Evaluation of Conceptual Knowledge: Review of the Experimental Study

Evgeny A. Eremin
Perm State Pedagogical University (Russia)
eremin@pspu.ac.ru

Abstract

The main aim of the research is to study how students link separate terms and concepts into common interrelated picture during learning a course. The evaluation was made before and after learning the discipline “Computer architecture”. Students were offered to combine pairs of basic related terms from the predefined list and to specify a type of every relationship. Computer processing of the obtained files with results allowed to form independent groups of interrelated concepts (clusters) and also to produce some statistical characteristics of students’ knowledge structure. Derived numeric gauge of knowledge organization can be accepted as some measure of learning success. It was discovered that according to experimental results several types of learners can be marked out.

Abstract

1. Introduction

Learning theory [1, 2] takes up the position that structured knowledge is much more preferable than solely isolated facts. Unfortunately traditional testing checks mainly the factual knowledge. As experiments show (see [3] for example), assessment based on knowledge integration is better than traditional multi-choice tests. Besides many publications, including the papers from previous edition of “The Future of Education” [4], show that a learning process can be described as linking of new concepts to knowledge structure, already existing in student’s mind. So my research was aimed at investigation of students’ ability to link separate terms and concepts into common interrelated picture during learning a course.

In 2008 I developed the experimental method for evaluation of knowledge structure’s entirety (detailed description can be found in [5]) and made the first attempt to essay it on my students. After some modification this method in 2009-2011 was used to estimate learning achievements of several student groups. The analysis of experimental results gave possibility to select the most suitable statistical characteristic of knowledge structure and to discover several interesting pedagogical findings.

All tested students studied at the physical faculty of our Perm State Pedagogical University; they had two different specialties: teachers of physics and informatics and educational specialists on ICT. Total number of students in all experiments amounts to 79.

2. Experimental method

Procedure of evaluation was organized the following way.

According to my personal preferences, introductory learning course about computer organization was selected for the investigation. Its knowledge domain consists of basic terms and interrelations between them. My study assumed, that the rate of digesting and mastery of these terms verified the success of the discipline. Details of building the lists from terms and types of relationships can be found in previous publications [6] and [5]. Here we’ll content ourselves with several examples of relation pairs: *processor – whole/part – arithmetic and logic unit* (ALU), *software – class/subclass – operating system*; *operating system – control – hardware*, *binary system – related topic – bit* and so on. Only general concepts were under consideration. Contrary, the list did not include the names of concrete operating systems, external devices and their manufacturers, and other similar data, less essential from the position of learning the main course’s regularities. Using the standard terminology from object-oriented programming, we may say, that classes of the concepts were under consideration, but not their instances.

It is worth to mention, that full list of discipline’s concepts contained 122 terms. Such size was large enough, because, as further experiments show, competent students usually used a little more than the half of terms in their answers. Two last groups were tested with reduced list from 80 concepts and results look quite similar.

Special software was written to support my experimental study. The first program fixed the associations between the concepts that students marked in a text file, suitable for further computer analysis. Another one helped experimenting teacher to analyze the results of knowledge checking and to get different statistics.
Combining all pairs, related to every concept, in the issue we get several disjoint groups of interconnected terms. In ideal case all concepts of the course must be interdependent, but real students' results are not so monolithic: at best they form 2-5 independent groups, but some students produce until 20 small groups.
Other details of testing procedure can be found in [5].

3. Discussion of results

3.1. Diagrams of concepts

To analyze the parcelling of linked concepts, we can use an original form of visual representation – diagrams of concepts' interconnection (peculiar map of knowledge entirety). A typical picture for one academic group from 12 students is presented in fig.1.

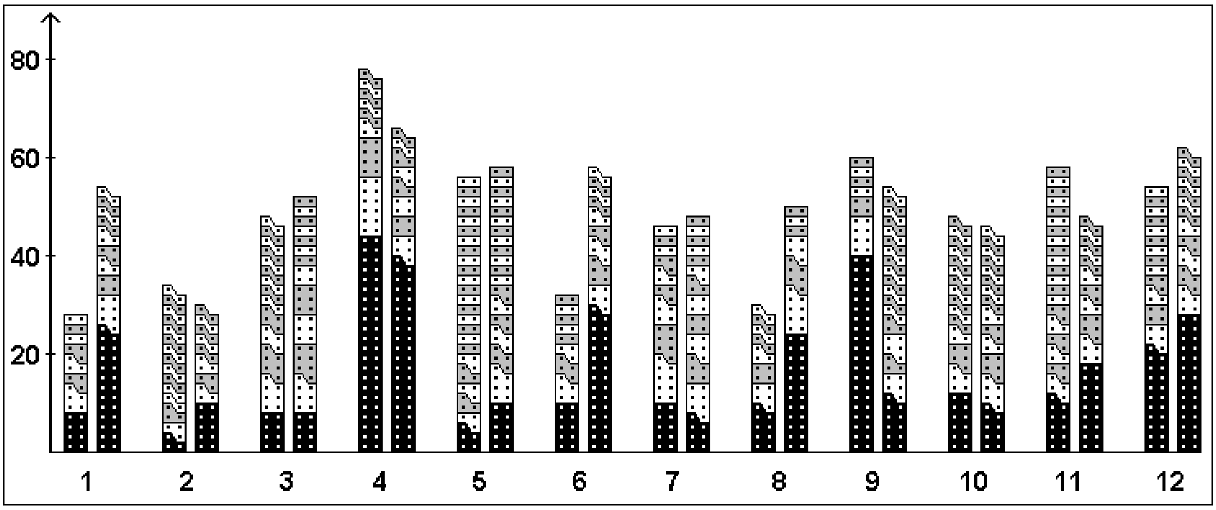

Fig.1. Diagram that shows groups of interconnected basic concepts (students are arranged according to their rating)

This "spotted" diagram is organized the following way. Pairs of columns we see on it represent input and output testing: the right bar in every pair indicates final results. Height of any column is proportional to the number of terms that student mentioned during assessment. Every dot in a column means one individual concept. All columns are divided into several areas; each of them images a group of interrelated concepts (cluster), learnt by student. For better visibility, neighbour areas are painted in white and grey colours. The black region in the bottom, which is always the largest, symbolizes the kernel of student's knowledge. As you can notice, all groups in every bar are regularized by size, so the smallest groups from 2-3 concepts (such groups may be interpreted as separate facts out of common picture) always reside at the top of a bar.
As it was mentioned above, in ideal case every diagram bar must be heterogeneous and black (consisting of the only group), and its height must include all the concepts of the course. Real picture, as you see from fig.1, is far from ideal one.
Students in the diagram are rank-ordered according to some rating: the criterion of such arrangement is a *time of finishing all the tasks*, given by the teacher. The students with small numbers *n* finished the course earlier; hence they are supposed to demonstrate better results in learning the course content. In opposite, "the slowest" students with larger values of *n* form the right part of the picture. It is important that students' rating is defined absolutely independent from pretest data: such approach helps to avoid the regression towards the mean effect.
For several groups of students we can confront our learning success measure with the results of traditional examinations. Calculated coefficient of linear correlation between the examination marks and proposed students' rating is negative and has high magnitude near –0.8, so it means that the

large rating value more probably will conform to low examination marks, i.e. these estimations are similar.

Studying such representation of interconnections, we may get several practical conclusions how successful was the digesting of learning content. For instance, the results for students number 3, 5 and 12 had slightly changed after studying the course; students 1, 6 and 8 demonstrate evident growth of conceptual knowledge, because the structure of their diagrams is rather better: the number of areas decreases, and their size, in opposite, increases. At the same time students 4 and 9 show down grade of factors: even the height of their bars, which depends from the number of digested concepts, fall off after the course. We may assume, that this paradox can be explained by students' intention to pass the last test quicker and get their long-expected credit.

3.2. Numeric characteristic

As analysis of the experimental data show, the most suitable numeric characteristic for interrelated groups of concepts is their average size. When a student selected *T* terms, which form *G* groups, this value is equal to *T/G*. Results for the same academic group is demonstrated in the left half of fig.2. Students are ranked the same way as it was described above for fig.1.

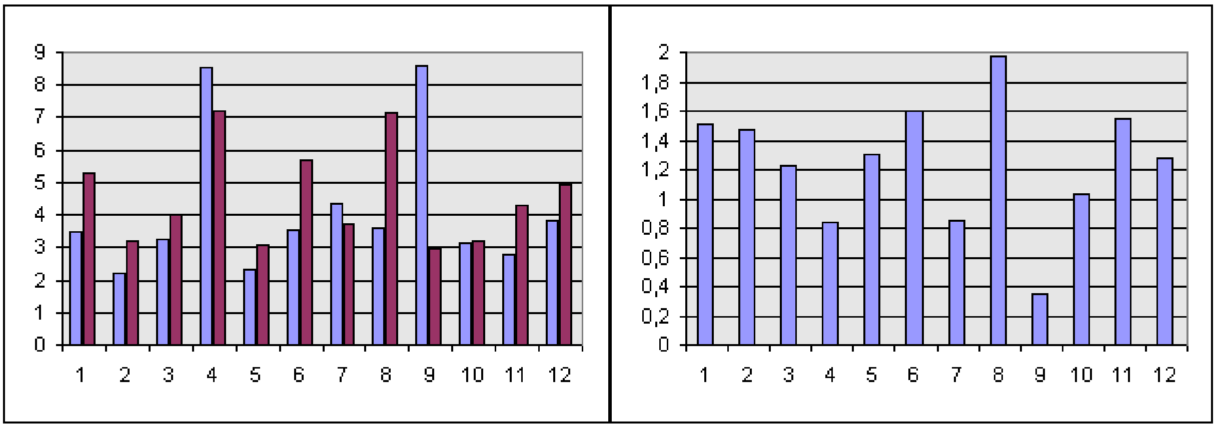

Fig.2. Average group size for pre- and posttest (left) and ratio of these results (right) for students, independently ranked according to their course results

If we compare fig.2 with fig.1, we can see that magnitude of *T/G* is a good integral characteristic of terms interrelation. When we divide a height of the posttest bar (right bars in each pair) by a height of the correspondent pretest one (the left bars), we get a relative measure of change K_{tg}, presented in the right part of fig.2.

Common sense suggests, that $K_{tg} > 1$, because every student must get some new knowledge after learning, however several students demonstrated lower values. Such paradox often occurs in pedagogical experiments [7].

Diagrams for other groups have similar view. We may notice that academic group is too small to exhibit clear statistic regularities. So we ought to combine the results of all students and analyse them together.

3.3. Combined diagram

To make the results of different groups comparable we must preliminary reduce rating numbers *n* to common scale, because groups have different size *N*. Let us introduce the formal variable *X* to describe student's rating. We assign value X = 0 to the first student with n = 1, and X = 1 – to the last one with n = N. For any other student with rating *n* we can calculate variable X_n, using formula

$$X_n = (n-1) / (N-1), \quad \text{where } n = 1, 2, \dots, N$$

Using the introduced above scale, we can put dots for growth coefficient K_{tg} (see right diagram in fig.2 for example), preliminary translating abscissa X_n for every student. The result scatter plot is shown in fig.3, where every point conforms an individual student.

Several different zones can be marked out in fig.3.

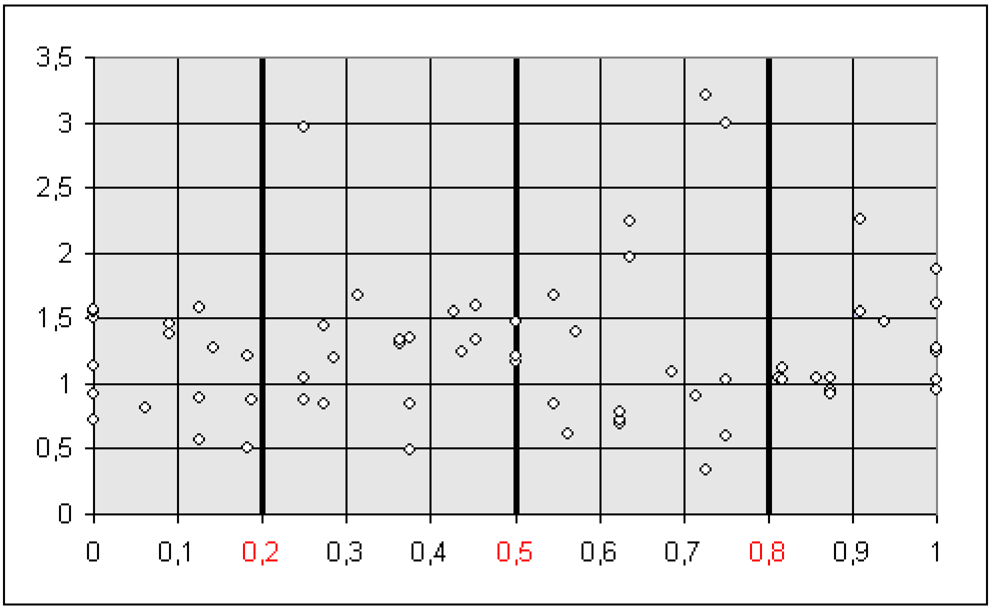

Fig.3. Combined scatter plot for growth ratio as a function of rating *X*

First zone *Z1* at X < 0.2 has noticeable dispersion of experimental dots, many of them are in the "incorrect area" $K_{tg} < 1$. This zone corresponds to advanced students who can easily do all teacher's tasks. Hence motivation strongly influences on their work: if they are not interested, they may give answers anyway, hurrying to finish testing and sometimes it leads to very low output results.

Second zone *Z2* at 0.2 < X < 0.5 has moderate but stable growth of value we consider (if neglect extreme point at X = 0.38). Zone Z2 is a place for middle students, who thoroughly do all tasks and usually get steady results.

Next zone *Z3* at 0.5 < X < 0.8 is again diffused, even more than Z1. I suppose the reason is that Z3 comprises a mix from two kinds of students: strong but studying bad (laziness, illness or other whys) and really weak students who can't study better. Therefore we have very high points and very low ones in this zone.

And the last (very amazing) zone *Z4* at X > 0.8 has all dots above 1 with high values in its right half. We can explain this phenomenon the following way. The low-perform students have nothing but hard work to complete the course, so they are obliged to increase their knowledge, doing their best. Similar stable achievements of low-score students were also reported in [8]. In my opinion, formulated conclusion has deeper sense: conceptual knowledge comes with hard intellectual efforts only. Hence this phenomenon must appear in the first instance for those learners, who need to put on all their abilities while reaching standard satisfactory level.

The last conclusion from fig.3 is that some students have extraordinary high results $K_{tg} > 2$, but their amount is low – close to value 5-6% reported in [9]. It is notable that these students usually do not demonstrate the best rating *X*.

4. Conclusion

This paper proposes the new method to measure educational achievements, which is based on estimation of knowledge structure's entirety. Several quantitative indicators were tested in experiments. Finally special procedure of grouping related concepts was developed to calculate the average size of isolated knowledge clusters that students remembered after completion of the course.

The analysis of experimental results leads us to some interesting pedagogical findings. First of all, some students, as it turned out, can structure knowledge essentially better than others, but unfortunately their amount is small. Another result (really surprising) is that the most evident increasing of numeric parameters belongs to low-perform students.

The joint diagram demonstrates four different zones with specific allocation of dots.

Proposed computer technology of testing is not difficult and teachers can try to use it for any other educational course with well-structured concept base.

References

[1] Anderson, L.W., Krathwahl, D.R. (2001). *A Taxonomy for Learning, Teaching and Assessing*. New York: Longman.

[2] Jonassen, D.H., Beissner, K., Yacci, M. (1993). *Structural Knowledge: Techniques for Representing, Conveying, and Acquiring Structural Knowledge*. Hillsdale, NJ: Lawrence Erlbaum Associates.

[3] Lee, H.S., Liu, O.L., Linn, M.C. (2011). Validating measurement of knowledge integration in science using multiple-choice and explanation items. *Applied Measurement in Education,* 24(2), 115-136.

[4] González, R.L., Luis Manuel Casas García, L.M.C., García, M.M., Masa, J.A. (2011). Possibilities of “Nuclear Concepts Theory" on Educational Research, a Review. In. *Proc. The Future of Education*, 326-330. Milan: Simonelli Editore – University press.

[5] Eremin, E.A. (2012). Original Experimental Method to Evaluate Conceptual Students’ Knowledge. *Procedia - Social and Behavioral Sciences*, 55, 1227-1232. http://www.sciencedirect.com/science/article/pii/S1877042812040815

[6] Eremin, E.A. (2007). Using Topic Map technology in the planning of courses from the CS knowledge domain. In *Proc. 7th Baltic Sea Conference on Computing Education Research*. CRPIT, 88, 179-182. Sydney: Australian Computer Society. http://crpit.com/confpapers/CRPITV88Eremin.pdf

[7] Lord, F.M. (1956). The measurement of growth. *Educational and Psychological Measurement*, 16(4), 421–437.

[8] Libarkin, J.C., Anderson, S.W. (2005). Assessment of learning in entry-level geoscience courses; results from the geoscience concept inventory. *Journal of Geoscience Education*, 53(4), 394-401.

[9] Kay, A. (1995). *Powerful Ideas Need Love Too!* Written remarks to a Joint Hearing of the Science Committee and the Economic and Educational and Opportunities Committee