A Compound Remote Associate Test Solver based on Language Data

Ana-Maria OLTEŢEANU [a,1] and Zoe FALOMIR [a]
[a] *Cognitive Systems, Universität Bremen, Germany*

Abstract. The Remote Associate Test (RAT) is generally used for measuring creativity in humans. In this paper, the initial parts of a creative cognitive problem-solving framework are implemented to solve the RAT automatically using a knowledge base of language data extracted from the Corpus of Contemporary American English. The results provided by the computational RAT are compared to those obtained in previous RATs carried out to human participants in order to hyphotesize on an associationist creative cognition paradigm.

Keywords. computational creativity, Remote Associate Test, cognitive systems, knowledge base, language corpus

Introduction

In the quest to achieve human-like artificial intelligence, some of the attributes considered the hardest to replicate are creativity [1] and creative problem-solving. Humans might be biased against attributing creativity to machines [2], possibly because of considering it a highly defining human trait. However creativity is not restrained to the human realm alone, as research on animal tool use [3] and frameworks for studying creativity in animals [4,5] show.

The study of cognitively-inspired computational creativity can benefit both artificial intelligence and cognitive science. For artificial intelligence it can show us the way towards more versatile, flexible and robust artificial agents and artificial cognitive systems. For cognitive science, it can provide us with the better understanding of our own creativity.

Various theoretical proposals have been made on the nature of creativity [6,7,8]. However, the field of creative problem-solving combines the generative powers of creativity with the constraints and evaluative functions implicitly involved in problem-solving, thus providing a well-balanced tool to study computational creativity.

From the cognitive systems perspective, one of the major unsolved questions about creative problem-solving is what kind of knowledge organization and processes endow the human cognitive system with creative abilities. Data on the particular profile such abilities take in human performance (i.e. specific errors, functional fixedness, ability to freely associate, incubation, insight) can be used to try to understand and model such

[1] Corresponding Author: Universität Bremen, Enrique-Schmidt-Str. 5, 28359 Bremen, Germany, E-mail: amoodu@informatik.uni-bremen.de

knowledge organization and processes. Running hypotheses on the type of knowledge organization and processes that are used by humans can be modeled in artificial systems. Thus any artificial system aiming to implement creative problem-solving in a cognitively-inspired manner should be matched against human abilities and performance data.

Various psychological tests have been used to measure creativity; some of those most widely used in the field are the Torrance Tests of Creativity (for a critical review see [9]). Some of these tests reach up to empirically studying insight [10,11]. However, insight problems take a long time to administer to humans, and would need significant amounts of problem-specific common-sense knowledge to be pre-given to the artificial system aiming at replicating the human performance. Thus a good starting point for testing any assumptions about principles of knowledge organization and processes in creative problem-solving would be a smaller task, which enlists abilities similar to insight problem-solving, though providing enough human data for a rich comparison of the performance of the machine. Due to the reasons stated above, the Remote Associate Test (RAT), initially proposed by Mednick [12], is used in this paper as a comparison point.

The following work is focused towards modeling a computational problem-solver that can answer this test (in its general form) in a cognitively inspired manner. The rest of this paper proceeds as follows: the RAT test and the principles of the proposed framework are presented in Section 1. The set-up of the proposed RAT problem-solver is explained in Section 2, together with the knowledge used, the system's knowledge organization and the system flow. Section 3 presents the obtained results, compares them to human normative data [14] and to general plausibility criteria in front of human judgement. A discussion of the results is presented in Section 4, followed by proposals of further work.

1. The Remote Associate Test (RAT)

From the perspective of a cognitive agent, knowledge organization is relevant for problem-solving, since encountering items in a similar context can build associations in the agent's knowledge base that can be further used to: (i) search for a problem solution, (ii) transform the initial objects in the problem, and (iii) formulate or reformulate the problem in a way that makes it solvable for the agent [13]. The aim of the approach presented here is to obtain a computational proof of concept for such a creative-problem-solving paradigm. This requires a system which can use such associations in problem solving, and solve problems in which humans use associations with comparable performance.

The Remote Associate Test [12] is such a test, meant to measure creativity and widely used in the literature [15,16,17]. The RAT takes the following form: given three word items, the participant has to find a fourth term, which is common or can be connected to all of them. For example, the following 3 items are given: *COTTAGE - SWISS - CAKE*; and the participant has to come up with a fourth related term. An answer considered correct in this case according to Mednick's studies[12] is *CHEESE*, because of the following associates: *cottage cheese, swiss cheese and cheese cake*.

Worthen and Clark [18] remarked that Mednick's test [12] includes a mix of structural and functional remote associates - where functional associates ellicit a non-language

relationship (i.e. between *bird* and *egg*) and structural associates triggered items previously associated in the same syntactic structure (i.e. *black* and *magic*). They proposed a remote associate test based on functional associates (FRAT). Different association mechanisms could be responsible for eliciting functional and structural associates and the mentioned creative cognition framework supports both[13]. This paper takes a different approach from [18], to study the relation between structural remote associates and knowledge organization. Thus the performance of the proposed system is compared to compound remote associates (i.e. those obtained from a syntactical compound - like a phrase or a compound noun), which are mainly structural associates, as defined by Bowden and Jung-Beeman [14].

Another reason to use the RAT as a benchmark to test initial mechanisms is its correlation with insight problems, previously demonstrated empirically [19]. Such a correlation might point to similar search processes in both types of problems. Due to the similarity of the manner in which the results are available to the solver - in the form of a pop-up effect, this may imply that the solution is acquired via a mechanism of *implicit* (parallel) memory search, which only makes its results available to the consciousness. In the associative theoretical framework we base our approach on[13], the knowledge organization helps the agent search its own memory in the creative problem-solving process, in order to find relevant but remote information, replace missing objects, see the problem in new solvable ways and propose solutions. Associations are pre-acquired by the knowledge base of the agent due to appearing in close spatial or temporal proximity, or due to sensory-driven organization in similarity-based feature spaces. The *'aha!'* effect happens when a group of implicit associations converges upon a possible solution or a new representation structure (i.e. a different way to see the problem).

The RAT problem-solver mechanism presented here tests part of these assumptions. The system has a knowledge base in which formerly encountered knowledge creates associative links. To find the solution, these links are brought together in an associative problem-solving process, and converge upon a solution. The actual mechanism will become clear in the following description of the RAT problem-solver set-up.

2. The Computational RAT Problem-Solver

The following section describes the knowledge the system uses (Section 2.1), the way in which the system learns and organizes this knowledge -fundamental to the purposes of this paradigm- (Section 2.2), and the system flow which proceeds the moment a query is made (Section 2.3).

2.1. The RAT Knowledge Base

To build the RAT Knowledge Base (RAT-KB) the proposed system is endowed with knowledge from language data, specifically n-grams parsed out of the publicly available, genre-balanced Corpus of Contemporary American English (COCA)[2]. First, the 1 million most frequent 2-grams of this corpus are pruned (based on the part of speech the data is classified as), in order to remove items not relevant for the RAT task. This step uses the tags the data came with, which are indexed according to the UCREL CLAWS7

[2]Corpus of Contemporary American English (COCA): `http://corpus.byu.edu/coca/`

Tagset[3]. As a result, approx. 200 000 items are obtained, those categorized with the tags displayed in Table 1.

Table 1. Tagset used for extraction of items from 2-grams of the Corpus of Contemporary American English.

Tag	Description	Example
FU	unclassified word	
FW	foreign word	*chateau*
JJ	general adjective	*blue*
ND1	singular noun of direction	*north*
NN	common noun, neutral for number	*sheep, cod*
NN1	singular common noun	*book*, *child*
NN2	plural common noun	*books*, *children*
RA	adverb, after nominal head	*else, galore*
REX	adverb introducing appositional constructions	*namely*
RR	general adverb	*down*
RT	quasi-nominal adverb of time	*now, tomorrow*
VB0	be, base form	finite, imperative, subjunctive
VVG -ing	participle of lexical verb	*giving, working*
VVN	past participle of lexical verb	*given, worked*

This dataset is not initially related to any RAT given to humans - that is, there is no a priori evidence on whether useful items for such task where present in the dataset or not.

2.2. Knowledge Aquisition and Organization by Association

The system is presented sequentially with each of the pre-selected 2-grams of the corpus. The system is endowed with three types of atomic knowledge structures: Concepts, Expressions and Links. When a 2-gram is presented to the system, it is registered as an Expression. The system then checks if it is aware of the concepts contained in this expression - any concept that is unknown is added to its Concepts list. A bidirectional Link is attached to each of the Concepts in the presented Expression. After a while, each Concept is thus connected by Links to all the other Concepts it has formed an expression with, thus forming a hub of incoming connections.

Note that some of the RAT queries might refer to compound nouns which will appear unsplit in the corpus. Thus, after all the expressions have been acquired by the system, the system proceeds to compare each Concept with other known Concepts, in order to obtain knowledge about compound words, which are not marked by the tagset in Table 1. If the system recognizes a Concept as *part-of* another Concept, it will then try to match the second part of the now assumed compound (lexical unit) to the other Concepts it knows. If the match is successful, this compound word is also added as an Expression, and a Link is set between its composing lexical units.

This concludes the knowledge acquisition and organization process. Now the system is ready to accept queries.

[3]For a complete list of the UCREL CLAWS7 Tagset see: `http://ucrel.lancs.ac.uk/claws7tags.html`

2.3. *Query System Flow*

Whenever a 3-item query is received, each of the 3 items is activated. Then all the Concepts which are Linked to the first 3 active items are activated. This implies activation of all the Concepts which have been previously observed in an Expression, independent of whether they appeared in the first or second position. Thus the second item in all 2-item Expressions which contain the initial query items become active too. This is intuitively shown in Table 2.

Table 2. Example activation of linked items for query COTTAGE, SWISS, CAKE.

(Cottage + *) OR (*+Cottage)	(Swiss + *) OR (*+Swiss)	(Cake + *) OR (*+Cake)
cottage **cheese**	Swiss Alps	cake batter
cottage garden	Swiss army	cake decorating
cottage industries	Swiss ball	cake flour
cottage ...	Swiss chard	cake layer
... cottage	Swiss **cheese**	carrot cake
... cottage	Swiss chocolate	**cheese** cake

It is worth noting that, every activation proceeds based on the Links of each of the 3 Concepts activated by the query, and no list of corpus Expressions is built anymore.

The system then checks for answers by searching its most activated concepts. Thus for the query illustrated in Table 2, the items $COTTAGE$, $SWISS$ and $CAKE$ have all activated the Concept $CHEESE$. This high activation allows the Concept $CHEESE$ to be considered as a response, on which the activation of the system converges.

Initially, the system is set up to offer as answer the first Concept found with the highest activation. Thus, in the case in which multiple items are activated from the three different concepts, the first one is chosen. As multiple items might be activated from all three concepts, a different answer to that obtained by Mednick's test [12] could be selected. Note that some expressions that compose the correct answer in Mednick's test might not be present in the extraction from the corpus, and other items might associatively arise as the answer. If no 3-item convergence is found, the system will propose the first encountered 2-item on which convergence has happened (a best-next principle of selection).

3. Experimentation and Results

For comparison of performance, the normative data from Bowden and M. Jung-Beeman [14] is used. The results show that out of the 144 items used in Bowden and M. Jung-Beeman's test, 64 are answered correctly[4] by the proposed system. Over 20 of other response items are plausible answers, some of which are displayed in Table 3. Plausible answers refer to responses that a human may deem viable, although they are not the answers considered as correct by Bowden and M. Jung-Beeman's normative data.

[4]Correctness in this case is considered as the exact answer provided by the system on its first try.

Table 3. Some of the plausible answers obtained by the computational RAT.

w_1	w_2	w_3	$Answer_1$ [14]	$\sum_{i=1}^{3} freq(w_i A_1)$	$Answer_2$	$\sum_{i=1}^{3} freq(w_i A_2)$
High	District	House	SCHOOL	47658	STATE	336
Health	Taker	Less	CARE	35632	RISK	1405
Cat	Number	Phone	CALL	10966	HOUSE	239
Chamber	Mask	Natural	GAS	5160	DEATH	272
Self	Attorney	Spending	DEFENSE	3945	BILL	516
Fight	Control	Machine	GUN	3624	POLITICAL	644
Off	Military	First	BASE	1406	PAY	239
French	Car	Shoe	HORN	268	COMPANY	474
Cry	Front	Ship	BATTLE	199	WAR	156
Change	Circuit	Cake	SHORT	179	DESIGN	203
Child	Scan	Wash	BRAIN	171	BODY	178
Mill	Tooth	Dust	SAW	No data	GOLD	No rel
Home	Sea	Bed	SICK	No data	WATER	No rel

Some of these plausible items are more interesting than others, and could be called surprising examples of different "creative response" or remote association convergence from the system. Others items more obviously arise from data regularity, bringing forth associates that are common to all three query items, but hardly interesting from a creative perspective. The system in its current form cannot differentiate between those two. A more rigorous description of the two cases can be that interesting items are items with which the 3 elements in the query form new concepts, while the "regular" items are attributes which are perhaps characteristic of many items (or form with the second element an attribute-concept pair). In this case taking into account the frequencies of such items or their part-of-speech tag might endow the system with the ability to differentiate between surprising and regular plausible answers.

The performance of the system can be also assessed by analyzing the knowledge acquired by the system. For example, an interesting issue is to check in how many cases all three expressions suggested by the query were inside the knowledge base of the system, and how did the system perform in case only two items where present.

The computational approach presented is a general implementation of associative principles and the data we compare to [14] only offers one correct answer. However, the results obtained compared to human data exceeded expectations, bringing strength to the framework's organizational principles. As Table 4 shows, in the cases where the system had all 3 items in its database, the system's accuracy of response compared to human data was at 97.92%, while when the system only knew two of the given expressions, it was finding the correct answers in 30.36% of the cases. Note that this is an extra, since humans are normally assumed to answer correctly the queries for which they know all three items. Although queries where only 2 items were known yielded only a 30.36% performance, this proved that associative principles can add robustness to the the system and help find solutions even in the case where knowledge is lacking.

For the other items in Bowden's data the system simply did not have enough knowledge to respond (for a full illustration of knowledge versus answers see Figure 1).

Table 4. Analysis of the accuracy data provided by the system.

Answer /items known	0 items	1 item	2 items	3 items	Total
Correct	0	0	17	47	64
Plausible	2	11	12	1	26
Not solved	4	23	27	0	54
Total	6	34	56	48	
Accuracy			30.36%	97.92%	

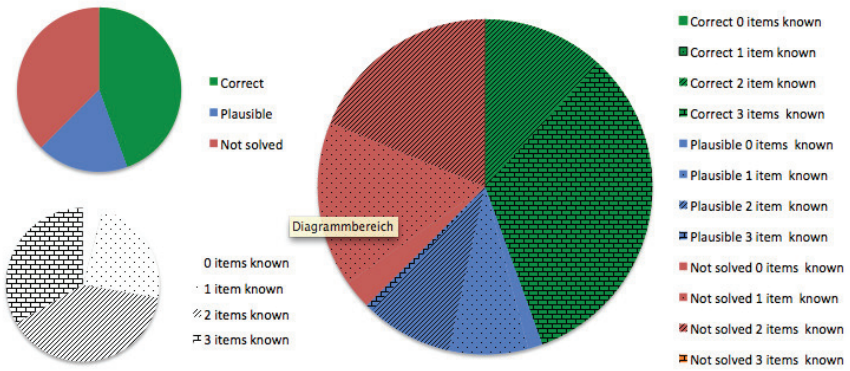

Figure 1. a) Left up - Correct, Plausible and not answered items; b) Left down - Number of known items per query; c) Right - Distribution of knowledge items over items answered

4. Discussion

In this section, the discussion is focused in explaining 3 aspects: (i) which part of the system is meant to represent a cognitive process, and which is an artifact of implementation; (ii) where this implementation fits into the theoretical cognitive framework, and how the system flow is explained from a cognitive perspective related to insight; and (iii) other interesting points ellicited by the performance of the current system.

The set-up of the knowledge base in the proposed system and the stimuli used are symbolic. This is not a commitment that cognitive processing in such cases would take part on a symbolic level alone, but an artifact of this particular implementation. The proof of concept presented in this paper uses lexical symbolic information because of ease of access to both training data (lexical corpuses) and human performance data in the literature (the RAT normative data). As future work, it is intended to introduce other types of sensory information - especially visual and spatial stimuli.

This implementation is a step further towards the automation of the theoretical framework by Olteţeanu [13] - specifically the use of associative links for problem-solving via convergence. The system proposed here models a convergent way of finding an answer, with a pop-up effect of such an answer. Thus, the query elicits the knowledge of the agent, an implicit search over that knowledge base happens, and the answers come up as a result of convergent activation from different initial conceptual points.

This is not problem-solving in its classical form. However, it is a form of search, in which the next possible states are the items associated with the objects offered in the query - which is taken to represent the initial problem state. The problem space becomes the cognitive space of all the associations the agent can find to the initial problem state in its knowledge base. In this case, three different lexical items find the fourth element (answer) because they converge on it by associative power. The initial items can act as initial constraints of the search, because of the way the knowledge base of the system is organised.

What is called here the pop-up or the "aha!" effect is the "sudden" appearance of the solution item (or a possible solution item) in the attention of the agent via a convergence of the associative mechanism. We argue that this is what happens at much higher levels of complexity, and with much deeper constraints, in insight problem-solving.

The computational RAT solver can sometimes find the correct solutions even if it only has two known items in its database. For obvious reasons, it is hard to compare the system's stored knowledge with the knowledge stored in the human counterpart, thus we can't be certain whether when humans don't find the right answer that happens because of lack of knowledge of the required expressions, or because of the difficulty of the associative search over remote items. The ability of the system to find the correct answer even in some of the cases in which only two items are known validates the associative principles used here as proving of interest for creative problem-solving in flexible conditions, with noisy or incomplete information.

However, in the case of knowing only two items, humans have an advantage. After having converged upon a possible answer from two different initial query items, a human can estimate the likelihood that the combination of the possible answer with the third item of the query is a valid semantical construction, or could be a new compound concept, even if that particular human has never encountered that particular concept before. The proposed system obviously doesn't have that constructive ability of matching a potential answer to a potential new concept, in order to estimate the answer's validity.

The RAT problem-solver's performance is satisfactory, although in the current implementation there are:

- No controls used to know whether the knowledge necessary to solve the queries is initially in the corpus or not. Thus the system might easily encounter query items which cannot be linked at all to the desirable solution.
- No frequency data used by the system, though this is available in the initial corpus. One might assume that the frequency with which various 2-grams appear in the corpus can be used as an indicator for how well-known that item is - the strength of its associative link(s) in the memory - which could boost the search. However the system performs quite well without this additional knowledge.

The system can easily come up with different answers in the form of other items associated to all three query items (the first converged upon item is proposed), or not find any correct answers when data for only two of the three items is stored in its knowledge base. The percentage of correct answers found on first convergence proves the associative principles proposed here are worthy of further investigation.

A further point of interest are the generative abilities of the system. A system that can solve RAT problems by organizing its knowledge to recognize associative combinations can also generate RAT problems.

Finally, note that the system was not built to specifically answer this dataset - as mentioned before, its knowledge comparative to the knowledge required by this dataset was calculated post-factum. The system was built in a general enough manner as to be able to attempt to answer any other datasets on tests applied to humans. The only restrictions come from the nature of the associations made by the system (thus structural associates queries will have a higher likelihood to be answered), and from the limitations of the knowledge base.

5. Conclusions and Further work

In this paper, an approach to a creative cognitive problem-solving framework is implemented to solve the Remote Associate Test (RAT) automatically using a knowledge base of language data extracted from the Corpus of Contemporary American English. The RAT is generally used for measuring creativity in humans. The results provided by the implementation of the RAT are compared to those obtained in previous RATs carried out on human participants in order to hyphotesize on an associationist creative cognition paradigm.

The experimental results showed that out of the 144 items used in Bowden and M. Jung-Beeman's test [14], using the COCA corpus, 64 are answered correctly, this is, provided by the system on its first try. Moreover, over 20 of other response items are plausible answers, that is, responses that a human may deem viable. The accuracy of response of the system compared to human data is at 97.92% in the cases where the system had all 3 items in its knowledge base, while 30.36% in the cases when the system only knew two of the given expressions. Humans are normally assumed to answer correctly the queries for which they know all three items, so this proved that associative principles can add robustness to the system and help find solutions even in the case where knowledge is lacking.

The RAT has not yet been studied thoroughly in the literature from a computational knowledge-organization paradigm. This puts us in the position to be able to suggest a large amount of further work. These suggestions can be narrowed to five categories:

1. **Addition of frequency data:** By relating the frequency data of the expressions in the corpus, to the ease with which humans can answer various RAT queries, we might hypothesize the way in which frequency of word and expressions usage influences ease of memory access for such tasks.
2. **Analysing semantic influences:** The frequency of expressions might not be the only item having an impact on the memory of RAT solvers. A second possibly contributing factor is the presence of two or all three of the query items in the same semantic category. Part of our data could be parsed to give some answers to this question in the future.
3. **Comparison to other categories of RAT problems:** Further differences between structural remote associates, functional remote associates as defined by Worthen and Clark [18] and the semantic associates that might result from the use of an ontology also need to be explored.
4. **Computationally building RAT - human performance prediction:** Due to the way it encodes data, the system proposed can solve RAT problems and could be able to reverse-engineer its own process: to thus set-up RAT tests for hu-

mans. Further analysis on what makes RAT problems difficult could help integrate quantitative or qualitative principles based on which the system proposed might categorize different levels of problems. This could be used for prediction of human performance, which would have the potential to falsify and refine the initial assumptions about the cognitive difficulty criteria.

Acknowledgements

This work was conducted in the scope of the project R1-[Image-Space] of the Collaborative Research Center SFB TR 8 Spatial Cognition. Funding by the German Research Foundation (DFG) is gratefully acknowledged by Ana-Maria Olteţeanu.

The support by the European Commission through FP7 Marie Curie IEF actions under project COGNITIVE-AMI (GA 328763), by the Universität Bremen and by the Collaborative Research Center SFB/TR 8 is also acknowledged by Dr.-Ing. Zoe Falomir.

References

[1] S. Colton and G. A. Wiggings, "Computational Creativity: The Final Frontier?", in *Proc. of European Conference on Artificial Intelligence*, pp.21–26, 2012.

[2] A. Eigenfeldt, A. Burenett and P. Pasquier, "Evaluating musical metacreation in a live performance context", in *Proc. of the Third International Conference on Computational Creativity*, pp.140–144, 2012.

[3] W. Köhler, *The Mentality of Apes*, New York: Liveright. (Originally published in 1925), 1976.

[4] J. C. Kaufman and A. B. Kaufman, "Applying a creativity framework to animal cognition", *New Ideas in Psychology*, Elsevier, vol. 22, no.2, pp.143–155, 2004.

[5] A.M. Bailey, W. McDaniel and R.K. Thomas, "Approaches to the study of higher cognitive functions related to creativity in nonhuman animals", *Methods*, Elsevier, vol. 42, no. 1, pp. 3–11, 2007.

[6] J.P. Guilford, *The nature of human intelligence*, McGraw-Hill, New York, 1967.

[7] M.A. Boden, *The Creative Mind: Myths and Mechanisms*, Routledge, 2003.

[8] L. Gabora, "Creative thought as a non Darwinian evolutionary process", *The Journal of Creative Behavior*, vol. 39, no. 4, pp. 262–283, 2005.

[9] K. H. Kim, "Can we trust creativity tests? A review of the Torrance Tests of Creative Thinking (TTCT)", *Creativity Research Journal*, v.18, no. 1, pp.3–14, 2006.

[10] N. RF. Maier, "Reasoning in humans. II. The solution of a problem and its appearance in consciousness", *Journal of Comparative Psychology*, v. 13, no. 2, pp.181,1931.

[11] K. Duncker, "On Problem Solving", *Psychological Monographs*, v. 58, no. 5, whole no. 270,1945.

[12] S. A. Mednick and M. Mednick, *Remote associates test: Examiner's manual*, Houghton Mifflin, 1971.

[13] A.-M. Olteţeanu, "From Simple Machines to Eureka in Four Not-So-Easy Steps.Towards Creative Visuospatial Intelligence", In V.C. Müller (ed.) *Philosophy and Theory of Artificial Intelligence, Synthese Library*, Berlin:Springer, 2014.

[14] E. M. Bowden and M. Jung-Beeman, "Normative data for 144 compound remote associate problems", *Behavior Research Methods, Instruments, & Computers*, vol.3, no. 4, pp. 634–639, 2003.

[15] P. Ansburg, "Individual differences in problem solving via insight", *Current Psychology*, v.19, no.2. pp. 143–146, 2000.

[16] E. M. Bowden and M. Jung-Beeman, "Aha! Insight experience correlates with solution activation in the right hemisphere", *Psychonomic Bulletin & Review*, vol. 10, no. 3, pp. 730–737, 2003.

[17] J. Dorfman, V. A. Shames and J. F. Kihlstrom, "Intuition, incubation, and insight: Implicit cognition in problem solving", *Implicit cognition*, pp. 257–296.

[18] B. R. Worthen and P. M. Clark, "Toward an improved measure of remote associational ability", *Journal of Educational Measurement*, v. 8, no. 2, pp. 113–123,1971.

[19] J.W. Schooler and J. Melcher, "The ineffability of insight", in T.B. Ward and R.A. Finker (eds.) *The creative cognition approach*, Cambridge, The MIT Press, pp. 249–268, 1995.