EJPCE
Eurasian Journal of Physics and Chemistry Education

Eurasian J. Phys. & Chem. Educ. 5(1): 39-56, 2013

journal homepage: http://www.eurasianjournals.com/index.php/ejpce

Chunking in Chemistry

Denis M. Zhilin[1,*] and Larisa E. Tkachuk[2]

[1]Moscow Institute for Open Education, Moscow, Russia
[2]School #6, Nyagan, Russia

Received: 02 October 2012- Revised: 30 December 2012- Accepted: 15 January 2013

Abstract

Experts (2 groups, 41 participants) and non-experts (7 groups, 130 participants) in chemistry were exposed to sequences of chemical symbols for 30 seconds and then, after a minute, wrote them down. Four sequences were real chemical equations and four sequences were fakes got from the equations by random mixing the symbols. Experts in chemistry recalled real chemical equations (a) much better than fake sequences and (b) better than novices in chemistry. Meanwhile, no significant difference was found between the experts and the novices in recalling the fake sequences. Besides, the novices remember both real and fake sequences symbol-by-symbol from left to right, the frequency of mistakes increasing in the same order. The experts remember real equations as a whole and chunk some chains in the fake sequences remembering them better than others. All these results provide evidence that experts chunk chemical information and allow extending the chunking-based theories to chemistry teaching.

Keywords: Chemistry Teaching, Chunking Theory

Introduction

Chunk-based theories describe the information processing mechanism with regard to the expertise. They classify memory into the working memory (sometimes considering the short-term memory as a part thereof, Yuan et al., 2006, and references therein) and the long-term memory. The working memory has a limited capacity (Miller, 1956), while the long-term memory is almost unlimited (Reid, 2008 and references therein). The working memory acts as a temporary area for activities such as mental arithmetic, reasoning, and problem solving (Baddeley, 1999), processing the information from the long-term memory. In the long-term memory, the information is stored in connected pieces that can be retrieved by a single act of recognition. These pieces are referred to as chunks. The chunks have strong associations with one another but weak associations with elements within other chunks (Gobert et al., 2001). Experts store more chunks (about 50000, Gobert & Simon, 1996b) in their long-term memory than novices and these chunks are much more complex. The expertise also allows quick recognition of familiar chunks in the long-term memory. While solving problems, experts use so-called macro-operators, chunking together a sequence of operations (Koedinger & Anderson, 1990, and references therein), thus solving rather complicated problems with limited capacity of working memory. For the model developed by Brooks & Shell (2006),

*Corresponding Author, Phone: +7-916-820-62-54 , E-mail: zhila2000@mail.ru

ISSN: 1306-3049,

expertise is thought of forming ever-larger knowledge chunks, and ability is related strongly to working memory capacity.

The chunking theory was founded by De Groot (1946/2008) and developed by Chase & Simon (1973). They found out that chess experts are able to reconstruct with high accuracy a game position that they had seen only for a few seconds. However, when the pieces are randomly placed throughout the board, experts perform only slightly better than do novices . It appears that experts' superiority in memorizing chessboard positions arises from the existence of a large store of intact and well-organized chess configurations or patterns (chunks) in the long-term memory. They also pointed out that chess experts investigate a small set of "good" alternative moves, while novices don’t make difference between “good” and “bad” ones. Additional research showed that expertise is a function of skilled pattern recognition (Cooke, 1992) and recall of domain relevant patterns (Ericsson & Staszewski 1989). Later, the chunking theory was extended to a template theory (Gobet & Simon, 1996; 1998). Templates, besides containing chunks, possess slots (variables that can be instantiated) in which some new information can be stored in a matter of seconds. Anderson (1996, 2004) developed ACT-R model, including operations with chunks.

Chunking-based theories of expertise were generalized to different domains such as computer programming, medical diagnosis, engineering (Gobet et al., 2001), architecture (Akin, 1986). Egan & Schwartz (1979) repeated the same experiment as Chase & Simon (1973) in relation to remembering of electrical circuits and obtained the same results for experts and electronics technicians. While solving physical problems, the experts choose appropriate schemas (that could be regarded as chunks) via categorization (Chi et al., 1981). Experts in geometry parse geometry problem diagrams into perceptual chunks and use macro-operators with step-skipping (Koedinger & Anderson, 1990). However, the expertise itself is strongly domain specific (McNeil et al., 2006; Sweller et al., 2007 and references therein).

Besides chunking-based theories of cognition, there are chunking theories of language acquisition and symbol sequences. In remembering symbol sequences (for example, telephone numbers), they are split into groups (for example, area, district and number, Kellett & Johnstone, 1980). Chunking operates by taking two or more items that *frequently occur* together and combining them into a single automatic chunk (such as “what’s this” and more complex). Professional interpreters use chunks rather than separate words (McWhinney, 2004 and references therein). Grammatical knowledge is implicitly encoded in a hierarchical network of chunks (Servan-Schreiber & Anderson, 1990). People also use chunking while learning symbol sequences within artificial grammar (Boucher & Deines, 2003). While measuring the response time in the task “name the letter before (or after) the exposed” and asking the participants from what point of the alphabet they began to search, Klahr et al. (1983) isolated several sequential parts, i.e., chunks. While searching for a letter, a person first chooses a chunk where it is located. We are not sure about the equivalence of “chunks” providing sequence remembering, on the one hand, and conceptual understanding within science on the other hand. However, they can merge namely in the field of chemistry, because chemical equations are a kind of language (or at least symbol sequences) and the linguistic concept of chunk would help us to interpret the results of our experiments. One more argument in favor of applicability of the linguistic concept to chemistry teaching is the finding of Howe (1971) that pupils rather memorize formulas and equations than get them by reasoning.

Recently, the psychological reality of chunks was also confirmed by neuroscience investigations (fMRI – functional magnetic resonance imaging in particular). The long-term memory is associated within the temporal lobes while short-term memory is within the frontal and parietal lobes (Campitelly et al. 2007 and references therein). Relying on these data, Campitelly et al. had shown that chess players recalled the game position seen for a short time ("short-term memory task") with participation of the long-term memory areas. For recalling geometrical drawings, they employed only short-term areas. Abe et al. (2007) let participants pre-learn 4-symbol sequences and then let them learn 8-symbol sequences, examining the brain activity by means of fMRI. When the 8-symbol sequence consisted of pre-learned 4-symbol ones, the temporal lobes were activated, and this was not observed while remembering totally new 8-symbol sequences. This also proves the involvement of the long-term memory in solving short-term tasks on remembering symbol sequences.

Although the up-to-date chunking-based models are quite complicated (e.g. Sweller, 2003; Anderson et. al, 2007; Reid, 2008), their core idea itself has at least two practical consequences:

- to teach a particular subject or to give a particular problem, a teacher should (a) trace that all the necessary amount of information is organized into a limited number of chunks and (b) develop all the necessary chunks beforehand (or make sure that they are developed);

- the expertise could be evaluated as complexity and number of chunks in a certain domain that is promising for evaluation (e.g. Sirhan, 2007; Walczak & Fishwick, 1997).

We teach chemistry, so we are interested in foundations and applications of the chunking-based theories within this domain. Basing on chunking theories, Reid (2008) offered (and checked) the idea of pre-learning some chemical concepts and proved its efficiency. Danili & Reid (2004) constructed teaching materials to minimize any limitations to learning caused by working memory, which improved teaching. El-Banna & Johnson (1986) measured the working-memory capacity by two kinds of tests and compared it with the performance in answering questions on chemistry of different complexity. They found an abrupt decrease of performance while increasing complexity. The complexity of question that students could not answer corresponded to their working memory capacity and this proved the role of working memory capacity in answering questions in chemistry. The Glasgow University group applied analysis of the required concrete and operational stages to estimate the difficulties of some topics in a syllabus (Johnstone, 2006 and references therein). Taasoobshirazi & Glynn (2009) modeled chemistry problem-solving strategies. They used the concept "scheme" ("well-organized, easily accessed, conceptual knowledge structure"), which seems to be similar to "chunk". However, neither the references therein nor their own work regard "scheme" as a piece of information proceeded by working memory *as a whole*.

The above-mentioned works prove the role of the working memory in studying chemistry and in solving chemical problems. However, they do not answer the question of whether chemical information is *chunked*, i.e., organized in well connected pieces of information that are operated as a whole. Meanwhile, domain specificity of expertise brings about the eternal suspicion that chunking-based theories are also domain specific. Hence, we need to check their applicability to a particular domain if we want to use it in that domain. Only if information in chemistry domain is chunked we could use all the findings of chunked-based theory in teaching chemistry. Then we should construct learning programs to develop chunks – from elemental to complex and hierarchical.

The easiest way to check whether chemical information is chunked it is to repeat experiments of Chase & Simon (1973) replacing the chess positions by something peculiar to chemistry (as Egan & Schwartz, 1979, did for electronic circuits). The most peculiar thing is a chemical equation. Therefore, we compared experts and novices in chemistry in remembering and recalling real chemical equations (referred below as "true equations") and the chains of the same symbols in random order (referred as "fake chains"). Our research question how experts recall real and fake sequences in comparison with novices. If (a) experts and novices recall fake chains equally to true equations and (b) experts recall true equations much better than novices, then chunking-based theories can be developed for the chemistry domain. Otherwise, they are not applicable.

Methodology

Design

In the experiment, we compared the performance in recalling real chemical equations and random (“fake”) sequences consisting of the same symbols for nine groups of participants with different level of expertise in chemistry. The total amount of retrieved information was calculated for each sequence (real and fake). The portion of participants of each group that correctly recalled a particular symbol was also calculated.

Participants

All the participants were from Russia, with Russian native language. However, all of them had been studying at least one European language at least for three years before, so they were quite familiar with Latin characters.

The nine groups of experiment participants ranged from totally ignorant to experts (Table 1). Seven of them were pupils of an ordinary Russian provincial school. The pupils of the 7th grade (N7) had not studied chemistry. They had studied elementary science course at 1-4th grades (1 hour a week) but it did not cover the topics of chemical reactions and chemical equations, so chemical equations were absolutely unfamiliar for them. So their expertise could be regarded as zero – they only knew the Latin alphabet (not chemical symbols). The 8th grade pupils (group N8) had been studying chemistry for half a year (2 hours a week), so their expertise could be regarded as beginning (in other words, they can be regarded as novices). However, the pupils had been selected to this class by interview, so their general cognitive abilities could be different from average. The 9th grade pupils had been studying chemistry for a year and a half (2 hours a week). Three groups were taken from three classes. Group N9s comprised motivated children with broad interests and the highest performance in chemistry among all the 9th grade groups. Their level of expertise we should estimate as medium. Group N9m comprised children with interests in the humanities who learned by diligence rather than by abilities – a group of medium performance. Group N9w included children with low motivation and bad performance in chemistry thus its expertise level is low. The 11th grade pupils (specialized in sociology and economics – N11e and specialized in informatics – N11i) had finished studying inorganic chemistry a year and a half before the experiment (at that time, they studied organic chemistry) and were not motivated in studying chemistry. So, their level of expertise could be regarded as being lower than that of the N9s group.

The rest two groups were 1st year students of M.V. Lomonosov Moscow State University, the Department of Chemistry (Ch group) and geochemists from the Department of Geology (GC group). Bearing in mind that (a) the M.V. Lomonosov Moscow State University is the best tertiary institution for chemistry in Russia, (b) the students are specialized in chemistry and (c) they had studied inorganic chemistry at the University for half a year before the experiment, they could be regarded as experts. The level of expertise of Ch group could be estimated as the highest, because they had passed the entrance exam in chemistry. The total number of participants was 171.

Table 1. The groups of participants (in order of *a priori* estimation of level of expertise)

Group	Number of participants	Grade	Age	Level of expertise
N7	19	7	12-13	Zero
N8	23	8	13-14	Low
N9w	14	9 (weak group)	14-15	Low
N11e	16	11 socio-economic	16-17	Low
N9m	22	9 (medium group)	14-15	Medium
N11i	21	11 informatics	16-17	Medium
N9s	15	9 (strong group)	14-15	Medium
GC	26	first-year University (Geochemistry)	17-18	Very high
Ch	15	first-year University (Chemistry)	17-18	The highest

Given sequences

Four real chemical equations and four fake chains were used; the letter “r” at the number of the sequence denotes real equation, “f” means fake sequence). The fake chains were obtained from the corresponding equations by randomly mixing symbols on the left of the equality sign and on the right (the symbols were not transferred from left side to right and *vice versa*). The coefficients were placed only after “+”, “=” signs or at the beginning of a chain (this is necessary to calculate information in a chain correctly) and “+” sign with a coefficient regarded as *one* symbol. The total amount of information contained in each equation (bits) was calculated based on the frequency table published elsewhere (Zhilin, 2010). In Russian educational tradition the states of chemicals (e.g. “liquid”, “solid” etc) are not depicted in the equations, so we did not depict them and did not take them into account calculating the amount of information. The sequences were exhibited in random order. This table also indicates the groups that should know the corresponding equation from previous learning.

Table 2. Sequences that were exhibited to the participants

#	Sequence	Familiar to	Order in exhibition	Number of Elements	Number of Symbols	Number of Bits
1r	$\mathrm{2NaCN+H_2O+CO_2 = Na_2CO_3+2HCN}$	Ch	7	5	22	94
1f	$\mathrm{2H{+}{+}CNO_2CNa_2O = NaH_3+2CCN_2O}$		4			
2r	$\mathrm{Na_2SO_4+Ba(NO_3)_2 = BaSO_4{\downarrow}+2NaNO_3}$	all but N7	6	5	24	117
2f	$\mathrm{()Na+3S_4NBaO_2O_2 = ONaBaO{\downarrow}SN_3+2_4}$		8			
3r	$\mathrm{2Al+2NaOH+6H_2O = 2Na[Al(OH)_4]+3H_2{\uparrow}}$	all but N7, N8	3	4	25	130
3f	$\mathrm{2OAl+2+6H_2OHNa = 2([{\uparrow}]Al_4+3OH_2NaH)}$		1			
4r	$\mathrm{K_3AsO_4+2KI+H_2SO_4 = 2K_3AsO_3+I_2+K_2SO_4+H_2O}$	Ch	2	6	33	149
4f	$\mathrm{{}_2+K_4OAs+2_3SHO_4IK = OK_2+HS_4KO_{23}+O_2As+2I_3}$		5			

All the equations were rather complex but this turned out to be reasonable; because they allowed discriminating the groups (see “Results and discussion”).

Procedure

Each sequence was presented for 30 seconds from overhead or media-projector. We used 30 s presentation time based on the chess experiments (Gobert & Simon, 2000). At this time the performance *versus* time curve is rather gradual, but the difference between experts and novices is quite significant. Then the participants waited for one minute and then they were given one minute to write the sequence on a sheet of paper. All the participants reported that one minute was enough to reproduce everything they have remembered.

Before the experiment we told the participants that we perform a kind of a memory test for scientific purposes. Then we described them the conditions: they look for 30 seconds at the sequence, then wait for one minute and then should write everything they remember in the same order. Then we told that the results will not influence their school marks and that anybody can refuse or quit at any moment without explanations. At least we made sure that all the participants can see all the symbols clearly by exhibiting a test chain.

Data treatment

The difference between two sequences can be described by Damerau-Levenstein distance (or edit distance) – the number of deletion, substitution, insertion and adjacent transposition operations needed to transform one string into another (Damerau, 1964; Levenstein, 1966). However, the more the distance between given and recalled sequence, the *worse* is the performance. To characterize the performance positively we used the number of correctly recalled symbols – the number of symbols in given sequence minus the number of errors. The errors were omissions of symbols (that should be corrected by insertions), insertions (that should be corrected by deletion), substitutions and adjacent transpositions. The symbol was scored as correct if it was a part of correct chain, namely, a chain of symbols present in the sequence. Omitted symbols were ignored. If a surplus symbol was inserted, the correct symbol next to it was not scored (to count insertion as an error). If a symbol was substituted by another one, it was not scored. If two adjacent symbols were transposed, only one of them was scored.

For example, if a participant wrote "$2H^{+}+NC_2NaO = 2C_3H+2CCNH$" instead of "$2H++CNO_2CNa_2O = NaH_3+2CCN_2O$" the symbols were scored as correctly recalled (marked by "1") according to

Table 3. In this example 14 symbols are considered to be recalled correctly. Each correctly recalled symbol carried a certain amount of information (bits), which was summed up to give the amount of information recalled for the sequence. The information contained in real equations will be referred to as "real information"; the information contained in fake sequences is called "fake information"; the amount of information recalled for all four real equations is the "total recalled real information"; the amount of information recalled for all four fake sequences is the "total recalled fake information". Sometimes we express the amount of recalled information as the portion of recalled information, i.e., the amount of recalled information divided into the amount of information contained in the particular sequences.

Table 3. An example of scoring correctly recalled symbols

Symbol in sequence	2	H	+	+	C	N	O	$_2$	C	Na	$_2$	O	=		Na	H	$_3$	+2	C	C	N	$_2$	O
Reported symbol	2	H	$^+$	+	N	C		$_2$		Na		O	=	2	C	$_3$	H	+2	C	C	N	$_2$	H
Score	1	1	0	1	1	0	0	1	0	1	0	1	1	0	0	1	0	1	1	1	1	1	0

In the symbol-by-symbol treatment, we calculated the frequency of symbol recalling by dividing the number of people in a group that recalled a symbol into the total number of people in this group.

Statistical hypothesis were checked in Statistical 7.0 program (StatSoft Inc.). The normality was tested using Shapiro-Wilk criterion for the 0.05 level unless indicated otherwise.

Results

Overall

In total, 10556 symbols of 17784 (59%) were omitted (not recalled) or recalled with errors (substitutions or transpositions). Besides, 228 symbols were inserted. This gives 10784 errors. The vast majority of errors (88%) were omissions of symbols, 8% of errors were substitutions, 2% – adjacent transposition, and 2% – insertions. This means that students rather forget information than confuse it.

Among all of the groups, only participants of N7 wrote wrong chemical symbols (for example, “B” instead of “Ba”). This means that all the students who had been studying chemistry percept chemical symbols of two characters as a unit.

No significant gender differences were observed either for real equations or for fake sequences (both, Mann-Whitney and Kolmogorov-Smirnov tests, p=0.05, Figure 1), so in what follows, we do not need to take gender differences into account.

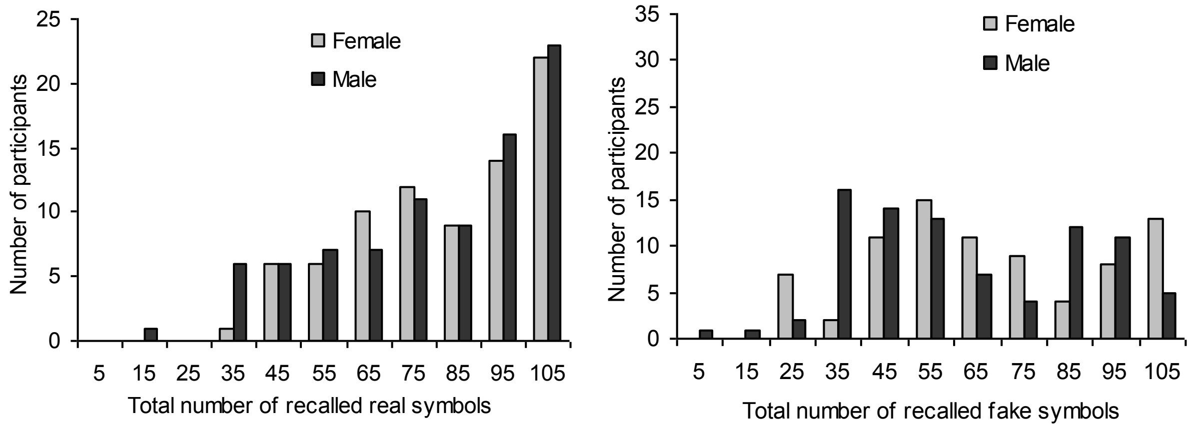

Figure 1. Distribution of total amount of recalled symbols among all the participants.

General recalling of information

The total recalled real and fake information for all groups of participants is compared in Figure 2. The normality of the results is presented in

Table **4**. The expert groups (GC and Ch) recalled almost all the real equations with small number of errors, which led to minor standard deviations (compared with other groups) and non-normal distribution of the results (it is not symmetric because the 100% recalling outweighs). They were significantly different from each other, because the Ch group performed almost ideally (ANOVA followed by Tamhane T2 test). Group Ch differed significantly from any of the non-expert groups, GC differed from all groups except for N8 (that will be discussed later). However, the amount of fake information recalled by expert groups was (a) distributed normally, (b) distributed much more widely and (c) not significantly different from any other group except for N8 or N11e.

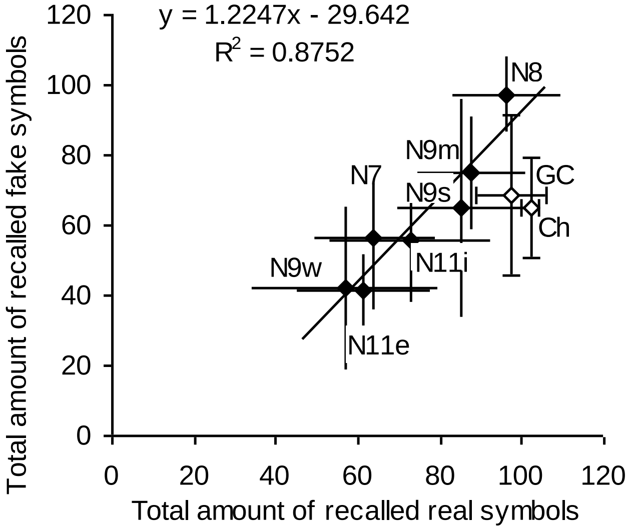

Figure 2. Total recalled real and fake symbols. Error bars denote standard deviations. Expert groups (GC and Ch) are excluded from the trend line and R^2 calculations.

Many experts (13 GC participants out of 26 and 6 Ch members out of 15) but not novices made a remarkable error: they had replaced "=" sign by "→". It is the only substitution that does not change the meaning of the equation (and was not regarded as a mistake while treating the data). So, even these results confirm that the idea of chunking is suitable for chemistry at least to the same extent as for chess.

Table 4. Distribution of recalled information for different sequences (Shapiro-Wilk criterion, p=0.05). N = "normal", U = "non-normal"

Group	Total real	Total fake	1r	1f	2r	2f	3r	3f	4r	4f
N7	N	N	N	N	U	N	N	N	N	N
N8	U	U	U	U	U	U	U	U	U	U
N9w	N	N	N	N	U	N	N	N	N	N
N11e	N	N	N	N	N	N	N	N	N	U
N9m	N	N	U	U	U	N	U	N	N	U
N11i	N	N	U	N	U	N	N	N	N	U
N9s	U	N	U	N	U	U	U	N	U	N
GC	U	N	U	U	U	N	U	N	U	U
Ch	U	N	U*	N	U*	N	U	U	U	N

*All the participants recalled all the information

For non-expert groups, the performance between recalled real and fake information is significantly correlated (Figure 2) and the trend lines are usually close to $y = x$. The results for expert groups worsen the correlation. While comparing the recalled true and fake information for individuals (e.g., Figure 3), we found significant (p=0.01) correlations for all non-expert groups except for N9s and N11e. Groups N9s (the strongest of non-expert groups) and GC displayed moderate correlations, whereas Ch and N11e showed no correlation. The correlation for non-expert group confirms that they employ the same mechanisms for memorizing real and fake information, whereas experts use different ones (more effective for real information). The moderate correlation in the case of GC together with the trend line far from $y = x$ (namely, $y = 1.2x - 0.5$) may be due to *several* of participants that used the same mechanism for fake and real sequences. However, we cannot explain the lack of correlation for N11e group.

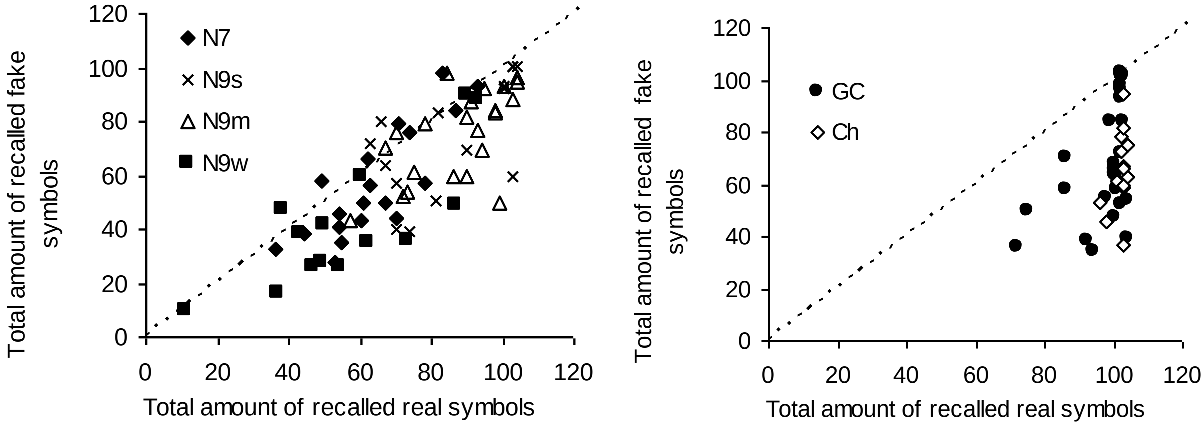

Figure 3. Total amount of recalled symbols in real and fake sequences for participants of four non-expert and two expert groups. The dotted line is $y=x$.

If we compare the recalled real and fake information for each real–fake pair (Figure 4), the picture will be almost the same, although less vivid. For sequence 3, the small difference between the expert groups and groups N9m, N9s is attributable to familiarity of a particular equation to these groups. For non-expert groups, real and fake recalled information for each pair of sequences (e.g. Figure 5) usually correlates. Among the 28 possible pairs (4 pairs for each group multiplied by 7 groups), strong correlation (p=0.01) was observed in 25 cases. For non-expert groups, no correlations were observed at all. This confirms our conclusions concerning the total recalled information, in particular, that non-experts use the same mechanism for remembering the real and fake information, while experts use different mechanisms.

Of particular interest was the N8 group, which remembered both real and fake sequences very successfully (with no significant difference between true and fake ones according to paired-sample t-test and sign test) and the distribution was not normal because 100% success outweighed. This means that they have either extremely high working memory capacity or extremely high focus of attention capacity limit (Cowan, 2001). We think that it is a result of selection procedure for that particular class, but cannot tell anything about the mechanism.

The other non-expert groups recalled the real sequences poorly but much better (according to both the paired-sample t-test and sign test) than the fake ones. For N7 group, it can be seen in Figure 3. This could mean that real equations have some structural features that facilitate remembering even for people who do not know chemistry at all. For example, indexes always follow symbols, "O" symbol is at the end of substance formulas, the length of chains between "+" signs is limited, etc.

One can also see that even the novices recalled about a half of presented information (some of them, almost all information). This caused some difficulties in discriminating experts and novices. This means that either the sequences should have been more complex (which is difficult to compose, because equation 4t is rather complex) or we should have given less time for memorizing.

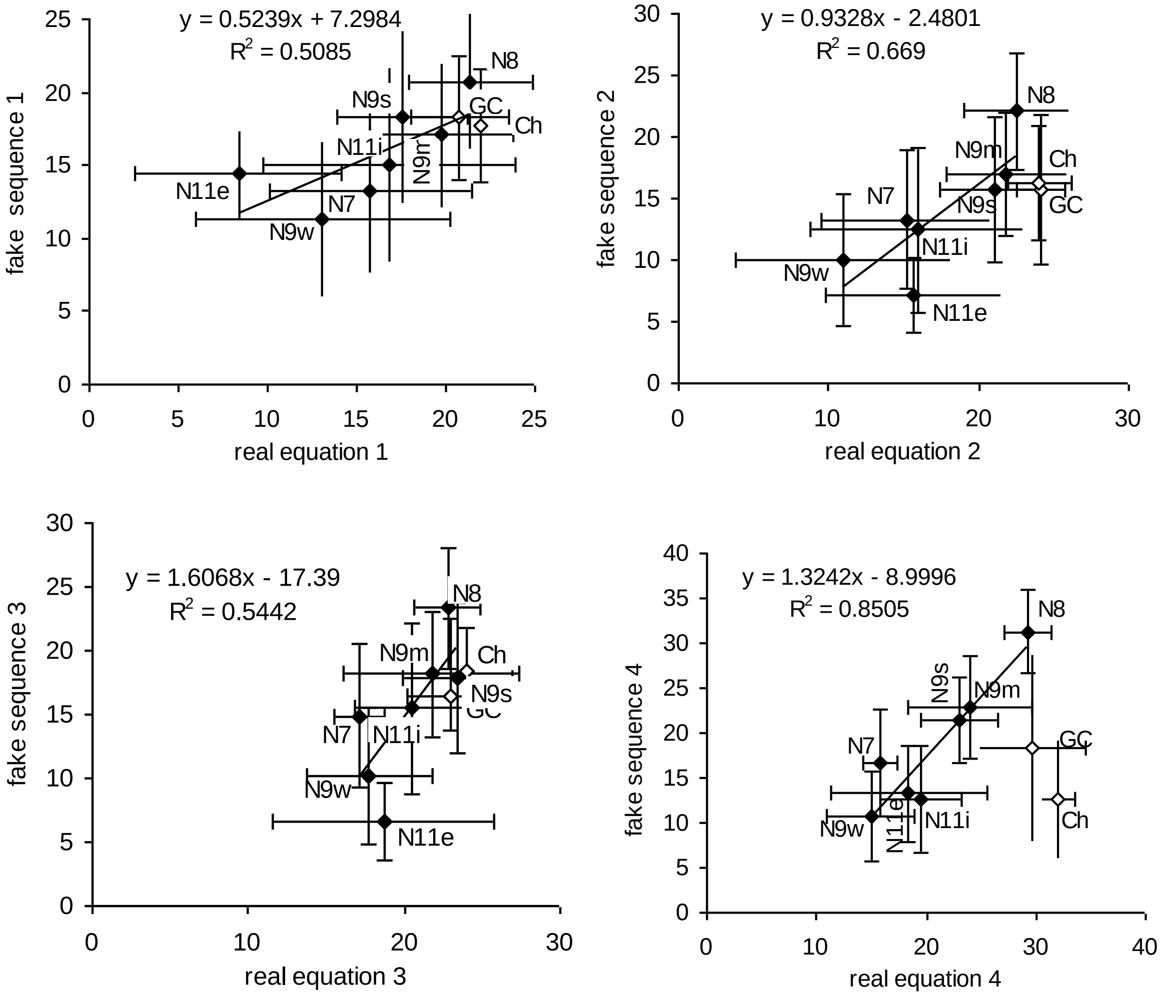

Figure 4. Total amounts of recalled symbols for pairs of real and fake sequences. Error bars denote the standard deviations. Expert groups (GC and Ch) are excluded from the trend line and R^2 calculations.

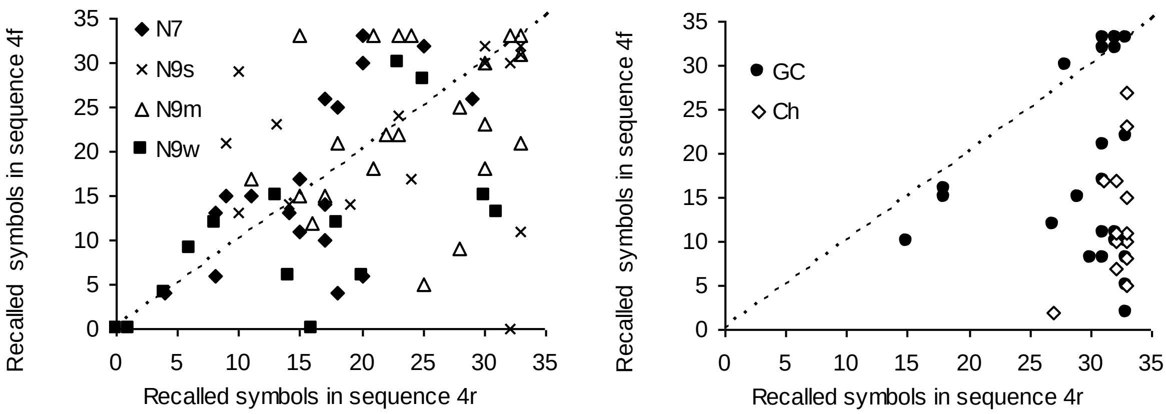

Figure 5. Total amount of recalled symbols for sequences 4r and 4f for participants of four non-expert and two expert groups. The dotted line is y=x.

Symbol-by-symbol recalling

To examine the difference between experts and novices in more detail and to learn something new about the mechanism of memorizing the investigated sequences, we considered how the participants recall equations symbol by symbol. Here we should deal with a group of symbols arranged one after another as a "chain".

Following the frequency of recalled symbols one after another, one can see that for fake sequences, they fit similar patterns irrespective of the group (e.g. Figure 6). Individual differences are very pronounced but upon averaging, the differences are smoothed. Moreover, for *each* sequence and *each* pair of groups, strong (p=0.01) correlation (with only three exceptions out of 144 combinations!) is observed (e.g., Figure 7). The correlation diagrams are usually similar for all of the sequences. This means that the participants utilize a common mechanism for remembering fake sequences.

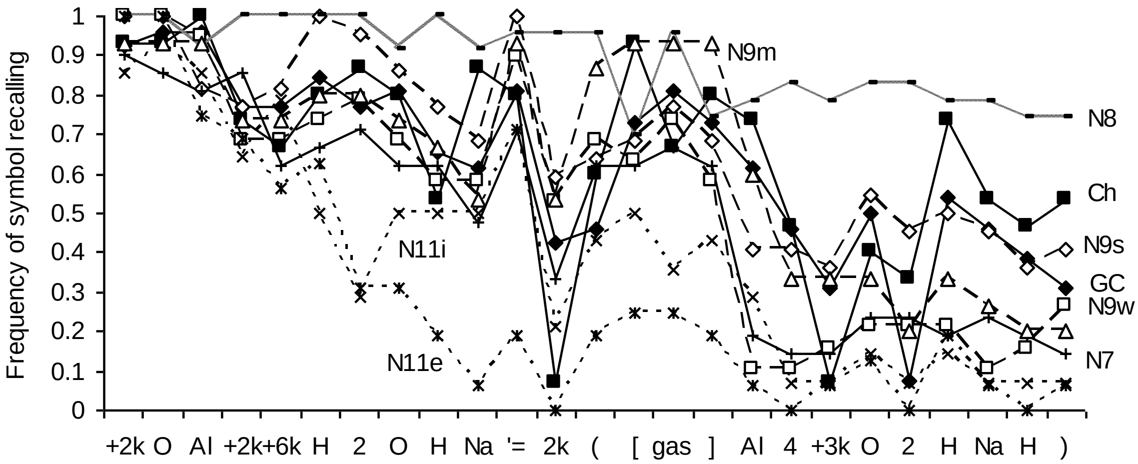

Figure 6. Frequency of symbol recalling in 3f sequence.

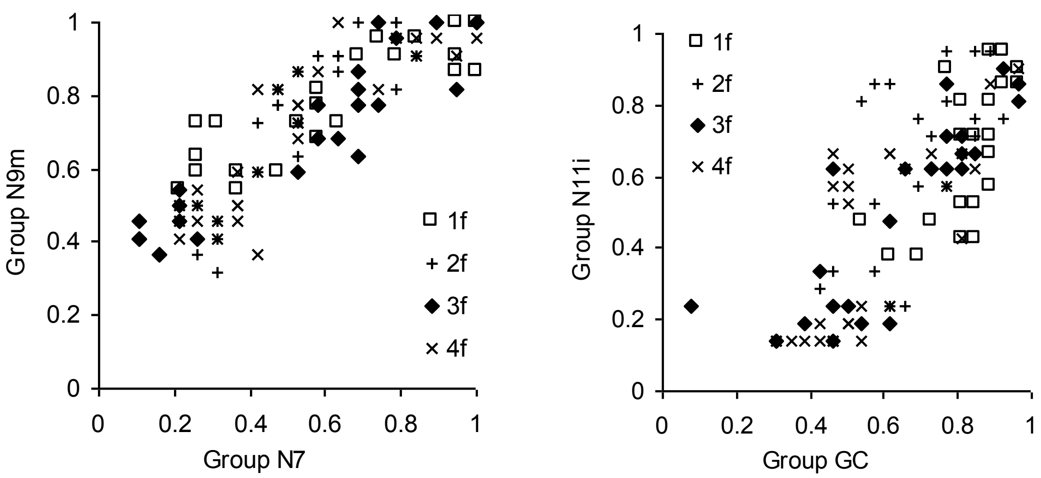

Figure 7. Correlation of frequencies of symbol recalling between two groups for fake sequences.

Even the first glance at Figure 6 reveals that the frequency of symbol recalling depends on its position in a sequence in accordance with a rather complicated pattern that is quite similar for non-expert groups. The similarity of the pattern is confirmed by strong correlations between the frequencies of recalling symbols that occupy the same position in different sequences (e.g., Figure 8). For non-expert groups, the correlation was strong ($p=0.01$) in 36 cases out of 42 (6 pairs of sequences multiplied by 7 groups) and moderate ($p=0.05$) in five cases.

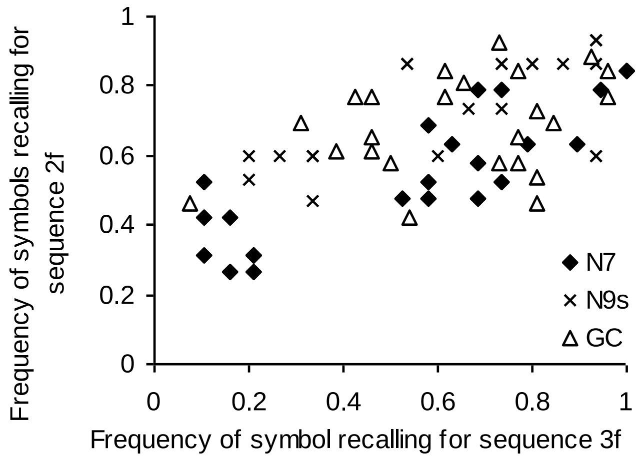

Figure 8. Correlation between frequencies of recalling symbols that occupy the same position for a pair of fake sequences (2f and 3f) for different groups. The last symbol of 3f sequence is omitted.

What kind of pattern is it? Its general features can be seen in Figure 9. The first 2 or 3 symbols are reproduced by almost everybody, then the frequency decreases up to the 6th-7th symbol, and then it fluctuates and decreases once again after about the 16th symbol. This indicates that symbols are generally remembered one-by-one from left to right.

However, there are several exceptions to the proposed model. The first exception is the “=” sign, which is recalled much more frequently than the surrounding symbols. It can be easily explained because there is only one “=” sign in the middle of each sequence, as for real equation. The other exception is a chain “([↑]” in 3f sequence, which is also recalled better. This cannot be explained by chunking it into a sequence of well-known cultural specific acronym (Wickelgren, 1964), because there is no such an acronym in Russian language. Neither it can be explained in terms of the phonological loop mechanism (Baddley, 1999), because it is unpronounceable. This chain could be chunked together only due to its pictorial symmetry, which relates to visual-spatial sketchpad (Baddley, 1999).

As opposed to the non-experts, the frequency of symbol recalling in fake sequences depends not only on the position it occupies. Among the 12 cases (6 pairs of sequences multiplied by 2 groups), we observed only one strong correlation (group GC, sequences 3f and 4f) and four moderate ones. The decrease in the frequency of recalling symbols along with the number of a symbol is not so gradual as for epy non-expert groups. They also recall “=” sign and “([↑]” chain better than the surrounding chains. Besides, there were other chains that were recalled better: “CCN” in 1f; “BaO_2O_2=” and “2_4” in 2f; “H_2O” and “HNa” in 3f; “SHO’ in 4f. This means that experts chunk the chains somehow even if they are random. Among the above-mentioned chains, only the chain “H_2O” is familiar to chemists, which corresponds to the idea of Wickelgren (1964). The chain “BaO_2O_2=” has no chemical sense but it is similar to those that do (for example, BaS_2O_3). The chains “SHO” and “CCN” are easy to pronounce, so their favor could be explained by phonological loop mechanism (Baddley, 1999). However, such chain as “SN_3” at the end of 2f sequence is also easily pronounced but was recalled worse than the following 2_4, which is unpronounceable. In this case, some other mechanism is employed. However, the common nature of this mechanism and the mechanism of chunking real information is under question.

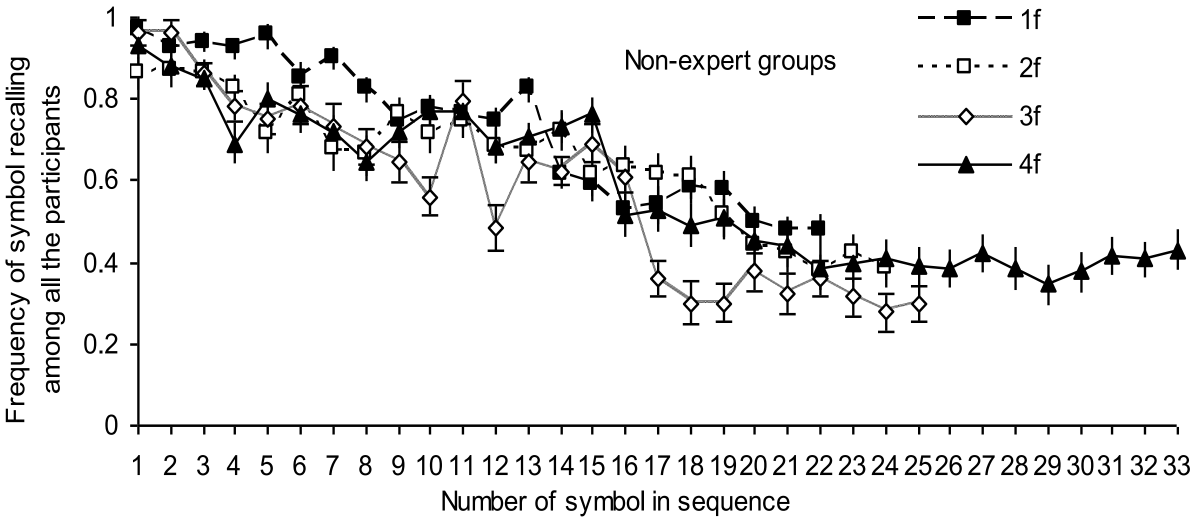

Figure 9. Frequency of symbol recalling among all the non-expert participants in fake sequences. The error bars denote the standard deviations between groups.

For recalling real equations by non-expert groups, the weaker the group is, the more closely the results follow the same pattern. For groups N7, N11i, N11s, all the pairs of sequences show strong or moderate correlations of frequency of recalling symbols depending on their positions in the sequence. For group N9w, no correlation is observed for 2r sequence, which is the easiest one for them. For groups N8 and N9m, strong or moderate correlation is observed in 5 pairs out of 6. And the strongest group, N9, showed significant correlations only for 2 pairs out of 6. This means that the stronger the group, the more pronounced the deviation of the mechanism of remembering real equations from the symbol-by-symbol mechanism. However, the symbol-by-symbol mechanism is still important, because there are strong correlations in the frequency of recalling a symbol occupying the same position between real and fake sequences (25 cases of strong correlation of 28, 4 pairs and 7 groups). Expert groups did not show any correlation of frequencies of recalled symbols occupying the same position for different sequences.

Discussion

The general pattern of recalling of real chemical equations and fake sequences of chemical symbols by experts and novices was the following: there is no significant difference between (a) experts and non-experts in recalling fake sequences and (b) recalling of true and fake sequences by non-experts. However, experts drastically outperform novices in recalling real chemical equations (Figure 2).

This is in line with the findings of Chase & Simon (1973) and Gobert & Simon (1998) for chess players: the difference between experts and novices is excellent in recalling domain-specific information. Now we can spread this conclusion to chemical domains.

Besides, the groups of experts (high performance in recalling real information and low for fake information) and novices (low in both), we should mark out the third group (high in both). In our study it was the group N8. We would like call it “rote learners” (in addition to conventional "experts" and "novices"). The problem of route learners have been arisen in the context of teaching Asian students (Watkins & Ismail, 1994; Saravanamuthu, 2008), but their cognitive abilities and mechanisms are still not investigated.

The correlations (with the trend line close to $y=x$) between the performance in recalling true and fake sequences both for novice groups (Figure 2) and novice individuals (Figure 3) shows, that the mechanism of remembering is the same (and the same that experts use for fake sequences, because their performance in recalling fake sequences does not differ from novices). However, the mechanism of remembering real equations by experts is quite different. Because chemical equations have no meaning for non-experts, we could state that there are different mechanisms in remembering meaningful and meaningless sequences.

The most probable mechanism for remembering meaningless sequence is symbol-by-symbol remembering. Using this mechanism people should remember as many symbols as their working memory capacity allows. Then (a) the frequency of recalling symbols with the same number in a sequence should correlate between different fake sequences and (b) the frequency of symbol recalling should gradually decrease along with the number of a symbol in a sequence.

The frequency of recalling symbols on the same positions shows moderate correlation (Figure 8). Generally the frequency of symbol recalling decreases along with its number in a sequence (Figure 9). This indicates that symbols are generally remembered one-by-one from left to right. When the working memory is overloaded, the remembered chain stops. This is in agreement with the findings of Kellet & Johnstone (1980), who examined the student scripts and showed that most errors are found at the right-hand side of each attempt and the interviewed students explained that they memorized every chemical symbol. However, there are very great variations in recalling different fake sequences for each particular participant with no correlations (except for some participants with good working memory that recall everything). The performance of a particular participant in recalling one sequence can not be predicted based on the results for other three sequences. Evidently, the success in recalling fake sequences depends on various random factors and can be predicted only statistically. Besides, people can chunk together easily pronounced or pictorially symmetric meaningless sequences and thus recall it better. Experts also can chunk together meaningless sequence if they resemble some meaningful (domain-specific) and also recall them better. Therefore, non-experts remember them one-by-one, chunking some of them. Experts also remember them one-by-one; however, they manage to chunk together more chains recognizing them as meaningful. At any rate, individual and sequence-by-sequence variations are rather pronounced and all of the conclusions are statistical rather than individual.

The mechanism of remembering meaningful sequences by experts is quite different. It is not limited by working memory capacity meaning that some other cognitive structures are employed in solving this task. We could suggest that experts in chemistry use chunks – complex units of information formed by previous learning. We don't know the nature of the chunks – it could be, for example, some representations of substances that are stored in the working memory. The number of such chunks for one equation is no more than seven that fits the working memory capacity (Reid, 2008 and refs.). Experts should also possess some rules that link components of chemical equations that relieve their working memory. Thus they can remember only some clue features (for example, the reagent formulas as chunks) and use the rules to restore everything else. However the relative importance of chunks and links between them ("rules") is still not clear and requires further investigations.

Implications and further investigations

The main problems of chemistry are (a) to predict what substances will form in the reaction and (b) what reactants and conditions should be used to obtain the necessary substances. Solution of both problems requires writing chemical equations. Our results show, that generally the working memory capacity is too small to remember the whole chemical equations (except for "route learners") meaning that route learning of chemical equations is useless for the majority of students. It means that while teaching chemistry necessary chunks and rules should be coherently developed. First the elemental concepts should be chunked and put to the long-term memory; then the links between the elemental concepts should be formed to chunk them to more complex ones (forming links between them) that also should be sent to the long-term memory and so on. Thus "context-based", "problem-based" "inquiry-based" and other modern chemistry courses that don't consider concept development (they could be indicated by easy transposition of chapters) should be inadequate (as it was generally shown by Kirschner et. al., 2006). However if they consider concept development, they would work.

However to propose a reasonable order of chunk formation one should understand the nature of chunk in chemistry. This is the matter for further investigations

Conclusions

1. The chunking theory is applicable in the chemistry domain. It requires that to prepare experts in chemistry the curricula should gradually develop chemistry-specific chunks from elemental to complex, following the reasonable load of working memory.

2. The nature of chemistry-related chunks is under question. The further investigations are required to propose reasonable order of chunking development.

3. Both experts and novices remember random sequences from left to right symbol-by-symbol. However experts manage to chunk some chains in the sequence together remembering them better than others. Several mechanisms of the chunking (with no universal one) can be employed: familiar chains; phonological loop; visual-spatial sketchpad. However, the similarity of this mechanism to chunking information in real equations is under question.

4. For further cognitive investigations, we suggest marking out a group of rote learners in addition to expert and novice groups. Rote learners can remember extremely high amount of information, irrespective of its meaning or structure. Their cognitive mechanisms and abilities can differ from those of experts and novices.

Acknowledgements

We kindly acknowledge Dr. Elena F. Kazakova and Dr. Igor V. Morozov (M.V. Lomonosov Moscow State University, Department of Chemistry) for their assistance in the experiment with expert groups and Dr. Zinaida P. Svitanko for helpful commentaries.

References

Abe, M., Hanakawa, T., Takayama, Yo., Kuroki, Ch., Ogawa, S. & Fukuyama, H. (2007) Functional Coupling of Human Prefrontal and Premotor Areas during Cognitive Manipulation. *Journal of Neuroscience*, 27, 3429 –3438.

Akin, O. (1986) *Psychology of architectural design*. London: Pion.

Anderson, J.R. ACT. A Simple Theory of Complex Cognition. (1996) *American Psychologist*, 51, 4, 355-365.

Anderson, J.R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., Qin, Y. (2004) An Integrated Theory of the Mind. *Psychological Review*, 111, 1036 –1060.

Anderson J.R. (2007) *How Can the Human Mind Occur in the Physical Universe?* Oxford University Press.

Baddeley, A.D. (1999). *Essentials of Human Memory*. Psychology Press.

Boucher, L. & Dienes, Z. (2003). Two ways of learning associations. *Cognitive Science*, 27, 807–842.

Brooks, D.W. & Shell, D.F. (2006). Working memory, motivation, and teacher-initiated learning. *Journal of Science Education and Technology*, 15, 17-30.

Campitelli, G., Gobet, F., Head, K., Buckley, M. & Parker, A. (2007). Brain localization of memory chunks in chessplayers. *International Journal of Neuroscience*, 117, 1641–1659.

Chase, W.G., & Simon, H.A. (1973). Perception in chess. *Cognitive Psychology*, 4, 55–81.

Chi, M.T.H., Feltovich, P.J. & Glaser, R. (1981) Categorization and Representation of Physics Problems by Experts and Novices. *Cognitive Science*, 5: 2, 121 — 152.

Cooke, N. J. (1992). Modeling human expertise in expert systems. In R. Hoffman (Ed.) *The Psychology of Expertise, Cognitive Research and Empirical AI* (pp. 29-60). New York. Springer.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87–185.

Damerau F. (1964). A technique for computer detection and correction of spelling errors. *Communications of the ACM*, 7: 3, 1964, 171-176.

Danili, E. & Reid, N. (2004). Some strategies to improve performance in school chemistry, based on two cognitive factors. *Research in Science & Technological Education*, 22, 203-226.

DeGroot, A.D. (1946/2008). *Thought and choice in chess. Amsterdam University Press*. Amsterdam Academic Archive.

Derek, J.M. (2002). The 7±2 Urban Legend. *MISRA C Conference 2002*. http://www.knosof.co.uk/cbook/misart.pdf

Egan, D.E. & Schwartz, B.J. (1979). Chunking in recall of symbolic drawings. *Memory and Cognition*, 7, 149-158.

El-Banna. H. & Johnstone, A.H., (1986). Capacities, demands and processes: a predictive model for science education. *Education in Chemistry*, 23, 80-84.

Gobert, F. & Simon, H.A. (1996). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology*, 31, 1–40.

Gobert, F. & Simon, H.A. (1996b). Chunks in chess memory: Recall of random and distorted positions. *Memory and Cognition*, 24, 493-503.

Gobet, F. & Simon, H.A. (1998). Expert Chess Memory: Revisiting the Chunking Hypothesis. *Memory*, 6, 225-255.

Gobet, F. & Simon, H.A. (2000). Five Seconds or Sixty? Presentation Time in Expert Memory. *Cognitive Science*, 24, 651–682.

Gobet, F., Lane, P.C.R., Croker, S., Cheng, P.C.-H., Jones, G., Oliver, I. & Pine, J.M. (2001). Chunking Mechanisms in Human Learning. *TRENDS in Cognitive Sciences*, 5, 236–243.

Howe, T.V. & Johnstone, A.H. (1971). *Reason or memory? The learning of formulae and equations.* Edinburgh, National Curriculum Development Centre Bulletin 1.

Johnstone, A. H. (2006). Chemical Education Research in Glasgow in Perspective. *Chemistry Education Research and Practice*, 7, 49-63.

Kellett, N.C. & Johnstone, A.H. (1980). Learning Difficulties in School Science – towards a Working Hypothesis. *European Journal of Science Education*, 2, 175-181.

Kirschner P.A., Sweller J, Klark R.E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist*, 41, 75–86.

Klahr, D., Chase, W.G. & Lovelace, E.A. (1983). Structure and Process in Alphabetic Retrieval. *Journal of Experimantal Psychology: Learning, Memory and Cognition*, 9, 462-477.

Koedinger, K. R. & Anderson, J. R. (1990). Abstract Planning and Perceptual Chunks: Elements of Expertise in Geometry. *Cognitive Science*, 14, 511–550.

Levenstein, V.I. (1966) Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Soviet Physics – Doklady*, 10, 707-710.

McNeill, K.L., Lizotte, D.J., Krajcik, J. & Marx, R.W. (2006) Supporting Students' Construction of Scientific Explanations by Fading Scaffolds in Instructional Materials. *The Journal of the Learning Sciences*, 15, 153–191.

McWhinney, B. (2004). A Unified Model of Language Acquisition. In J. Kroll & A. De Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 49-67). Oxford, UK. Oxford University Press.

Miller, J.A. (1956). The Magical Number Seven, Plus or Minus Two. Some Limits on Our Capacity for Processing Information. *Psychological Review*, 101, 343-352.

Reid, N. (2008). A scientific approach to the teaching of chemistry. What do we know about how students learn in the sciences, and how can we make our teaching match this to maximise performance? *Chemistry Education Research and Practice*, 9, 51–59.

Saravanamuthu, K. (2008) Reflecting on the Biggs-Watkins theory of the Chinese Learner. *Critical Perspectives on Accounting*, 19, 138-180.

Servan-Schreiber, E. & Anderson, J.R. (1990). Learning Artificial Grammars with Competitive Chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 592-608.

Sirhan, G. (2007) Learning Difficulties in Chemistry: An Overview. *Journal of Turkish Science Education*, 4, 2-20.

Sweller, J. (2003). Evolution of Human cognitive architecture. In B. Ross (Ed), *The Psychology of learning and motivation*, Vol. 43, (pp. 215-266), San Diego: Academic Press.

Taasoobshirazi, G. & Glynn, S.W. (2009) College Students Solving Chemistry Problems: A Theoretical Model of Expertise. *Journal of Research in Science Teaching*, 46, 1070–1089.

Watkins, D. & Ismail, M. (1994). Is the Asian Learner a Rote Learner? A Malaysian Perspective. *Contemporary Educational Psychology*, 19, 483–488.

Walczak, S. & Fishwick, P. (1997). A quantitative analysis of pattern production and its relationship to expert performance. *Journal of Experimental & Theoretical Artificial Intelligence,* 9, 83-101.

Wickelgren, W.A. (1964). Size of rehearsal group and short-term memory. Journal of Experimental Psychology, 68, 413–419.

Yuan, K., Steedle, J., Shavelson, R., Alonzo, A. & Oppezzo, M. (2006) Working memory, fluid intelligence, and science learning. *Educational Research Review*, 1, 83–98

Zhilin, D. (2010) How much Information do Chemical Equations Content? *Abstract book of 3rd World Conference on Science and Technology Education. Tartu, June, 28 – July 2 2010* (pp. 230-232). http://www.mioo.ru/projects/1119/x3/zhilin-inf_in_chemeq.pdf.