MODIFICATION OF SCORING SCHEMES USING DECOMPOSITION PROCEDURES ON STATISTICAL DATA

Monika Hanáková,¹ Aba Teleki,² Boris Lacsny³

Abstract: This paper presents a method of modifying original scores to obtain independent random variables. It includes an analysis of the consequences of using such a method. The paper also describes the mathematical background of the method in detail and discusses the possible use of the method in identifying student or participant assessments that are over- or underrated. The method distinguishes performances of students and assesses their written solutions using a scoring scheme. In this study, it is used to analyze the competence of participants in the Physics Olympiad competition. Scoring schemes that are appropriately set by an author for a physics problem present the participant scores as independent random variables. The assessment solutions are analyzed using analytical tools (such as covariant matrix) for the dependence of random variables. The evaluators of the participants' solutions were highly qualified professionals. Nevertheless, the study found statistical evidence of minor distortion in the evaluations, though this was found to only marginally affected the ranking of participants.

UDC Classification: 37.01/.02; **DOI:** <http://dx.doi.org/10.12955/cbup.v5.997>

Keywords: Physics Olympiad, Assessment, Scoring, Independent Random Variables, Ranking

Introduction

Assessment of students plays an important role in physics education as it provides an essential feedback for all its participants, including instructors and students, as well as participants in various physics competitions. One of the most-often used assessment methods is evaluating written solutions of physics problems and tasks. According to Gaigher (2007), problem-solving is considered a reliable way to demonstrate conceptual understanding in physics for purposes of evaluation.

Physics Olympiad (PhO) is a worldwide competition where PhO participants solve physics problems, and their performance is evaluated according to their written solutions. Its history in Slovakia started in the school year 1958–59 and has been an important part of Slovakia's education system since then. The main objectives of PhO in Slovakia, as defined by the Committee of PhO, are to develop the problem-solving and experimental skills of primary and secondary school students talented in physics, create competitive environments for them, and encourage them to study physics or a related science (Slovak Committee of Physics Olympiad, 2010).

An assessment of each PhO participant is crucial as it influences the participant's results. Currently, the PhO participants are rated by points (maximum of 10 for a physics problem), which they achieve according to the scoring scheme proposed by the authors of the physics problems. Each physics problem is divided into tasks that are scored separately. For example, a score for the solution of the j -th task is defined as a random variable X^j and scored as either '0' for an incorrect or '1' for the correct answer. In general case, $X^j \in [0,1]$ – for example, one can assign the value 0.5 to any partially correct solution which is nor absolutely incorrect nor absolutely correct. This scaling is used rather than specific scores because of its universality. This paper describes the basics of a statistical method to quantify participant assessments using data and results of solutions provided by the participants of the competition, PhO.

Data and Methodology

The first step was to analyze a case using a proposed modification to the scoring of a single physics problem. That is, we analyzed the effect on scores and rankings of a particular PhO participant in regard to solving one specific physics problem. The modified overall scores gained for solving all four physics problems in each category were determined as the sum of modified scores gained for particular physics problems.

We defined scores of X^j for solving the j -th task (upper index) from a set of 's' tasks as random (column) vector variable, where the k -th element X_k^j is a random variable for solving the j -th task by

¹ Faculty of Natural Sciences, Constantine the Philosopher University in Nitra, Slovakia, monika.hanakova@ukf.sk

² Faculty of Natural Sciences, Constantine the Philosopher University in Nitra, Slovakia, ateleki@ukf.sk

³ Faculty of Natural Sciences, Constantine the Philosopher University in Nitra, Slovakia, blacsny@ukf.sk

the k -th participant (lower index). We assumed that the random variables, X_k^j , were defined on the same probability space $(\Omega, \mathcal{A}, P)$, where Ω denoted the set of all possible outcomes (containing all possible solutions of the task, including all incorrect, incomplete, and correct solutions). Therefore, it was accepted that Ω was the same for all tasks and all participants. A set of subsets (events) of Ω was defined as, $\mathcal{A}$, i.e., an event was a particular solution of the task. The measure of probability, P , mapping any event into the interval $[0,1]$, i.e., $P(\mathcal{F}) \in [0,1]$ for any $\mathcal{F} \in \mathcal{A}$. In practice, the participants were a limited ensemble with their solutions assessed by scores, where X_k^j was $P(\mathcal{F}_k^j)$, where $\mathcal{F}_k^j$ was the event, $P(\mathcal{F}_k^j)$, representing the solution of j -th task by the k -th participant, and P was the probability measure by the author of the scoring scheme and evaluators. The probability of the measure was distorted by inadequacies of the authors scoring scheme and by unsatisfactory decisions of evaluators. In the case of the PhO, this influence was relatively small due to the professionalism of authors and evaluators. Mathematical statistical tools identified strong evidence of the limitations in the author's scoring scheme, and in some cases, probable anomalies in the assessment of the evaluators (Hanáková, & Teleki, 2017).

The basic concept was that the random vector, $\mathbf{X}^j$, as an independent variable provided objective scoring for the PhO participants with the same knowledge or skills. However, Pearson's correlation coefficients (Evans, 1996; Markechová, Stehlíková, & Tírpáková, 2011; Spiegel, 1998) of random vectors $\mathbf{X}^j$ showed that many were dependent variables. Following this, a covariant matrix, Σ , was created to analyze variables $\mathbf{X}^j$ and $\mathbf{X}^k$. This method was selected in place of the Pearson's correlation coefficients to avoid nonlinear equations in the optimization procedure.

In the optimization procedure, the scoring scheme was modified to obtain independent random variables, ξ^k .

The modified random variables corresponded to the modified tasks. However, examining these modified tasks in their explicit forms was a highly complex problem and beyond the purpose of this article.

The covariance matrix, Σ , was defined by matrix elements as follows:

$$\text{cov}(\mathbf{X}^j, \mathbf{X}^k) = \Sigma^{jk}(\mathbf{X}) = E[(\mathbf{X}^j - \mu^j)(\mathbf{X}^k - \mu^k)], \quad (1)$$

where,

$$\begin{aligned} E[\mathbf{X}^j] &\text{ denoted the expected value of } \mathbf{X}^j, \text{ and} \\ \mu^j &= E[\mathbf{X}^j]. \end{aligned} \quad (2)$$

We estimated $E[\mathbf{X}^j]$ by the average of $\frac{1}{N} \sum_{i=1}^N X_i^j$ and therefore, the covariance matrix elements were estimated using Equation 3.

$$\Sigma^{jk} = \frac{1}{N} \sum_{i=1}^N X_i^j X_i^k - \frac{1}{N^2} \sum_{i=1}^N X_i^j \sum_{r=1}^N X_r^k, \quad (3)$$

where,

$$N = \text{number of assessed pupils and students.}$$

The covariance matrix Σ , is a real and symmetric matrix, i.e., it is a Hermitian matrix (Horn and Johnson, 1985) and therefore Σ is diagonalizable by an orthogonal transformation $\mathbf{A}$, $(\mathbf{A})_{kj} = a_{kj}$ (Equation 4).

$$\Sigma_D = \mathbf{A} \Sigma \mathbf{A}^T \quad (4)$$

Covariant matrix diagonalized: $\Sigma_D^{kk} = D^k$, $\Sigma_D^{jk} = 0$ for $j \neq k$.

The transformation, $\mathbf{A}$, was not unique in the manner described below and defined the above mentioned new probability vectors, ξ^k (Equation 5a).

$$\xi^k = \sum_{j=1}^N a_{kj} \mathbf{X}^j \quad (5a)$$

The inverse transformation $\mathbf{B} = \mathbf{A}^{-1} = \mathbf{A}^T$, $(\mathbf{B})_{jk} = b_{jk}$ was

$$\mathbf{X}^j = \sum_{k=1}^N b_{jk} \xi^k \quad (5b)$$

where,

$$b_{jk} = a_{kj}, \quad (6)$$

$$\sum_{j=1}^N a_{jk} b_{kl} = \sum_{j=1}^N a_{jk} a_{lk} = \sum_{j=1}^N a_{kj} a_{kl} = \delta_{jl}, \text{ and} \quad (7)$$

$$\delta_{jl} = \text{Kronecker-delta} (\delta_{jj} = 1 \text{ and } \delta_{jl} = 0 \text{ for } j \neq 0).$$

It was convenient to save a different notation for the matrix $\mathbf{A}$ and its inverse $\mathbf{B}$ and show, that

$$\text{cov}(\xi^j, \xi^k) \equiv \Sigma^{jk}(\xi) = \sum_{l,r=1}^N a_{jl} a_{kr} \text{cov}(\mathbf{X}^l, \mathbf{X}^r) = \sum_{l,r=1}^N a_{jl} a_{kr} \Sigma^{lr}(\mathbf{X}) = \Sigma_D^{jk}. \quad (8)$$

The diagonalization procedure defined the new random vectors, ξ^j , with a chance that they were statistically independent ($\Sigma_D^{jk} = 0$ for all $j \neq k$). It was assumed that they were independent random vectors.

The transformation matrix $\mathbf{A}$ was obtained by the Schur decomposition procedure (Horn and Johnson, 1985), which solved the eigenvalue equation:

$$(\Sigma - \lambda_j \mathbf{I}) \mathbf{a}_j = 0, \quad j = 1, \dots, s \quad (9)$$

where $\mathbf{a}_j$ was a normalized eigenvector ($|\mathbf{a}_j| = 1$) associated with eigenvalue λ_j . The set of eigenvectors $\mathbf{a}_j$ formed an orthonormal basis in the s -dimensional vector space. Multiplying Equation 9 by $\mathbf{a}_k$ provided the matrix elements of the diagonal covariance matrix Σ_D (Equation 10).

$$\mathbf{a}_k \cdot \Sigma \mathbf{a}_j - \lambda_j \delta_{kj} = 0, \quad \lambda_j = \Sigma_D^{jj} \quad (10)$$

The following defined the transformation matrix $\mathbf{A}$:

$$(\mathbf{A}_{jk}) = a_{jk} = (\mathbf{a}_k)_j. \quad (11)$$

The normalized eigenvectors are given up to a multiplicative factor ϵ_j , ($|\epsilon_j| = 1$). Also, the order of the eigenvectors is arbitrary.

In the first step, the eigenvector $\mathbf{a}_{j_1}$ and the k_1 -th coordinate, $a_{k_1 j_1}$, is defined with the property:

$$(\mathbf{a}_{j_1})_{k_1} = |a_{k_1 j_1}| = \max_{j,k \in K_1} |a_{jk}| > 0, \text{ where } J = \{1, 2, \dots, s\}. \quad (12)$$

Where $a_{k_1 j_1} < 0$, the transform $\mathbf{a}_{j_1} \rightarrow -\mathbf{a}_{j_1}$ is achieved to obtain $a_{k_1 j_1} > 0$.

In the second step, the new sets of indices were defined (Equation 13a).

$$J_2 = J \setminus \{j_1\} \quad \text{and} \quad K_2 = J \setminus \{k_1\} \quad (13a)$$

Also, the eigenvector, $\mathbf{a}_{j_2}$, and the k_2 -th coordinate, $a_{k_2 j_2}$, were defined with the property:

$$(\mathbf{a}_{j_2})_{k_2} = |a_{k_2 j_2}| = \max_{j \in J_2, k \in K_2} |a_{jk}| > 0. \quad (13b)$$

Where $a_{k_2 j_2} < 0$, the transform $\mathbf{a}_{j_2} \rightarrow -\mathbf{a}_{j_2}$ was achieved to obtain $a_{k_2 j_2} > 0$.

This procedure was repeated to obtain in the r -th ($r > 3$) step for a new set of indices (Equation 14a).

$$J_r = J_{r-1} \setminus \{j_{r-1}\} \quad \text{and} \quad K_r = K_{r-1} \setminus \{k_{r-1}\} \quad (14a)$$

Also, the eigenvector, $\mathbf{a}_{j_r}$, and the k_r -th coordinate, $a_{k_r j_r}$, were defined with the property:

$$(\mathbf{a}_{j_r})_{k_r} = |a_{k_r j_r}| = \max_{j \in J_r, k \in K_r} |a_{jk}| > 0. \quad (14b)$$

Where $a_{k_r j_r} < 0$, the transform $\mathbf{a}_{j_r} \rightarrow -\mathbf{a}_{j_r}$ was achieved to obtain $a_{k_r j_r} > 0$.

Next, the eigenvectors were permuted to obtain Equation 15.

$$j_1 = 1, j_2 = 2, \dots, j_s = s. \quad (15)$$

At this point the transformation matrix $\mathbf{A}$ and its inverse $\mathbf{B}$ was unambiguous.

The new score ξ_k^j of the k -th participant gained in the j -th task was calculated (Equation 16).

$$\xi_k^j = \sum_{r \in J} a_{jr} X_k^r \quad (16)$$

The full score, Ξ_k , for the k -th participant was formed from Equation 17.

$$\Xi_k = \sum_{j \in J} \xi_k^j \quad (17)$$

Notably, the probability vectors, ξ^j , were scalable in the following meaning: the new probability vectors, as defined by Equation 18, had also diagonal covariant matrix, $\text{cov}(\zeta^j, \zeta^k)$ and where the ξ^j -s represented independent probability variables, ζ^j -s, represented independent probability variables too.

$$\zeta^j = c_j \xi^j \quad (18)$$

This scaling property was critical in the forming the new scoring schemes. A given scoring scheme can motivate participants (or students) by lowering the maximal value of the scoring in tasks, which were excessively difficult for the participant. Another scoring scheme could emphasize the participants with the highest competency.

The above-described tool was tested successfully on participants of PhO in 2016, and the rescaling of ξ^k was defined as follows:

$$\zeta^k = \frac{10}{S(\xi)} \xi^k \quad (19)$$

In this equation, $10/S(\xi)$ was a scaling factor with 10 the maximum original score for complete and accurate solving of a physics problem. The appropriate choice of the value, $S(\xi)$, guaranteed the same maximum of 10 for the modified assessment score, calculated by using ξ -s, (i.e., $\sum_k \zeta_{\max}^k = 10$).

The influence of modifying the overall ranking of PhO participants was analyzed by calculating the differences between the original and the modified scores (ΔP) and rankings of the PhO participants (ΔR). Linear correlations between random variables X^j and X^k were identified from the values of the covariance matrix in Equation 1. The described modification procedure was applied to obtain independent random variables (modified scores of PhO participants).

The effects of modification on results of particular PhO participants showed differences in scores (ΔP) and ranking (order) (ΔR). Figure 1 shows the maximum differences in scores corresponding to maximum differences in ranking and the frequency of modified cases with non-zero differences in scores $f_{1P}(\Delta P)$ or ranking $f_{1P}(\Delta R)$ per one participant.

The effects of modification on results of particular PhO participants were analyzed to observe:

- differences in scores ΔP and ranking (order) ΔR ;
- whether the maximum differences in scores correspond to maximum differences in ranking;
- frequency of modified cases with non-zero differences in scores $f_{1P}(\Delta P)$ or ranking $f_{1P}(\Delta R)$ per one participant.

Results and Discussion

Quantitative results of characteristics of the analysis for four physics problems solved in D and E categories are presented in Figure 1 and Figure 2, and in more detail, can be found in Hanáková and Teleki (2017). The PhO participants were ranked according to their original overall scores, with a score of one denoting the best.

Figure 1: Effect of modification on ranking of Physics Olympiad participants after solving one particular physics problem in D category

Source: Authors

Figure 2: Effect of modification on ranking of Physics Olympiad participants after solving one particular physics problem in E category

The maximum difference in D category in scores is $\Delta P = +3$ and the maximum difference in ranking, as shown in Figure 1, is $\Delta R = -8$. The PhO participant's position in ranking changed from 9th to 17th place, where the negative number indicates the result of the PhO participant was worse after modification of the scores. The frequencies characterized above have the values $f_{1P}(\Delta P \geq 0) = 2.1$; $f_{1P}(\Delta R \geq 0) = 0.6$.

Figure 3: Effect of modification on overall scores of the Physics Olympiad participants in D and E categories

The maximum difference in scores in E category is $\Delta P = +3$, and the maximum difference in ranking (Figure 2) is $\Delta R = +20$ (which indicates an improvement for the PhO participant after modification of scores). The frequencies characterized above have the values $f_{1P}(\Delta P \geq 0) = 2.8$; $f_{1P}(\Delta R \geq 0) = 1.7$. Non-zero differences in ranking were observed for each physics problem.

These results show that in observed cases a higher frequency of differences in the scores $f_{1P}(\Delta P \geq 0)$ resulted in a higher frequency of differences in the ranking $f_{1P}(\Delta R \geq 0)$.

The modified overall scores of the PhO participants in D and E categories increased when compared with the original scores, in all cases (Figure 3). We assume that this could be explained by non-zero linear correlations between original scores (random variables).

Conclusion

The performance of the PhO participants was quantified and then compared according to the scoring scheme applied for assessing their written solutions to the physics problems. The main objective of this article was to underline the need for improving the scoring scheme and to present and describe a

statistical tool that could provide a more objective assessment of participant solutions to the physics problems. We found non-zero linear correlations between the probability vectors $\mathbf{X}^j$ and $\mathbf{X}^k$ that were determined using a covariance matrix. This result was considered relevant towards proposing a modification of the scores to provide independent random variables. The study identified certain cases of original scores changing after modification in the range of -0.4 to 3.0 points for individual physics problems. The ranking of the PhO participants also changed after modification in the range of -12 to $+20$ places for particular physics problems. As the modification by way of the proposed tool was focused on the results of PhO participants, the context of the modified random variables remains a matter for future research. This paper describes, in detail, a proposed statistical tool as a basis for developing a suitable method of modifying scores as well as analyzing over- or underrating of performance in solving physics problems. Finally, because of its universality, this tool can be applied not only to the Physics Olympiad but also to other cases where solutions to physics problems are assessed.

Acknowledgements

We would like to express our sincere gratitude to University Grant Agency of Constantine the Philosopher University in Nitra, Slovakia for support of project UGA VII/7/2017 and KEGA 039UKF-4/2017 (Lifelong support of secondary school teachers for great teaching – physics experiments).

References

- Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences*. CA: Brooks/Cole Publishing, Pacific Grove
- Gaigher, E. (2007). Exploring the Development of Conceptual Understanding through Structured Problem-solving in Physics. *International Journal of Science Education*. <http://dx.doi.org/10.1080/09500690600930972>
- Hanáková, M., & Teleki, A. (2017). Modification of scoring scheme in Physics Olympiad and its possible utilization in analysis of assessment. *AIP Conference Proceedings: 20th International DIDFYZ 2016 Conference: From the Roots to Contemporary Education*. DOI: 10.1063/1.4974382
- Horn, R.A. & Johnson, C.R. (2013). *Matrix Analysis*. 2nd ed., Cambridge University Press. ISBN 978-0-521-83940-2. (section 2.3 and 4.1)
- Markechová, D., Stehliková, B., & Tirpáková, A. (2011). *Štatistické metódy a ich aplikácie*. Nitra: Prírodovedec [Statistical Methods and Their Applications.]
- Slovenská komisia fyzikálnej olympiády, (15 March 2010) Organizačný poriadok Fyzikálnej olympiády No. MŠSR-9409/2010-913 [Slovak Committee of Physics Olympiad. Organisation Manual of Physics Olympiad]
- Spiegel, M. R. (1998). *Problems of Probability and Statistics* New York: McGraw-Hill