Reliable multi method assessment of metacognition use in chemistry problem solving

Melanie M. Cooper*[a], Santiago Sandi-Urena[b], Ron Stevens[c]

Received 30th September 2007, Accepted 29th November 2007
First published on the web 28th January 2008
DOI: 10.1039/b801287n

Metacognition is fundamental in achieving understanding of chemistry and developing of problem solving skills. This paper describes an across-method-and-time instrument designed to assess the use of metacognition in chemistry problem solving. This multi method instrument combines a self report, namely the Metacognitive Activities Inventory (MCA-I), with a concurrent automated online instrument, Interactive MultiMedia Exercises (IMMEX). IMMEX presents participants with ill defined problems and collects students' actions as they navigate the problem space. Artificial neural networks and hidden Markov modeling applied to the data collected with IMMEX produce two assessment parameters: the *strategy state*, which is related to the metacognitive qualities of the solution path employed, and the *ability* which is a measure of the problem difficulty students can properly handle. The ability values are significantly correlated with the MCA-I scores, and groups of students who performed using more metacognitive state strategies had significantly higher mean MCA-I values than those using fewer metacognitive strategies. This evidence is indicative of convergence between the methods. This instrument can be used diagnostically to guide the implementation of interventions to promote the use of metacognition; it takes little instructional time, is readily available and allows for the assessment of large cohorts.

Keywords: chemistry problem solving, metacognition, assessment, multi-assesement methods, Metacognitive Activities Inventory (MCA-I), Interactive MultiMedia Exercises (IMMEX)

Introduction

The influence and relevance of metacognition in learning and problem solving has been extensively demonstrated (Veenman et al., 1997; Georghiades, 2000; Pintrich, 2002; Schraw et al. 2005), and the findings suggest that it may even be more important for problem solving success than aptitude (Swanson, 1990). It has also been suggested that it may play a compensatory role for cognitive skills and motivation in the learning of chemistry (Schraw et al., 2005). Despite the numerous definitions encountered in literature, probably the most common description for metacognition is *knowledge and regulation of one's own cognitive system* (Brown, 1987). It may be more easily understood as "*awareness of how one learns; awareness of when one does and does not understand; knowledge of how to use available information to achieve a goal; ability to judge the cognitive demands of a particular task; knowledge of what strategies to use for what purposes; and assessment of one's progress both during and after performance*" (Gourgey, 2001). Metacognition differs from cognition in its being necessary to *understand how a task is performed* whereas cognition is necessary to *simply perform the task* (Schraw, 2001). This crucial characteristic makes the role of metacognition in chemistry learning fundamental to achieve deeper and fruitful understanding (Rickey and Stacey, 2000). In accordance with this argument, Gilbert (2005) has described the use of metacognition in the processes of visualization, which he refers to as "*metavisualization*", as necessary, and asserts the prevalent role of metavisual skills in the learning of science.

There are two main metacognition components generally identified: metacognitive knowledge or knowledge of cognition, and metacognitive skillfulness or regulation of cognition (Davidson, 1995; Schraw and Moshman, 1995). Knowledge of cognition refers to the explicit awareness of the individuals about their cognition, that is: knowing about things (declarative knowledge), knowing how to do things (procedural knowledge) and knowing why and when to do things (conditional knowledge). Regulation of cognition is the executive component that comprises the repertoire of activities used by individuals to control their cognition (Schraw et al., 2006).

College instructors interested in developing problem solving skills through facilitating of metacognition use can benefit from having an adequate assessment instrument to determine changes in the use of metacognitive activities (Rickey and Stacey, 2000). Such an instrument that responds to the current need for reliable ways of measuring this and related constructs has been recently reported (Cooper and Sandi-Urena, 2008). The Metacognitive Activities Inventory, MCA-I, is a self report developed by Cooper and Sandi-Urena

---

[a] *259 Hunter Laboratories, Chemistry Department, Clemson University, Clemson SC 29634,E-mail: cmelani@clemson.edu*
[b] *259 Hunter Laboratories, Chemistry Department, Clemson University, Clemson SC 29634, E-mail: gsandiu@clemson.edu*
[c] *The IMMEX Project, 5601 W. Slauson Ave, Suite 255, Culver City, CA 90230, E-mail: immex_ron@hotmail.com*

This journal is © The Royal Society of Chemistry 2008

that allows for rapid assessment of large numbers of chemistry students at higher education institutions. It can be administered and analyzed easily and rapidly at any time during the instructional cycle.

Like most designs reported for similar purposes, the MCA-I uses a single instrument. In a recent review on the assessment of metacognitive skills, Veenman (2005) stressed the potential of methodologies that use more than one instrument, that is, *multi method designs*, especially those which use different types of instruments administered at different times in relation to the performance of the task, namely *across method and time design*. The same author also suggested that the use of concurrent instruments (administered as the task is performed) is more effective for the assessment of metacognition than those that are prospective (before task performance) or retrospective (after task performance). A handful of reported studies, focused especially on text reading and studying, did use multiple methods but most showed little or no concordance between them (Pintrich, 2002). No attempts using multiple methods have been made to investigate problem solving at the tertiary level.

Concurrent assessment of metacognition in science has been traditionally done by using instruments that are very time consuming and require individual evaluation of participants. Predominant methods, such as think aloud protocols, systematic observations, and analysis of note taking, are very informative for the researcher but not as useful for the practitioner. On the other hand, the array of prospective and retrospective procedures, questionnaires and scales allow a rapid assessment of a large number of participants. However, even if they refer to problem solving, these instruments rely on the recollection of habitual performance or of a recent task and not on the actual deployment of the skills. In addition to reliance on the student's capability of reconstructing and recalling experiences, other issues that present a challenge for self report are the selection of a reference point and social desirability. In these cases, participants' responses may be affected by their own expectations and the perceived expectations of others (Thorndike, 2005). Multi method assessment design presents itself as an effective solution to tackle the shortcomings of using instruments separately.

The primary goal of this research was to develop an assessment of metacognitive skillfulness in college chemistry problem solving that utilizes two instruments: a prospective traditional self report tool, MCA-I, followed by a computer based instrument capable of gathering solution strategy information at the time the student works through the problem. Taken together, these two instruments give insight into both what the students think that they do during problem solving, and also what they actually do as they solve a problem. By using an across method and across time design, construct validity is tested to its limit; convergence between the two instruments would address the potential disadvantages of using a self report and would allow investigation of the problem solving metacognitive activity of large numbers of students.

Instruments

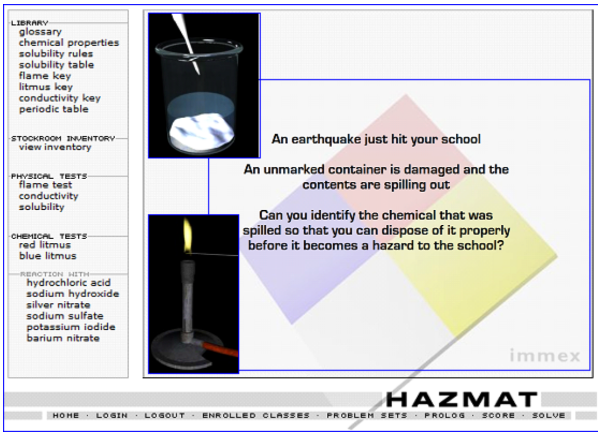
Metacognitive Activities Inventory, MCA-I

The design, validation and characteristics of the MCA-I have been described elsewhere (Cooper and Sandi-Urena, 2008). This 27 item self report instrument assesses students' metacognitive skillfulness when solving chemistry problems and may be used as a diagnostic tool in deciding to implement interventions (Appendix). Respondents select their agreement with the items from a 5 point Likert scale (1, strongly disagree to 5, strongly agree). The score is reported as a percentage of the maximum number of points attainable. Evidence gathered indicates that this inventory is robust, reliable and valid for the intended purpose.

Interactive MultiMedia Exercises, IMMEX

IMMEX is a web based platform that has been described in depth (Underdahl et al., 2001; Stevens and Palacio-Cayetano, 2003; Stevens et al., 2004; Stevens et al., 2005; Cooper et al., 2007) and that has been extensively used in gathering of student performance and problem solving strategy information (Case, 2004; Stevens et al., 2004; Nammouz, 2005; Cox, 2006). Typically, an ill defined problem is presented by using a meaningful real life type scenario. Each problem type, or *problem set*, contains multiple *cases* or *clones*. For research purposes, participants are asked to solve at least five cases of one problem set. Students are able to design their own problem solving strategy as they navigate through the problem space analyzing and processing the information they request. The problem space contains necessary background, as well as information specific to the problem. IMMEX uses an HTML tracking feature to create a record of the items selected, their sequence and the time each item was under consideration. This information can be modeled to partially reconstruct the strategy. Artificial neural networks, ANN, and Hidden Markov Models, HMM, are used to cluster a large number of performances in a predetermined number of strategies, also called *states* (Stevens et al., 2005; Cox, 2006). Evidence indicates that for a given problem type, individuals stabilize on one state after working on five cases (Case, 2004; Stevens et al., 2005; Cox, 2006).

The problem selected for this work, *Hazmat*, is based on inorganic qualitative analysis and has 38 different clones (unknown substances). The prolog for Hazmat is shown in Figure 1. Background or '*library*' items contain information such as a glossary, solubility tables, flame color key, and so forth; whereas information specific to the unknown includes tests that students can request (flame tests, precipitation tests, solubility) and physical properties. When test items are selected, students are presented with a short animation from which they can extract the result of the test. Students have then the possibility of considering their understanding and interpretation of results to continue their navigating of the environment. For instance, if a given test's interpretation solves the identity of the anion, an efficient problem solver will most probably not request more precipitation tests. Students select those items from the problem space that they deem necessary to arrive at a solution. In a training phase,



An earthquake just hit your school

An unmarked container is damaged and the contents are spilling out

Can you identify the chemical that was spilled so that you can dispose of it properly before it becomes a hazard to the school?

Fig. 1 The prologue for Hazmat, an IMMEX qualitative inorganic analysis problem set.

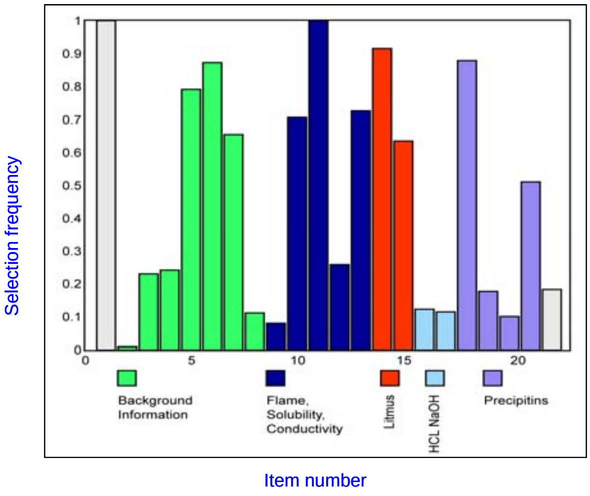


Fig. 2 Sample neural network node.

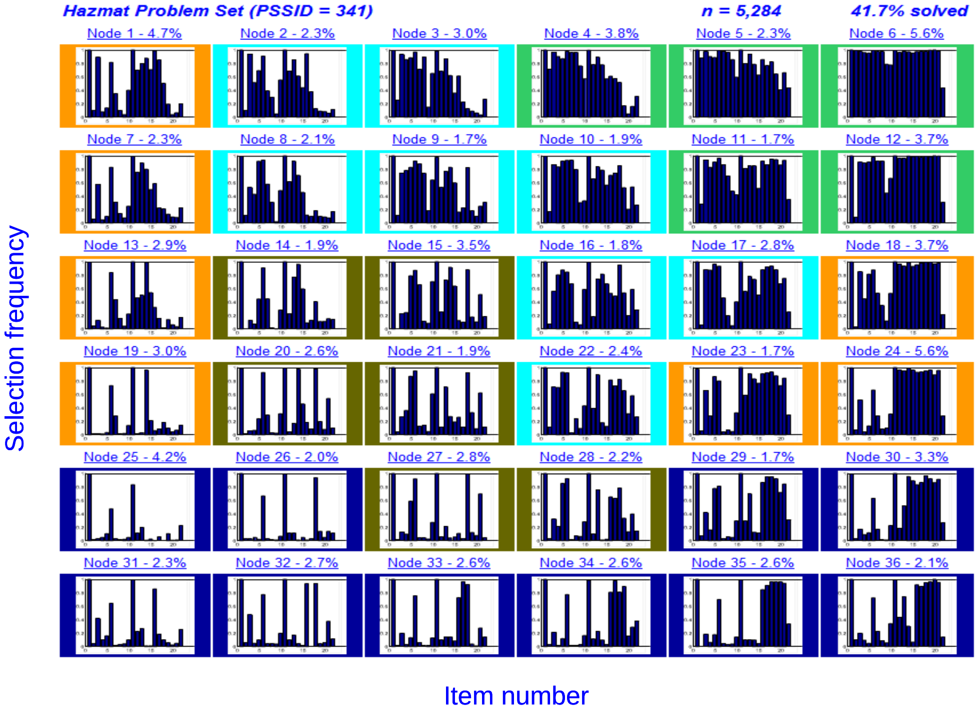


Fig. 3 ANN topological map.

ANNs are fed the problem space items chosen by students (input) in a large number of performances. Based on their pattern recognition ability and self organizing capability, ANNs cluster similar performances in a set number of output nodes which then represent different approaches or strategies employed by the students. These nodes are histograms that describe the probability (vertical axis) of a given item (horizontal axis) to be chosen in a given strategy type. Figure 2 illustrates a single output node obtained from the ANN analysis. For the sake of simplicity, the labels for individual items are omitted and instead types of items are described and color coded. It has been found that a total of 36 nodes are adequate for most IMMEX problem sets (Stevens et al., 2004; Cox, 2006). This analysis produces a topological map, Figure 3, in which geometric distance acts as a metaphor for similarity between strategies. For instance, nodes in the upper right corner of Figure 3 represent strategies where the number of items selected is very high, whereas nodes in the bottom left corner show a much more discerning item selection. Once appropriately trained, the ANNs learn to identify new performances and place them in the node that best fits their strategy.

States are reached through HMM analysis, and can be seen as clusters of nodes that emerge as related strategies. Based on thousands of performances, five states have been identified for Hazmat; Figure 4 shows the ANN nodes related to each of these five states which are also color coded in Figure 3. The probability of individuals to move away from the states

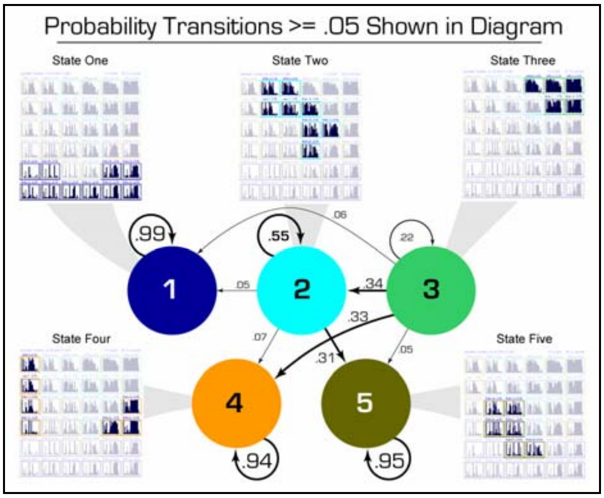


Fig. 4 Hazmat strategy states and nodes associated with them.

(probability of transition) is shown in Figure 4. Individual strategies or nodes can be analyzed in terms of the number of items chosen and their type (for instance, chemical tests, physical properties or library items) and relevance to the case in study. This in-depth analysis of the nodes associated with each state in conjunction with the probability of transition, allow characterization in terms of the implied use of metacognition (Table 1). For example, strategy state 1 represents participants who move rapidly to furnish an answer with little consideration of the background information and without running tests thought to be crucial by experts. Also, there is not noticeable consistency of the items chosen, suggesting random picking of information. Students in this strategy state have a high probability (p=0.99) of remaining in it in subsequent cases, despite the fact that they are informed that their responses are incorrect. This strategy is associated with lack of planning skills, poor ability to sort out items based on their relevance, and poor monitoring and evaluating skills. Therefore, it is characterized as the lowest in metacognition use. At the other extreme, participants in strategy state 5 use an adequate number of items to solve the problem, invariably choose those of high relevance (for example, flame test), consult the background information and remain in this strategy having realized it is effective and efficient (p=0.95).

For the purposes of this work, strategy states 1 and 3 are classified under “low metacognition use, L” (Table 1). These strategies are more prevalent in the solution of the first case attempted, while the students are *framing* the problem. Those participants who do not move away from these states are less metacognitive. States 2 and 4 are “intermediate, I”. In the initial case, these states are not common but are populated later on by students moving specially from state 3 (with a 34% probability to State 2 and 33% probability to State 4 – see Table 1). Careful analysis of the nodes associated with them reveals that the main difference between these two intermediate states is the nature of the information used as indicated by the space items with higher selection frequency. The relative frequency of different types of items can be seen

Table 1 State description, Hazmat

State	Description	Strategy descriptor*
1	Limited, few items used.	L
2	Equal use of background and test items.	I
3	Prolific use of problem space items.	L
4	Many tests, little use of background information.	I
5	Efficient, relatively few items including relevant ones.	H

* L: low; I: intermediate; H: high.

directly from the output node, as exemplified in Figure 2. Strategy 2 uses about the same proportion of tests and library items, whereas strategy 4 is data driven, with less use of background. State 5 is considered “high, H”; as pointed out above, this strategy is the most efficient.

The IMMEX performance data can also be modeled using item response theory, IRT, to obtain a second piece of valuable information: *student ability* (Hambleton et al., 1991). This parameter can be viewed as a measure of the level of case difficulty that a given student can solve. Since not all Hazmat cases are of the same difficulty level (i.e. determining the identity of sodium chloride is considerably easier than solving nitric acid), a simple comparison of correctness might be misleading. Ability calculation considers the different difficulty of the items, hence enabling reliable comparisons of students’ performance; it uses a relative scale where higher values correspond to higher student ability. This parameter allows us to investigate the correlation with state efficiency (Cox, 2006) and MCA-I scores.

Methodology

All participants were students registered in the General Chemistry 1 Laboratory course at a USA southeastern research university, and all signed informed consent forms and were assigned identification numbers. Administration of the Metacognitive Activities Inventory, MCA-I, took place during the first week the laboratory sections met. Hard copies of the instrument were used and responses were entered on optical reader answer sheets. Typically, completion of the instrument took about 15 minutes. Incomplete inventories and those in which a verification item was wrong were discarded. Participants were instructed to solve six cases of the Hazmat problem set the same day of the inventory administration and were given a full week to complete the online assignment. A total of 209 students completed both assessments; all others were excluded from the analysis. Hazmat data were modeled by the IMMEX Project as described previously, thereby obtaining state and ability reports for each participant. SPSS 14.0 was utilized for descriptive statistics of the inventory administration, and to run analysis of variance studies for ability and MCA-I scores by state. The same software package was used to measure the correlation between ability and MCA-I score and to conduct frequency distribution analysis.

Table 2 MCA-I and ability by strategy state (N = 209)

Strategy (N, sample %)	%MCA-I	Ability
Low (45, 21.5)	74.1	43.8
Intermediate (145, 69.4)	75.2	45.5
High (19, 9.1)	80.7	49.3
Mean	75.5	45.5

Table 3 Combination of strategy levels and self reported MCA-I levels

MCA-I group *	State descriptor *		
	L	I	H
L	L L	L I	L H
% within MCA-I	22.6	74.2	3.2
% within State	15.6	15.9	5.3
I	I L	I I	I H
% within MCA-I	23.5	68.6	7.8
% within State	80.0	72.4	63.2
H	H L	H I	H H
% within MCA-I	8.0	68.0	24.0
% within State	4.4	11.7	31.6

* L: low; I: intermediate; H: high.

Results and discussion

Table 2 shows the mean values for the % MCA-I and the ability (IRT) by Hazmat strategy. For both, % MCA-I and ability (IRT), the trend is towards higher mean values for more efficient strategies, with the mean values for the high metacognitive strategy significantly different from the other two groups at the 0.05 level. The MCA-I and the ability (IRT) were significantly correlated at the 0.01 level, although the correlation coefficient is not particularly high (r=0.2).

The results of this study show that there is considerable convergence between the two instruments employed to assess metacognition use by General Chemistry students (Table 2). The three indicators employed, MCA-I score, ability and strategy, are in mutual agreement and in accordance with the expectations derived from the theoretical framework. Students classified as Hazmat low metacognitive strategy users had the lowest MCA-I score and showed the lowest mean ability, whereas students who used the most efficient Hazmat strategies, had statistically significantly higher corresponding measures.

It is important to emphasize that the Hazmat strategy states had been described in the literature previous to this study (Stevens et al., 2004). Even though the magnitude or strength of the relationship between MCA score and the ability is not high, one must remember that the significance of the relationship is as important in the interpretation of the results (Ott and Longnecker, 2001). The significant correlation between the ability and % MCA-I at the 0.01 level supports the convergence of the instruments.

A 2005 review of pivotal importance by Veenman (2005) concluded that "*little or no correspondence between prospective and retrospective statements on the one hand, and actual, concurrent behavior on the other*" was revealed. He pointed out reasons why prospective and retrospective statements may be inadequate (i.e. concerns about the reconstruction and verbalization of skills), but the overriding focus of this paper is the need for multi method research on metacognitive skills as a source of evidence for convergent validity, that is, the agreement between scores on tests intended to assess the same construct (American Educational Research Association, 1999). The report presented here contributes sound evidence in that direction by developing of an across method and across time design for the assessment of metacognitive skillfulness in college chemistry problem solving. Convergence between these instruments reduces the reported shortcomings of self report designs and eliminates the time limitation of traditional concurrent assessments.

Another significant contribution in itself is the use of available technology for the concurrent assessment of metacognition use. IMMEX allows for the collection and recording of strategy information through direct execution of metacognitive skills without interference or disturbance by the researchers. Traditional concurrent assessments usually require of environment that is not naturalistic and participants are aware of being under observation. Using IMMEX, students choose the physical environment and time to work on the problems. Other possible disadvantages of traditional methods that are removed by IMMEX are: verbalization differences, calibration of raters, inter-rater reliability issues, and the bias factor originated from researchers doing the data coding and analysis, since IMMEX performances are modeled in an automated fashion. IMMEX data collection and modeling capability allows for the investigation of hundreds or thousands of students. This potential use makes IMMEX a powerful instrument in the concurrent analysis of metacognition and related constructs.

Although, as we have shown, most students show convergence between self assessed metacognitive activity, and their IMMEX problem solving strategies, there are some cases in which these two parameters do not seem to converge. As important as the cases that demonstrate convergence are, those that do not correlate may even be more important for the designing of specific in-class interventions. In order to conduct distribution analyses, the MCA-I scores are divided into three groups:

- a low or "L group"; those participants below the mean value minus one standard deviation,
- a high or "H group" participants with scores above the mean value plus one standard deviation,
- an intermediate or "I group" composed by those whose score is between these extremes.

Table 3 shows the possible combinations of the strategy descriptors (H, I, L as defined in Table 1) and the self reported metacognition groups H, I, L. The columns in Table 3 correspond to the strategy descriptors, the rows to the metacognition groups, and the cells represent the crosstabulation of frequency. For example, cell labeled "LL" shows that 22.6% of the participants who self reported as low metacognition users performed in the low metacognition strategy group (% within MCA-I). Conversely, 15.6% of the total that performed in the low metacognition strategy group had reported to be low metacognition users (% within State). Top figures across a row add up to 100% (within MCA-I); bottom figures down a column add up to a 100% (within State)

The assignment of MCA-I groups is somewhat arbitrary, and given that the distribution of the scores approaches normality, any choice of cut off points will almost inevitably lead to adjacent values being assigned to different groups. Arranging the data in this array produces nine metacognitive awareness groups which allow the identification of students who are overestimating or underestimating their problem solving abilities. Each metacognitive group is described by two letters, the first one representing the MCA-I group, the second the strategy descriptor. The three top right cells in Table 3 (LI, LH, IH) correspond to overestimates (green), the bottom left cells (IL, HL, HI) to underestimates (yellow) and the groups situated on the diagonal that separates these two (LL, II, HH) are concordant. This representation of the data allows teachers to identify those students whose actions do not correlate with their beliefs about what they are doing. For example, those students who believe they are highly metacognitive but according to their actual performance are not (HL-group), may be more resistant to participation in appropriate interventions than those who are more aware of their limited skills (LL-group).

Students in the HL-group may be more familiar with well-defined problems (where following a sequence of pre-established steps may lead to successful performance), may have a clear strategic understanding but not efficient strategic performance, or may be easily de-motivated. Students falling in the LH-group, those who report low metacognition but were efficient solving problems online, make up a small percentage of the study (0.5%). One could venture that this underestimate of their abilities is caused by using a very rigorous reference point to reply to the inventory which may be consequence of their self-image. The HH-group, students who performed efficiently having previously scored high in the MCA-I, amounts to 31.6% of the high metacognitive performing participants (Table 3). It must be kept in mind that the efficient group itself is only 9% of the total sample (Table 2). It follows then that the HH-group constitutes only about 3% of the total participants. Knowledge of the distribution of students in these concordant and over and underestimation subgroups can assist the practitioner in the designing and implementation of interventions. For instance, decisions can be made upon group composition so that students in the HH-group, who may be high achievers, can be used as peer leaders allowing for their modeling of strategies. Alternative, interventions could be tuned for the different groups and these students could be challenged with more difficult tasks preventing them from stalling in their individual progress and from losing motivation. This analysis of groups does not pretend to be exhaustive but it is an example of the diagnostic power of the multi method instrument.

This paper describes the convergence of two instruments for the assessment of metacognition use in chemistry problem solving. The prospective MCA-inventory consumes very little instructional time while the concurrent assessment (Hazmat) is readily available and easily fits in any General Chemistry curriculum. The access to a reliable, efficient, multi method assessment is of great significance for practitioners. It allows rapid collection of relevant information that informs the implementation of metacognitive interventions tuned to students' metacognitive level.

Work in progress includes the development of instructional interventions to improve student problem solving, use of this multi method to measure changes in student ability, strategy, and use of metacognition, and further investigation of students whose problem solving behavior and self reported activities do not correlate.

Appendix – Metacognitive activities inventory

Statement

1. I read the statement of a problem carefully to fully understand it and determine what the goal is.
2. When I do assigned problems, I try to learn more about the concepts so that I can apply this knowledge to test problems.
3. I sort the information in the statement and determine what is relevant.
4. Once a result is obtained, I check to see that it agrees with what I expected.
5. I try to relate unfamiliar problems with previous situations or problems solved.
6. I try to determine the form in which the answer or product will be expressed.
7. If a problem involves several calculations, I make those calculations separately and check the intermediate results.
8. I clearly identify the goal of a problem (the unknown variable to solve for or the concept to be defined) before attempting a solution.
9. I consider what information needed might not be given in the statement of the problem.
10. I try to double-check everything: my understanding of the problem, calculations, units, etc.
11. I use graphic organizers (diagrams, flow-charts, etc) to better understand problems.
12. I experience moments of insight or creativity while solving problems.
13. I jot down things I know that might help me solve a problem, before attempting a solution.
14. I find important relations amongst the quantities, factors or concepts involved before trying a solution.
15. I make sure that my solution actually answers the question.
16. I plan how to solve a problem before I actually start solving it (even if it is a brief mental plan).
17. I reflect upon things I know that are relevant to a problem.
18. I analyze the steps of my plan and the appropriateness of each step.
19. I attempt to break down the problem to find the starting point.
20. I spend little time on problems for which I do not already have a set of solving rules or that I have not been taught before.
21. When I solve problems, I omit thinking of concepts before attempting a solution.
22. Once I know how to solve a type of problem, I put no more time in understanding the concepts involved.

23. I do not check that the answer makes sense.
24. If I do not know exactly how to solve a problem, I immediately try to guess the answer.
25. I start solving problems without having to read all the details of the statement.
26. I spend little time on problems I am not sure I can solve.
27. When practising, if a problem takes several attempts and I cannot get it right, I get someone to do it for me and I try to memorize the procedure.

References

American Educational Research Association, (1999), *Standards for educational and psychological testing*, American Educational Research Association, Washington, DC.

Brown A., (1987), Metacognition, executive control, self-regulation and other more mysterious mechanisms. In F.E. Weinert and Kluwe, R.H. (Eds.), *Metacognition, motivation and understanding*, Lawrence Erlbaum, Hillsdale, New Jersey, pp. 65-116.

Case E.L., (2004), *The effect of collaborative grouping on student problem solving in first year chemistry*, Doctoral Dissertation, Clemson University, Clemson, South Carolina.

Cooper M.M., Cox C.T. Jr., Nammouz M., Case E. and Stevens R., (2007), An assessment of the effect of collaborative groups on students' problem solving strategies and abilities, *J. Chem. Educ.*, "In Press".

Cooper M.M. and Sandi-Urena S., (2008), Design and validation of an instrument to assess metacognitive skillfulness in chemistry problem solving, *J. Chem. Educ.*, submitted for publication.

Cox C.T., (2006), *An Investigation of the effects of interventions on problem solving strategies and abilities*, Doctoral Dissertation, Clemson University, Clemson, South Carolina.

Davidson J.E., Deuser, R. and Sternberg, R.J., (1995), The role of metacognition in problem solving. In J. Metcalfe and A.P. Shimamura (Eds.), *Metacognition: knowing about knowing*, MIT Press, Cambridge, Massachusetts, pp. 207-226.

Georghiades P., (2000), Beyond conceptual change learning in science education: focusing on transfer, durability and metacognition, *Educ. Res.*, **42**, 119-139.

Gilbert J.K. (2005), Visualization: a metacognitive skill in science and science education. In J.K. Gilbert (Ed.), *Visualization in Science Education*, Springer, The Netherlands, pp. 9-27.

Gourgey A.F., (2001), Developing students' metacognitive knowledge and skills. In H.J. Hartman (Ed.), *Metacognition in learning and instruction: theory, research and practice*, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp.17-32.

Hambleton R.K., Swaminathan H. and Rogers H.J., (1991), *Fundamentals of item response theory (measurement methods for the social science)*, SAGE Publications, Newbury Park, California.

Nammouz M.S., (2005), *A study of the effects that grouping laboratory partners based on logical thinking abilities have on their problem solving strategies in a general chemistry course*, Doctoral Dissertation, Clemson University, Clemson, South Carolina.

Ott R.L. and Longnecker, M., (2001), *An introduction to statistical methods and data analysis*, Duxbury, Pacific Grove, California.

Rickey D. and Stacy, A.M., (2000), The role of metacognition in learning chemistry, *J. Chem. Educ.*, **77**, 915-920.

Pintrich P.R., (2002), The role of metacognitive knowledge in learning, teaching, and assessing, *Theor. Pract.*, **41**, 219-225.

Schraw G. and Moshman. D., (1995), Metacognitive theories, *Educ. Psych. Rev.*, **7**, 351-371.

Schraw G., (2001), Metacognition in basic skills instruction. In H.J. Hartman (Ed.), *Metacognition in learning and instruction: theory, research and practice*, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp.3-16.

Schraw G., Brooks D.W. and Crippen, K.J., (2005), Using an interactive, compensatory model of learning to improve chemistry teaching, *J. Chem. Educ.*, **82**, 637-640.

Schraw G., Crippen K.J. and Kendall H., (2006), Promoting self-regulation in science education: metacognition as part of a broader perspective on learning, *Res. Sci. Educ.*, **36**, 111-139.

Stevens R. and Palacio-Cayetano J., (2003), Design and performance frameworks for constructing problem-solving simulation, *Cell Biol. Educ.*, **2**, 162-178.

Stevens R., Soller A., Cooper M.M. and Sprang M., (2004), Modeling the development of problem solving skills in chemistry with a web-based tutor. In Lecture Notes in Computer Science Series, *Intelligent Tutor Systems*, Springer, Berlin, pp. 580-591.

Stevens R., Johnson D.F. and Soller A., (2005), Probabilities and predictions: modeling the development of scientific problem-solving skills, *Cell Biol. Educ.*, **4**, 42-57.

Swanson H.L., (1990), Influence of metacognitive knowledge and aptitude on problem solving, *J. Educ. Psych.*, **82**, 306-314.

Thorndike R.M., (2005), *Measurement and evaluation in psychology and education*, Pearson Prentice Hall, New Jersey.

Underdahl J., Palacio-Cayetano J. and Stevens R., (2001), Practice makes perfect, *Learn. Lead. Techn.*, **28**, 26-31.

Veenman M.V.J., Elshout J.J. and Meijer J., (1997), The generality vs. domain-specificity of metacognitive skills in novice learning across domains, *Learn. Instr.*, **7**, 187-209.

Veenman M.V.J., (2005), The assessment of metacognitive skills: what can be learned from multi-method designs? In C. Artelt and B. Moschner (Eds.), *Lernstrategien und Metakognition: Implikationen fur Forschung und Praxis*, Waxmann, Berlin, pp. 3-29.

This journal is © The Royal Society of Chemistry 2008