effects of music tempo and loudness level on treadmill exercise, *Ergonomics*, 49 (15), 1597-610. doi: 10.1080/00140130600899104

Grühn, W., Galley, N., & Kluth, Ch. (2006). Do Mental Speed and Musical Abilities Interact? *Annals New York Academy of Sciences 999 (1)*, 485-496. doi.org/10.1196/annals.1284.059

Hu, L. & P. Bentler. (2009). Cut off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: *A Multidisciplinary Journal*. Lawrence Erlbaum Associates, Inc., 6(1), 1-55. doi: 10.1080/10705519909540118

Holas, M. (1985). *Úvod do hudební diagnostiky: určeno pro učitele hudební výchovy, učitele hudební nauky na LŠU, učitele HV na SPgŠ*. Praha: Státní pedagogické nakladatelství.

Kos, B. (1975). Metodika výzkumu rytmických schopností sportovce. *Teorie Praxe tělesné výchovy,* 23 (11).

Kos, B. (1975). Výzkum rytmických schopností sportovce. *Teorie Praxe tělesné výchovy,* 23 (12).

Lance, C., Butts, M. & L. Michels. (2006). The Sources of Four Commonly Reported Cut off Criteria: What Did They Really Say? *Organizational Research Methods. Sage Publications*, 9(2), 202-220. doi: 10.1177/1094428105284919

Měkota, K., & J. Novosad. (2005). *Motorické schopnosti*. Olomouc: Univerzita Palackého.

Mihule, J., & K. Appelt. (1963). *Rytmus a jeho experimentální výzkum.* In Sborník ITVS UK 4, Praha: Univerzita Karlova.

Mihule, J., & D. Šťastná. (1993). *Rytmická gymnastika*. Praha: Karolinum.

Moseley, S. (2004). When “Too Little” Is Not Enough: Lessons Learned and Applied from Teaching One-Credit-Hour Dance Methods Courses for Physical Education Majors. *Research Quarterly for Exercise and Sport, 75 (1.).*

Novotná, V. et al. (2012). *Gymnastika jako tvůrčí akt*. Praha: Karolinum.

Seashore, C. E. (1915). The measurement of Musical talent. *Music quarterly*. 1(1), 129-148. doi: 10.1093/mg/l.1.129

Seashore, C. E. (1936). *Objective Analysis of Musical Performance.* New York.

Sedlák F., & H. Váňová. (2013). *Hudební psychologie pro učitele*. Praha: Karolinum.

Shmulevich, I., & D. J. Povel. (2000). Complexity measures of musical rhythms. In P. Desain & L. Windsor, *Rhythm perception and production:* Lisse, NL: Sweets & Zeitlinger, 239-244.

Urbánek, T., & M. Šimeček. (2001). Teorie odpovědi na položku. *Československá psychologie: Časopis pro psychologickou teorii a praxi*. Praha: Academia, 45(5), 428-440. ISSN 0009-062X

Váňová, H., & J. Skopal. (2007). *Metodologie a logika výzkumu v hudební pedagogice*. Praha: Karolinum.

Corresponding author:

Mgr. Alena Kašparová
Department of Gymnastics
Faculty of Physical Education and Sport
Charles University
José Martího 269/31
162 52 – Praha 6 – Veleslavín
Česká republika
e-mail: akasparova@ftvs.cuni.cz
tel.: 774 853 192

Article received: 19.8.2020
Article accepted: 1.12.2020