Measuring the students’ level of knowledge on the basis of other students’ subjective opinions

I U Zulkarnay¹, A I Dashkina² and D A Tarkhov²,³,⁴

¹Bashkir State University, Ufa, 450076, ul.Z.Validi, 32, Russian Federation
²Peter the Great St. Petersburg Polytechnic University, 195251, St.Petersburg, Polytechnicheskaya, 29, Russian Federation
³E-mail: dtarkhov@gmail.com

Abstract: This paper develops a new methodology of measuring the students’ level of knowledge on the basis of other students’ subjective opinions. It is done a research into the correlation between the subjective and objective indicators of the educational process. One of the key issues behind controlling the state of social processes is the fact that it is difficult to measure them objectively. In this case it is much easier to obtain subjective indicators. It implies assessing the values we are interested in by people. If we manage to measure the objective indicators with an acceptable degree of accuracy on the basis of the subjective ones, it will be much easier to study sociodynamic phenomena and processes. The methodology includes building a neural network function. The method allows the teacher to evaluate the progress made by a group of learners in a less time-consuming way.

1. Problem

When it comes to controlling and assessing social processes, one of the main problems is obtaining the objective indices for them. It is much easier to get subjective indices, people assessing the values we are interested in. If we manage to identify the objective indices on the basis of the subjective ones with adequate precision, it will be much easier to study sociodynamic phenomena and processes. Our research focuses on studying the correlation between the objective and subjective indices in the educational process.

There is a large body of literature on the measurement theory and a metrological perspective and on how they are applied in measuring social processes. Thus, Mari, L., & Wilson, M. give comparative discussion of deterministic and probabilistic measurement models on the example of social measurement in terms of the International Vocabulary of Metrology, giving an introduction to Rasch models in a metrological context they analyze measurement validity [1]. Maul, A., Wilson, M., & Torres Irribarra, D. try to formulate a philosophically coherent account of how measurement works in both the physical and human sciences. They argue that “…measurement is concerned with gaining knowledge about the real world, that it is an empirical process, and that it is shaped by the pragmatic concerns of the agents involved” [2]. Pendrill, L. studies how the metrology of human-based observations can be tackled with a measurement system analysis approach, focusing on the side where Man acts as a measurement instrument [3]. He discusses the problem of decision risks and suggests that through the Rasch invariant measure approach it is possible to provide a proper treatment of ordinal data. In this

⁴ To whom any correspondence should be addressed.
case Rash models are similar to neural network ones, which provides a basis for successful application of neural networks in studying sociodynamic phenomena.

Our current research described in this article focuses on studying the correlation between the objective and subjective indices in the educational process.

2. Research Methodology
We measured how efficiently the students learned professionally orientated English depending on how the teaching process was organized. Our pedagogical experiment is based on the application of the hierarchical teaching system and mentorship, which prove to be much more effective than a conventional system of organizing the learning process, especially if the learners in the group have different levels of knowledge. In the hierarchical teaching system the knowledge is transferred not only from a teacher to learners, but also from one student to another, the system can be referred to as a hierarchical one.

The experiment was conducted at St. Petersburg Polytechnic University. It involved three groups of third-year students specializing in economics and management. At the end of the experiment we prepared a final test for the two experimental groups and the control one in order to check their progress. The test included multiple choice lexical tasks based on the limited amount of professional vocabulary studied in the course of the experiment. The students were also supposed to summarize a text based on the same lexical material.

The students in group 2 were with a high level of knowledge, whereas the learners in group 1 were with a low level of foreign-language competence. Mentorship meant that the students from group 2 helped the students from group 1 in their learning process.

The students from group 2 worked in the capacity of mentors in collaboration with the students from group 1. As group 2 outnumbered group 1 (20 and 7 students respectively), there were more mentors than mentees. The students from all the groups did the reading comprehension exercises, summarized the texts and discussed the cases referring to the texts. While doing the assignments, the mentors from group 2 helped the mentees from group 1 and the teacher checked the results of their work in class.

The students in control group 3 worked individually, without collaborating with each other, which means that the students with a high level of knowledge did not help the ones with a low level of foreign-language competence. They also summarized and discussed the texts, and did reading comprehension exercises.

At the beginning of the experiment we ran an input test aimed at checking the students’ knowledge before the series of the experimental classes. At the end of this series of classes we ran a final test in order to assess objectively how the knowledge had changed.

We ran the same final test for the two experimental groups and the control one. Each of the three groups which took part in the experiment was given the multiple choice lexical tasks based on the vocabulary studied in the course of the experiment. The students were also supposed to summarize a text based on the lexical material studied in the course of the experiment.

Apart from the above mentioned testing, the students from group 2 assessed the knowledge of the students from group 1 on a five-point scale. The latter, in their turn, assessed the knowledge of the students from group 2. Besides, all the students’ knowledge was assessed by the teacher.

3. The results of the experiments
The final test enabled us to determine how well they had acquired the vocabulary and how efficiently they summarize professionally-orientated texts. When we evaluated the results of summarizing the text, we considered the following criteria: rendering the ideas, grammar, vocabulary, sentence structure and independent judgment. The maximum score for each criterion was 5. The final test also included 10 multiple-choice sentences (one point per sentence). The maximum score in the assignment with true or false statements was 5.

Apart from that, we asked the students from group 1 to assess the progress of their mentors from group 2 in professionally-orientated English (table 2). The mentors (the students from group 2), in their
turn, had to assess their mentees from group 1 (Table 1). We wanted to discern the correlation between the students’ subjective peer assessment and the objective assessment based on the results of the tests run by the teacher. The results of assessing the knowledge in group 3 we see in Table 3.

Table 1. The results of assessing the knowledge in group 1 (The values are in the form of fractions: above the slash (/) – the results of the final test, below the slash (/) – the assessment by the students from group 2).

| №   | Overall result of the input test | Summarizing | Grammar | Vocabulary | Sentence structure | Opinions and judgments | Filling out the gaps | True or false statements | Overall result: final test / stud.’s assesm. |
|-----|---------------------------------|-------------|---------|------------|-------------------|------------------------|----------------------|--------------------------|-----------------------------------------------|
| 1   | 21                              | 4/5         | 4/3     | 4/3        | 4/2               | 3/4                    | 10                   | 5                        | 34/17                                      |
| 2   | 26                              | 5/5         | 5/4     | 54         | 5/3               | 4/5                    | 10                   | 5                        | 39/21                                      |
| 3   | 22                              | 4/3         | 4/3     | 4/3        | 4/5               | 3/4                    | 9                    | 5                        | 33/18                                      |
| 4   | 22                              | 4/3         | 4/3     | 4/3        | 4/1               | 3/3                    | 8                    | 3                        | 30/13                                      |
| 5   | 23                              | 5/3         | 3/2     | 4/4        | 4/4               | 4/3                    | 10                   | 4                        | 34/16                                      |
| 6   | 19                              | 5/5         | 4/3     | 4/5        | 4/3               | 4/5                    | 10                   | 5                        | 36/21                                      |
| 7   | 22                              | 3/3         | 3/2     | 4/1        | 4/3               | 3/2                    | 8                    | 4                        | 29/11                                      |

Table 2. The results of assessing the knowledge in group 2 (The values are in the form of fractions: above the slash (/) – the results of the final test, below the slash (/) – the assessment by the students from group 1).

| №   | Overall result of the input test | Summarizing | Grammar | Vocabulary | Sentence structure | Opinions and judgments | Filling out the gaps | True or false statements | Overall result: final test / stud.’s assesm. |
|-----|---------------------------------|-------------|---------|------------|-------------------|------------------------|----------------------|--------------------------|-----------------------------------------------|
| 1   | 36                              | 5/5         | 5/5     | 5/4        | 4/4               | 5/4                    | 10                   | 5                        | 39/22                                      |
| 2   | 33                              | 4/4         | 5/5     | 5/4        | 5/3               | 5/5                    | 9                    | 5                        | 38/21                                      |
| 3   | 38                              | 5/5         | 4/5     | 5/5        | 5/5               | 5/5                    | 10                   | 5                        | 39/25                                      |
| 4   | 39                              | 5/5         | 5/5     | 5/5        | 4/5               | 5/5                    | 10                   | 5                        | 39/25                                      |
| 5   | 35                              | 5/4         | 5/5     | 5/5        | 4/4               | 5/5                    | 10                   | 5                        | 39/23                                      |
| 6   | 32                              | 5/4         | 4/4     | 5/5        | 4/5               | 5/4                    | 10                   | 5                        | 38/22                                      |
| 7   | 37                              | 5/5         | 4/5     | 5/5        | 5/5               | 5/5                    | 9                    | 5                        | 39/25                                      |
| 8   | 37                              | 5/5         | 5/5     | 5/5        | 5/5               | 5/5                    | 9                    | 5                        | 39/25                                      |
| 9   | 33                              | 5/5         | 4/5     | 5/5        | 5/5               | 5/5                    | 10                   | 5                        | 39/24                                      |
| 10  | 40                              | 5/4         | 5/5     | 5/5        | 5/5               | 5/4                    | 10                   | 5                        | 40/23                                      |
| 11  | 39                              | 5/4         | 5/5     | 5/5        | 5/5               | 5/4                    | 10                   | 5                        | 40/23                                      |
| 12  | 39                              | 5/5         | 5/5     | 5/5        | 5/5               | 5/5                    | 10                   | 5                        | 40/25                                      |
| 13  | 30                              | 4/4         | 4/4     | 5/4        | 4/4               | 5/4                    | 9                    | 5                        | 36/20                                      |
| 14  | 37                              | 5/4         | 4/5     | 5/5        | 5/4               | 5/4                    | 10                   | 5                        | 39/22                                      |
| 15  | 35                              | 4/4         | 5/5     | 5/5        | 5/5               | 4/5                    | 10                   | 5                        | 38/24                                      |
| 16  | 32                              | 5/5         | 4/5     | 5/5        | 4/5               | 5/5                    | 10                   | 5                        | 39/24                                      |
| 17  | 35                              | 4/4         | 5/5     | 5/5        | 4/5               | 5/5                    | 10                   | 5                        | 38/24                                      |
| 18  | 35                              | 5/4         | 4/5     | 5/5        | 4/5               | 5/5                    | 10                   | 5                        | 38/24                                      |
| 19  | 32                              | 5/4         | 4/5     | 4/5        | 4/4               | 5/4                    | 10                   | 5                        | 37/22                                      |
| 20  | 28                              | 4/4         | 5/5     | 5/4        | 4/4               | 4/4                    | 9                    | 5                        | 36/21                                      |
Table 3. The results of assessing the knowledge in group 3.

| №  | Overall result of the input test | Summarizing | Grammar | Vocabulary | Sentence structure | Opinions and judgments | Filling out the gaps | True or false statements | Overall result of the final test |
|----|----------------------------------|-------------|---------|------------|-------------------|------------------------|-----------------------|--------------------------|---------------------------------|
| 1  | 28                               | 4           | 3       | 4          | 3                 | 1                      | 8                     | 4                        | 27                              |
| 2  | 8                                | 1           | 1       | 1          | 1                 | 0                      | 3                     | 3                        | 10                              |
| 3  | 7                                | 1           | 1       | 1          | 0                 | 0                      | 2                     | 1                        | 6                               |
| 4  | 30                               | 4           | 4       | 4          | 3                 | 3                      | 8                     | 4                        | 30                              |
| 5  | 37                               | 4           | 5       | 4          | 5                 | 5                      | 10                    | 5                        | 38                              |
| 6  | 9                                | 0           | 1       | 1          | 1                 | 0                      | 2                     | 3                        | 8                               |
| 7  | 30                               | 3           | 4       | 4          | 4                 | 4                      | 6                     | 5                        | 30                              |
| 8  | 22                               | 2           | 4       | 5          | 4                 | 1                      | 5                     | 4                        | 25                              |
| 9  | 23                               | 2           | 3       | 2          | 3                 | 1                      | 5                     | 4                        | 20                              |
| 10 | 20                               | 3           | 3       | 4          | 3                 | 2                      | 3                     | 2                        | 20                              |
| 11 | 26                               | 4           | 5       | 5          | 4                 | 2                      | 4                     | 4                        | 28                              |

4. Analysis

Tables 1-3 let us build all figures below to see the progress of students. In figure 1, figure 2 and figure 3 there are the results (which are total for all the criteria) for both the input and the final tests. For illustrative purposes we arranged the results in ascending order and smoothed them with the neural network.

![Figure 1](image1.png)

**Figure 1.** The results of the input and final tests (lower and upper ones respectively) for experimental group 1, N – students in the scaled sample.

It is obvious from figure 1 that the progress in the experimental group is substantial: the results of the final test are much better than those of the input one. It indicates the efficiency of collaborative learning.

Figure 2 illustrates that the students in group 2 also made a remarkable progress due to their teaching activities. However, the students whose scores in the input test were high did not show considerable improvement in the final test, which is illustrated with the graphs converging in one point. The results of the final test are approaching the saturation, which means that the students with mastery of English should have been given more challenging assignments.
In Figure 3 it is quite obvious that there was very little difference in the results of the input and the final test, which indicates that the traditional teaching and learning process (without students’ collaboration and with the focus on learning the language instead of concentrating on learning the professionally-orientated content), was not very efficient.

![Test results](image1)

**Figure 2.** The results of the input and final tests (lower and upper ones respectively) for experimental group 2, N – students in the scaled sample.

![Test results](image2)

**Figure 3.** The results of the input and final tests (lower and upper ones respectively) for group 3, N – students in the scaled sample.

The results that have been obtained illustrate that it is possible to develop a reliable method of evaluating objective indices using subjective data. However, developing such a method and explaining the reasoning behind it requires further research.

This method can be integrated into the educational process because it allows the teacher to evaluate the progress made by a group of learners in a less time-consuming way. Instead of doing a test, the students will just fill out a short evaluation table with the results of each other’s peer assessment. As the result of that test can be done less frequently.

The teacher needs to assess the level of students’ achievement and performance in order to provide operational management of educational process (the pace of learning, choosing the check points,
determining the level of complexity of the current tests, etc.). Interviewing a few students may appear to be insufficient, whereas interviewing or testing a great number of students is time-consuming. For this reason we suggest that peer assessment by the students collaborating on the basis of the hierarchical system should be used for continuous monitoring. The results of verifying these assumptions on the basis of the data from tables 1 and 2 are considered further.

**Figure 4.** a) The combined results of the test in the mentees’ group as the function of the mentors’ assessment; b) the combined results of the test in the mentors’ group as the function of the mentees’ assessment.

The first relationship (figure 4a) appeared to be linear. The root-mean square error for it accounted for 1.12, the evaluation of the root-mean deviation of the sample was 3.41. If we take into account the error which resulted from the discreteness of assessment and its subjective nature, the result is quite acceptable.

When we considered the second relationship (figure 4b), we doubted whether it would be possible to obtain any meaningful results because we did not expect the students with a rudimentary level of foreign language knowledge to be able to assess the advanced students. However, we managed to come to some conclusions. In order to achieve that, the sampling was smoothed on the basis of the method of least squares through neural network function. The resulting relationship was the following (formula 1):

\[
y = 38.5 + 1.05 \text{Tanh}(0.01x - 2.32) + 1.51 \text{Tanh}(1.13x - 24.2)
\]  

(1)

Neural networks are a standard tool for finding the implicit relationships in the data [6]. The root-mean square error for formula 1 accounted for 0.754, the evaluation of the root-mean deviation of the sample was 1.15. If we take into account the error which resulted from the discreteness of assessment and its subjective nature, the result is quite acceptable.

This formula refers only to the qualitative relationship, but not the quantitative one, and it gives a vivid illustration of the data. This relationship is similar to the relationship that results from applying Rasch methodology and approaches the minimum value at the beginning of the interval and the maximum one at its end. Here it is essential to mention that the mentees’ results cannot exceed 40, so the students from group 2 with the highest level of knowledge had approximately the same results of the final test, because their score was close to the maximum possible value. However, their mentees’ assessments differed from each other. The mentors who were interested in their mentees’ results got the highest scores, whereas the mentors who were unwilling to explain the material got the lowest ratings.

The graph on figure 4b can be divided into two parts. The first part indicates that the subjective and the objective evaluations tend to change in one direction. The second part of the graph is virtually horizontal, which means that the objective evaluations were practically similar, whereas the subjective assessments were different. The similarity of the objective evaluations (i.e. the results of the final test)
can be explained by the fact that most of the students got a maximum or nearly maximum score, which means that the test does not allow distinguishing between their levels of competence. According to the teacher, the real level of the students’ foreign language proficiency does not always correspond to the results of the peer assessment. The students who were interested in their partners’ progress were given higher scores by their mentees.

Further we describe the correlations between the individual students’ results of the final tests and the assessment of their performance by the students who cooperated with them. The criteria of assessment (summarizing the plot, grammar, vocabulary acquisition, etc.) were considered separately. It is impossible to establish a clear correlation for most of the criteria. The most interesting results are considered below:

**Figure 5.** The mentors’ results as the function of their assessment by: a) the mentees; b) grammar vocabulary.

Figure 5 illustrates that there was a certain degree of correlation between the actual results of the test done by the mentors and the subjective assessment of the same results by the mentees. It is obvious that the mentees had difficulty assessing the mentors’ performance. On the other hand, it is easier to assess vocabulary and grammar in quantitative terms than other aspects of learning a foreign language.

**Figure 6.** The mentees’ results as the function of their assessment by: a) the mentors; b) grammar summarizing.

The graph on the left (figure 6a) describes the performance in grammar and shows a clear correlation between the mentees’ objective results of the test and the mentors’ subjective assessment. Such a high degree of consistency can be explained by the fact that the assessment was made by the students who have a mastery of English and who can make judgements about the mentees’ level of knowledge.

![Diagram](image-url)
Moreover, as we have already mentioned, the students’ achievements in grammar can easily be assessed in quantitative terms. The graph on the right (figure 6b) shows the correlation between the mentees’ progress in summarizing texts and the mentors’ assessment of this progress. The correlation is not as clear as in grammar, because it is difficult to describe summarizing in quantitative terms. However, the mentors managed to make expert assessment of the mentees’ progress.

The results given above allow us to come to two main conclusions: Mentors make more objective evaluations of the mentees, whereas the mentees’ evaluations are less accurate. The results are more reliable when we deal with clearer criteria such as grammar.

5. Conclusions
The results that we have obtained reveal that it is difficult to develop a reliable method of assessing the objective indices on the basis of the subjective data. In order to succeed in applying such methods a researcher needs to define the desirable result. It is also important to determine whether it is possible to achieve the result on the basis of the data that are readily available or the data that can be obtained. The easiest way is to make conclusions about quality on the basis of the objective data. For instance, it is easy to come to the general qualitative conclusion about the efficiency of learning in different groups and the effectiveness of the teaching methods on the basis of figures 1-3.

Figure 4 illustrates that quite distinct qualitative conclusions can be made on the basis of the subjective data. Figures 5 and 6 indicate that it is difficult to conduct more detailed studies. More accurate results can be obtained when we deal with more formalized data (such as the level of proficiency in grammar). If the data are less formalized, it is more difficult to assess them because assessing the results of the tests by the teacher can be different from the results of the peer assessment (in this case – by other students).

This method can be applied in higher education, especially in open-source learning, in which students peer assessment is essential. However, developing such a method and explaining the reasoning behind it requires further research, for example designing models using the techniques [4, 5]. Neural networks were also used in measuring the efficiency of applying a hierarchical teaching system based on mentorship in several groups of students [7].

This method can be integrated into the educational process because it allows the teacher to evaluate the progress made by a group of learners in a less time-consuming way. Instead of doing a test, the students will just fill out a short evaluation table with the results of each other’s peer assessment. As the result of that the tests can be run less frequently.

References
[1] Mari L and Wilson M 2014 May An introduction to the Rasch measurement approach for metrologists Measurement 51 315-27
[2] Maul A, Wilson M and Torres I D 2013 On the conceptual foundations of psychological measurement Journal of Physics Conference Series 459(1)
[3] Pendrill L 2014 December Man as a measurement instrument [Special Feature] NCSLi Measure: The Journal of Measurement Science 9(4) 22-33
[4] Blagoveshchenskaya E A, Dashkina A I, Lazovskaya T V, Ryabukhina V V and Tarkhov D A 2016 Neural Network Methods for Construction of Sociodynamic Models Hierarchy Lecture Notes in Computer Science (Springer International Publishing Switzerland) ed L Cheng et al pp 513–520
[5] Tarkhov D A, Shanshin I K and Shakhanov D O 2016 The reversed problem of migration streams modeling Journal of Physics: Conference Series 772
[6] Haykin S 1999 Neural Networks: A Comprehensive Foundation (Prentice-Hall Inc) p 199
[7] Dashkina A and Tarkhov D 2017 The application of neural network modeling in organizing a hierarchical teaching system based on mentorship Complex Systems: Theory and Applications ed R Martinez (New-York: Nova Publishers) chapter 4 pp 95-116