

Factors Underlying Metacognitive Judgements in Foundation Chemistry

Kgadi Mathabathe ^{1*}

¹ University of Pretoria, Groenkloof Campus, SOUTH AFRICA

Received 20 July 2018 • Revised 18 December 2018 • Accepted 21 December 2018

ABSTRACT

Poor performing students have been found to generally report unjustified high levels of confidence in their academic performance. The aim of this study was to elucidate the reasons students provide as justifications for their perceived academic performance in a chemistry test. As part of a larger mixed methods study, students enrolled in a foundation chemistry course ($N = 91$) were required to justify their level of confidence in the accuracy of their responses during a pre- and posttest set on a specific chemistry topic. The qualitative data was analysed for the purposes of the current study and is presented in this paper. The students' responses shed light on the explanations students give as justifications for their confidence judgements and the influence of teaching on these explanations.

Keywords: metacognitive judgements, overconfidence, foundation chemistry, underlying factors

INTRODUCTION

When we are prompted to make judgements on how we perceive our ability or how well we know something or how well we have performed in a particular task, the judgements we report are called metacognitive judgements (Dunlosky, Serra, Matvey and Rawson, 2005; Fernandez-Duque & Black, 2007; Koriat & Bjork, 2005; Rosenthal, 2000). Metacognitive judgements are extensively investigated partially due to the fact that mastering the skill of accurately making them, may result in the effective regulation of self-paced study which is necessary in a university environment (Dunlosky et al., 2005). Nelson and Narens (1990) identified several types of metacognitive judgements namely ease-of-learning judgements (EOL), judgements of knowing or judgements of learning (JOL), feeling-of-knowing judgements (FOK) and confidence judgements. The current study is part of a larger study focusing on metacognitive judgements in an academic development programme, particularly confidence judgements about mastery of a difficult chemistry topic (Mathabathe & Potgieter, 2014).

An academic development programme is an intervention by a tertiary institution to address under-preparedness of incoming students for the mainstream programmes offered by that institution. The context of the current study was an academic development programme offered with the goal of ultimately widening access to mathematics and science-related careers (Potgieter, Dawidowitz, & Mathabatha, 2007). South Africa has a school system that largely produces students who are under-prepared for tertiary science studies. Teaching in the academic development programmes is specialised to give students the necessary support to address under-preparedness. However, many of the students in these programmes have been found to report unjustified exaggerated confidence judgements in their mastery of foundational chemistry concepts and skills (Potgieter et al., 2007). Some scholars have argued that overly optimistic confidence judgements may be necessary to motivate students to tackle the more difficult tasks but some have argued that not knowing what they do not know may put students at risk of failure by negatively influencing decisions regarding study time management and self-regulated learning (Ehrlinger, 2008; Mathabathe & Potgieter, 2014). A lot has been reported about the tendency of students to make inaccurate judgements about their performance in various domains (Beyer, 1999; Ehrlinger, 2008; Nowell & Alston, 2007; Schaefer, Williams, Goodie, & Campbell, 2004).

Contribution of this paper to the literature

- This study contributes to the growing literature on bias in self-evaluation in teaching and learning contexts.
- While many studies reported from the perspective of educational psychologists have focused on quantifying the level of bias in self-evaluation this study contributes to the field of chemistry education by reporting on the impact of tertiary teaching and feedback on the quality of judgements made by students to justify their academic performance.
- Few studies have focused on determining the reasons given by participants as justification for their judgements of academic performance and how these are influenced by teaching.

THE LARGER STUDY

The larger study was conducted to determine how students in an academic development programme evaluate their competence to solve chemistry problems and to determine the impact of tertiary teaching and feedback on the quality of judgements made. The larger study took place in the second semester of a South African university's academic development programme. The study received ethical clearance from the institution where data was collected. The purposively selected sample consisted of 91 multiracial participants who voluntarily participated in the study and for whom complete records were obtained, i.e. 35 male students (38%) and 55 female students (60%) and one record with gender information omitted. The ages of participants ranged between 17 - 25 years ($M = 19$). Pretest data were collected before a 3-week period of instruction on stoichiometry and posttest data after instruction and ample opportunity for guided and unguided problem-solving. The students shared commonalities in that they were all taught stoichiometry by the same lecturer in the large group lecture setting. The students received further support by way of small group tutorial sessions with plenty of problem-solving and constant feedback opportunities. Even though they had different instructors with different teaching styles in the tutorial sessions, the instructors worked collaboratively in terms of the material used as well the quantity, format and content of tests and tasks given to the students.

Stoichiometry was chosen on the basis that it was part of the chemistry curriculum for the programme and it had been identified to be amongst the topics that pose the most difficulty for first year chemistry students (Huddle & Pillay, 1996; Potgieter et al., 2007). A decision to give a test on a difficult topic prior and post instruction was deemed appropriate to deliberately investigate bias in performance evaluation before and after instruction. Pretest and posttest data were collected through a 19-item test instrument with acceptable reliability (Mathabathe & Potgieter, 2014). Each item in the instrument consisted of three tiers to enable the simultaneous collection of both quantitative and qualitative data. For each item the first tier consisted of a multiple choice question on stoichiometry followed by a Likert scale from 0% to 100% with 10% intervals on which participants were asked to indicate their confidence in the accuracy of the chosen response in the first tier. Lastly, in the third tier participants were asked to explain their choice of confidence indicators, this constituted qualitative data reported in this paper. For a student who was 80% certain that the answer was correct, we assumed that he/she expected that there was an 80% likelihood of the answer actually being correct. The average of all of the confidence ratings that were chosen by a student gave a good indication of expected or perceived performance in the test. To determine accuracy of performance evaluation, actual performance (test scores out of 19 converted to a percentage) were subtracted from the perceived performance (percentage average confidence scores).

Taking into consideration the difficulty of the topic, the level of preparedness of the students in our sample and the format of the test (multiple-choice) an acceptable margin of error was determined and set as the equivalent of three out of 19 answers judged incorrectly which translates into a judgement error of 15.8%. Subjects whose average confidence scores exceeded the test scores by more than 15.8% were labelled as overconfident (OC). The realistic group (R) were subjects whose average confidence scores were between 15.8% and -15.8% (-15.8% and 15.8% included). Subjects whose test scores exceeded their average confidence scores by more than 15.8% were labelled as under-confident (UC). Using these criteria the majority of students were found to be overly optimistic about their performance in both the pretest (69%) and the posttest (71%). With only 31% and 26% being realistic about their performance in the pretest and posttest respectively and only two students emerging as underconfident in the posttest.

The global view of results obscured the finer details about students who were able to show an improvement in terms of accuracy in performance evaluation after instruction and those whose ability to do so deteriorated. Five performance evaluation subgroups were defined on the basis of accuracy of evaluation in the pre- and posttests. The five subgroups were labelled first by their pretest and then their posttest category. A large percentage of the students showed no improvement in their accuracy of performance evaluation (OC-OC subgroup, 55%). The number of students who acquired the skill of reporting accurate self-evaluations of their performance (OC-R

subgroup, 14%) was similar to the number of students who reported overly positive self-evaluations of their performance upon exposure to teaching (R-OC subgroup, 16%). Only two of the 28 students who were realistic in their pre-test performance evaluation became underconfident in the posttest (R-UC subgroup, 2%). The R-UC subgroup was deemed too small for meaningful inferences to be made and was therefore excluded in the comparative analysis of performance evaluation subgroups.

Important to note from the findings of the larger study and relevant for the purpose of the current study is that although the four subgroups could be assumed to be comparable in terms of prior knowledge in stoichiometry the students in the four subgroups differed significantly in terms of learning gain (Hake, 1998). Students in the R-R and OC-R subgroups distinguished themselves by demonstrating the quality of being able to gain from the teaching and learning process and realistically evaluate their performance in the posttest. The OC-OC subgroup never mastered the metacognitive skill of accurate performance evaluation, yet they demonstrated moderate learning gain while the R-OC subgroup did not gain anything from the teaching and learning experience yet became excessively confident in their mastery of stoichiometry. It became necessary therefore to understand how students, who were consistently overconfident or became overconfident in the posttest, differed from the realistic students in terms of the factors they cited when motivating their choice of confidence ratings. We also wanted to explore the influence of teaching on these factors. The research literature was consulted in order to formulate a theoretical framework that offers an explanation for the factors students rely on when constructing metacognitive judgements.

THE CURRENT STUDY: FACTORS STUDENTS RELY ON WHEN MAKING METACOGNITIVE JUDGEMENTS

Modern education research was introduced to the construct of metacognitive experiences through the publication of Nelson and Naren (1990)'s theory of monitoring and control. The theory focused on the interaction between metacognitive monitoring and control. Metacognitive monitoring entailed processes that enabled individuals to observe and reflect on, their own cognitive processes (Flavell, 1979) whereas metacognitive control could be observed in the decisions and judgements individuals consciously or unconsciously made based on the outcome of their monitoring. Monitoring is revealed by asking participants to make judgements about their memory, knowledge, learning or comprehension. Control on the other hand is revealed by the actions an individual engages in as a result of the monitoring, for example decisions about which items to study and the amount of time allocated to study (Schwartz & Perfect, 2002). Accurate monitoring is therefore necessary because it informs the decisions taken for corrective measures. A learner who possesses metacognitive abilities should then be able to plan and select relevant strategies, monitor the progress of learning, make metacognitive judgements about progress towards desired outcomes, correct errors, evaluate the effectiveness of learning strategies and change strategies and learning behaviours when necessary (Ridley, Schutz, Glanz, & Weinstein, 1992). However, when faced with solving a problem, novice learners often times do not try to examine a problem in depth; they are satisfied just to scratch the surface. They do not try to reflect on the reasons for their success or failure and use this feedback to improve performance.

The theoretical framework of Nelson and Narens (1990) gives an overview of different factors that guide the monitoring and control processes that occur when a student studies for an upcoming examination and when a student retrieves information during an examination. The framework consists of three stages, namely the acquisition, retention and retrieval stages. The acquisition stage takes place prior to studying for the examination. The retention stage occurs when a student is busy studying for the test and the retrieval stage is when the student is taking the test and information is being retrieved. The current study focuses on factors students refer to or rely on when making confidence judgements during the retrieval stage, i.e. test-taking.

Consistent with Nelson and Narens (1990)'s framework we propose that the metacognitive monitoring processes students may experience while taking a test are as follows: a question asked in the test elicits either a feeling of knowing (FOK) or feeling of not knowing (FOnK) (Jing, Kazuhisa, & Yuejia, 2003; Winnie & Nesbit, 2010). When an FOK is elicited the students may claim a judgement of knowing (JOK), proceed to search their long-term memory (LTM) for an answer, report the answer as correct and make a positive confidence judgement based on recalled information or alternatively the student may give an answer and positive confidence judgement based on how they feel (FOK). Failed attempts to retrieve a potential answer from memory should lead to no output. Lack of or negative confidence in the accuracy of response should be based on reasons that a tip-of-tongue experience is not elicited by the question or information given in the question or lack of knowledge. A JOK made without consulting information in the LTM is based on merely a subjective conviction while a JOK made after searching LTM should be substantiated with objective evidence such as possession or demonstration of task-related declarative, procedural or conditional knowledge. JOKs based on feelings or incorrect information may lead to incorrect outputs. Positive confidence judgements indicated for incorrect answers reported based on perceived JOK or subjective conviction may result in bias in performance evaluation.

Nussinson and Koriat (2008) make a distinction between two factors people may rely on when making metacognitive judgements. When making metacognitive judgements people may rely on the knowledge or information they retrieve from their LTM (information-based judgements) or on subjective feelings (experience-based judgements). People, however often make metacognitive judgements based on subjective feelings rather than on the information they retrieve from memory (Koriat, 1993; Koriat & Bjork, 2006; Nussinson & Koriat, 2008). In their study of calibration accuracy of college students in classroom contexts, Hacker, Bol and Bahbahani (2008: 101) observed “differences by performance level in open-ended explanations for calibration judgements”. Higher-performing students compared to lower-performing students relied less on subjective factors when making metacognitive judgements of performance.

Investigating the nature of justifications students make to substantiate their confidence judgements is important because if the judgements are substantiated with objective evidence this gives an indication that students know what they know and what they do not know. This is necessary if they are to identify areas for improvement. However, if judgements are substantiated with vague evidence this may be an indication of the Kruger-Dunning effect (Kruger & Dunning, 1999) that students may be unskilled and unaware. Thus not able to draw from the experience, the information necessary to improve their future performance. In the current study we set out to determine factors students allude to when justifying confidence judgements. The findings reported in this paper represent our attempt to answer the following research questions:

- What are the factors that students in the performance evaluation subgroups allude to rely on when making performance evaluations?
- What shifts, in terms of reliance on these factors, are observed, for each performance evaluation subgroup, after the teaching of stoichiometry?

METHODS

Research Design

An embedded experimental mixed methods approach (Harrits, 2011; Leech & Onwuegbuzie, 2011; Maxwell & Loomis, 2003) was followed for the larger study. Quantitative data was required to measure performance and to determine accuracy of performance evaluation. Qualitative data presented in this paper, was collected to enrich our understanding of the factors underlying metacognitive judgements (Nelson & Narens, 1990).

Analysis of Free-Response Explanations

Students were required to explain their choice of confidence judgement rating in each third tier of the stoichiometry test instrument. Providing students with this opportunity enabled us to collect data which could assist in shedding light on the factors informing metacognitive judgements of performance. The fact that one subgroup achieved a learning gain significantly higher than the rest (OC-R) and one subgroup did not achieve any gain at all (R-OC) highlighted the need to probe further for information that could explain the strengths and deficiencies in metacognitive skills that gave rise to such a difference. The qualitative data proved to be a rich resource for this purpose.

A software package called ATLAS.ti was used to systematically organise the students' free responses according to their pre-post performance evaluation subgroups. Following thematic analysis representations of patterned responses or meaning were inductively identified within the data set (Braun & Clarke, 2006). Four hermeneutic units were created, namely OC-OC, OC-R, R-R and R-OC. This was done to make it possible to separately analyse the students' responses according to their pre-post performance evaluation subgroups.

Qualitative data analysis entailed reading, re-reading and systematically coding responses in each hermeneutic unit. To avoid multiple assignments of codes to the same response, each response had to be carefully read to find the message it conveyed and coded accordingly. To reduce the codes into a reasonable number of categories, the researchers first read through the codes to find themes that were recurring in the data. Related codes were combined into themes. Each theme was assigned an identifying name using descriptive words from the text to establish a category. In this way categories were identified and allowed to emerge from the codes. For greater conceptual clarity categories that were linked to the same concept were grouped together to create super-categories. The researchers clustered the responses by themes and created a total of twenty-seven categories which were further grouped into eight super-categories. **Table 1** shows the categories and super-categories that were generated from the emerging codes. In order to demonstrate the nuances of statements in each super-category examples of typical responses are provided for each category. Two additional super-categories were added resulting in ten super-categories. Super-category nine (SC 9) labelled 'No explanations', covered all the entries where students omitted explanations for their choice of confidence judgement ratings. Super-category ten (SC 10) labelled as

Table 1. Super-categories and categories generated from emerging codes

| Super category (SC) | Categories (C) |
|-----------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| DECLARATIVE KNOWLEDGE (SC1) | C7: What the person thinks is required to solve a problem. E.g. <i>'used conversion factor method'</i> C9: Uncertainty due to deficit in general declarative knowledge. E.g. <i>'Not sure of theory'</i> C19: Perceived possession of declarative knowledge required to solve the problem. E.g. <i>'know theory'</i> |
| PROCEDURAL KNOWLEDGE (SC2) | C8: Uncertainty due to deficit in general procedural knowledge (approach and method). E.g. <i>'not sure how to work out'</i> C15: Certainty based on the perceived possession of specific procedural knowledge that can be named and demonstrated – (balancing/calculations). E.g. <i>'sure of calculation' or student has shown how answer has been worked out or provided a balanced equation.</i> C22: Lack of specific procedural knowledge. E.g. <i>'don't know how to intermingle particles'</i> C23: Perceived possession of relevant skills to solve the problem. E.g. <i>'Can calculate limiting reactants'</i> C27: Perceived possession of test-taking skills. E.g. <i>'used process of elimination'</i> |
| GLOBAL EVALUATION OF ANSWER (SC3) | C1: Estimation of chance. E.g. <i>'50% chance of getting answer right'</i> C2: Subjective feeling of doubt or uncertainty. E.g. <i>'Doubtful'</i> C3: Informed guess. E.g. <i>'Guess'</i> C5: Lack of confidence. E.g. <i>'don't feel confident'</i> C6: Vague judgement of how answer looks in comparison with given responses. E.g. <i>'it's right'</i> C10: Certainty due to a subjective, vague feeling not based on any evidence. E.g. <i>'sure of answer'</i> C11: Certainty due to a subjective feeling based on perceived correctness of answer. E.g. <i>'believe answer is correct'</i> C12: Estimation of confidence in approach or answer. E.g. <i>'confident with work'</i> C13: How own answer compares with given multiple choice options. E.g. <i>'closest value to own answer'</i> C14: Certainty due to an appeal that answer makes sense or is logical. E.g. <i>'Answer makes sense'</i> C17: Unreflective, almost defensive evaluation. E.g. <i>'just know'</i> C18: Global feeling that answer is right. E.g. <i>'this is how I feel'</i> |
| FEELING OF UNPREPAREDNESS (SC4) | C4: Feeling of general inadequacy and unpreparedness. E.g. <i>'haven't studied yet'</i> |
| EXTERNAL FACTORS (SC5) | C20: Deficits of the question. E.g. <i>'vague question'</i> C24: Familiarity with question or concept. E.g. <i>'done in high school'</i> C25: Unfamiliarity with question or concept. E.g. <i>'never seen such question before'</i> |
| METACOGNITIVE STRATEGIES (SC6) | C16: Metacognitive strategy. E.g. <i>'double checked solution'</i> |
| MATH SKILLS (SC7) | C21: Uncertainty due to inadequate mathematical skills. E.g. <i>'No idea how to calculate'</i> |
| LACK OF MEMORY (SC8) | C26: Lack of memory. E.g. <i>'Forgot how to work out'</i> |

'Others' represented all the responses that could not be coded such as incomplete sentences, incomplete equations etc. To ensure reliability of the method of coding and the validity of the designated codes, the researcher and an independent coder coded the responses to the open-ended third tier items and solved their coding discrepancies by mutual agreement. The inter-rater reliability reached 0.83 in the pretest and 0.91 in the posttest. The two kappa values fall within a range of values indicative of strong agreement (Fleiss, 1981).

RESULTS

While data is being coded, ATLAS.ti counts the frequency with which the codes occur. For each performance evaluation subgroup, the number of times a category emerged was first determined. These numbers were converted into percentage values. The frequencies of occurrence of responses in the categories belonging to a single super-category were added together to represent the frequency of occurrence of responses in that super-category. During the analysis of qualitative data it emerged that some categories had very low incidences of occurrence and therefore in deciding whether the frequency with which a specific category or super-category was cited was significant or not, an arbitrary cut-off value of 5% was adopted. Based on this criterion, the super-categories labelled Declarative knowledge (SC1), Procedural knowledge (SC2), Global evaluation of answer (SC3), and External factors (SC5) emerged as major super-categories. The incidences of occurrence of super-category 4 (The feeling of unpreparedness) were of marginal prominence while those of Metacognitive strategies (SC6), Mathematics skills (SC7) and Lack of memory (SC8) super-categories showed incidences of occurrence which were too scarce to be

Table 2. Frequencies of occurrence of super-categories of performance evaluation subgroups in the pre- and posttest

| Super-categories (SC) | OC-OC (50) | | OC-R (13) | | R-OC (15) | | R-R (11) | |
|----------------------------------------|------------|----------|-----------|----------|-----------|----------|----------|----------|
| | Pre (%) | Post (%) | Pre (%) | Post (%) | Pre (%) | Post (%) | Pre (%) | Post (%) |
| SC1 Declarative knowledge | 5.1 | 7.9 | 4.5 | 5.7 | 6.0 | 6.0 | 5.3 | 11.0 |
| SC2 Procedural knowledge | 40.2 | 34.6 | 44.1 | 34.4 | 37.2 | 27.4 | 43.1 | 35.9 |
| SC3 Global evaluation of answer | 33.8 | 38.7 | 35.6 | 31.6 | 34.4 | 43.9 | 38.3 | 32.1 |
| SC4 Feeling of unpreparedness | 1.1 | 2.5 | 2.8 | 6.1 | 3.5 | 2.5 | 2.0 | 3.8 |
| SC5 External factors | 7.6 | 7.8 | 10.1 | 17.4 | 10.5 | 9.5 | 3.8 | 6.2 |

considered in our discussion of qualitative data. In **Table 2** all percentage values for response prevalence are rounded off to one decimal place to simplify comparisons.

Stoichiometry is a very mathematical part of chemistry, dealing with calculations of masses and sometimes volumes of reactants and products involved in a chemical reaction. In stoichiometry students are not only required to demonstrate understanding of chemical reactions, but they must also be able to apply a thorough understanding of the principles involved in ratio and proportion calculations (Ben-Zvi, Eylon, & Silberstein, 1988; Huddle & Pillay, 1996). Solving stoichiometry problems usually requires predominantly procedural knowledge. Stoichiometry has a very minimal component of memory and recall problems. It is for this specific topic necessary to go beyond recall of declarative knowledge. In fact it was not surprising to find the prevalence of responses in the SC1 (Declarative knowledge) to be so small across all subgroups in both the pre- and posttest. The prevalence of this super-category stayed at a level of 5 or 7% and only for the R-R subgroup did it ever attain a level of 11%.

How students in the four subgroups justified their choice of confidence indicators in the pre- and posttest

The discussion that follows will focus only on responses with a prevalence of 5.0% or higher. In the interpretation of frequency data it will be helpful to bear in mind that some super-categories clearly reflect a subjective judgement (SC3 and SC4); some are rationally motivated (SC1 and SC2), whereas a single super-category seems to be based on external factors such as feelings of not knowing due to unfamiliarity with the question or the features of the question itself and cognitive feelings (SC5).

OC-OC Subgroup

The results for the OC-OC subgroup indicate that judgements made on the basis of declarative knowledge (SC1) and those made based on external factors (SC5) increased marginally in the posttest (SC1 – Pre: 5.1%, Post: 7.9%; SC5 – Pre: 7.6%, Post: 7.8%). Examples of typical SC1 responses are “*Because I know about limiting reactants*” and “*I just think so, but I know nothing about a mole*”. The following statements are typical examples of SC5 responses: “*Because the diagrams are a bit confusing*”, “*I have not encountered such a question before so I find it hard to answer this one*”. The prevalence of objective judgments based on procedural knowledge (SC2) observed in responses like “*I actually performed the calculations to find the exact answer*” decreased from 40.2% in the pretest to 34.6% in the posttest and the global evaluation of answers observed in explanations such as “*I chose 80% because I am not fully sure of the answer but I believe it is so*”, indicating subjective feeling of doubt and uncertainty, increased in the posttest as compared with the pretest (SC3, Pretest: 33.8%, Posttest: 38.7%).

OC-R Subgroup

The prevalence of judgements based on procedural knowledge decreased by almost 10% (SC2 – Pretest: 44.1%; Posttest: 34.4%), while that of subjective judgements remained stable. An example of a typical SC2 response is, “*Because I did calculations to find the correct answer*”. The prevalence of judgements based on a global evaluation of the answer (SC3 – Pretest: 35.6%; Posttest: 31.6%), observed in statements like “*Just a feeling I have*”, decreased by roughly the same margin as the increase documented for the feeling of unpreparedness (SC4 – Pretest: 2.8%; Posttest: 6.1%), observed in statements such as “*We didn't do this section in detail yet*”. This subgroup showed the largest increase in the prevalence of SC5 responses, i.e. from 10.1% in the pretest to 17.4% in the posttest. Several examples are included for this super-category, External factors, in order to demonstrate the nuances of these statements: “*Question not clearly understandable*”, “*Because we have done such questions in the small group lecture and I understood*”, “*Did almost the same question this morning and I got it right*”, “*Quite sure about my answer, learnt it in high school as well as here*” and “*Answer was simple to find*”. The increase in SC5 responses may indicate that after teaching the students in the OC-R subgroup, being the subgroup that showed the highest learning gain, were able to recognise and reveal when they were or were not familiar with the question, when they did or did not understand what was being asked or when the question was easy to solve. This awareness may have placed them in a position to better evaluate their performance during the posttest.

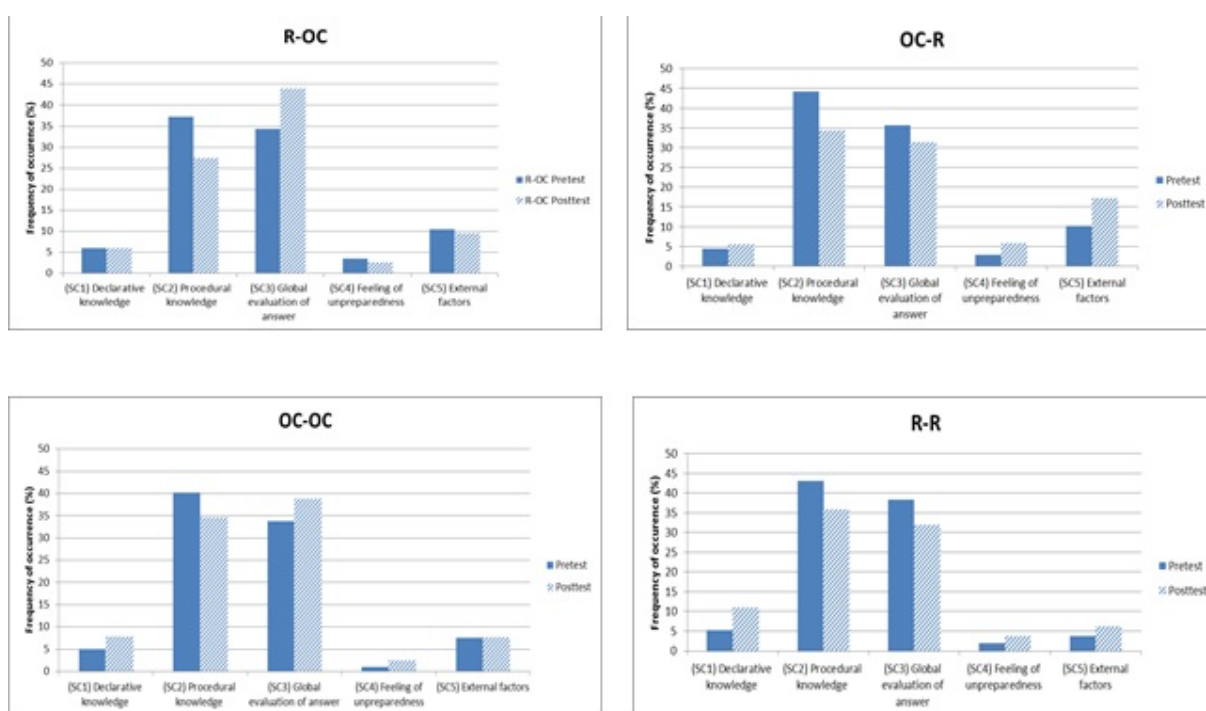


Figure 1. Shifts observed in the prevalence of super-categories for the OC-OC, OC-R, R-OC and R-R subgroups

R-OC Subgroup

The prevalence of judgements based on procedural knowledge decreased (SC2 – Pre 37.2%; Post: 27.4%) and the global evaluation of answers increased by a similar margin (SC3 – Pre: 34.4%; Post: 43.9%). Examples of typical SC2 and SC3 responses are, “Balancing of that equation gives $4Cu_2O$ ” or “Can balance the equation” and “I am doubting” or “I am guessing” respectively. The prevalence of judgements based on external factors was fairly stable (SC5 – Pre: 10.5%, Post: 9.5%). Examples of typical SC5 responses for this subgroup are “Because two possible answers are present. They both have equal chance” and “There are just so many numbers to work with”.

R-R Subgroup

The following shifts were observed in the super-categories of this subgroup. The prevalence of judgements based on declarative knowledge more than doubled in the posttest (SC1 – Pre: 5.3%; Post: 11.0%). Examples of typical SC1 responses were “I know the definition” or “Because I know the law of conservation of mass and I also know how to balance an equation”. The large increase in SC1 responses was unique to the R-R subgroup. A small increase was also observed for judgements based on external factors (SC5 – Pre: 3.8%; Post: 6.2%), evident in responses such as “I am not familiar with reactions in diagrams” or “Because that is what I learned so far in my chemistry classes”. The prevalence of judgements based on procedural knowledge decreased by 7% (SC2 – Pre: 43.1%; Post: 35.9%) and so did the prevalence of judgements based on a global evaluation of the answer (SC3 – Pre: 38.3%; Post: 32.1%). Examples of typical SC2 and SC3 are “I balanced the equation before looking at the given options and I found the answer I worked out” or “I balanced the equation first and therefore calculated the moles of reactants using the moles of the product” and “Because I have a feeling that it’s correct” respectively.

Shifts in the factors students relied on when making confidence judgements

Information on the frequency with which codes per categories occurred enabled us to determine which super-categories and categories were prevalent or dominant in each pre-post performance evaluation subgroup first in the pre- and then the posttest. This was also helpful in tracing changes in the factors students in the different pre-post performance evaluation subgroups relied on when judging their performance after instruction. **Figure 1** shown below gives an indication of the shifts in the prevalence of super-categories after instruction for the students in the OC-OC, OC-R, R-OC and R-R subgroups, respectively.

The two super-categories that feature most prominently for all subgroups are SC2 which is rationally based and SC3 which is not. It was confirmed through the use of the *McNemar* statistical test (Adedokun & Burgess, 2012) that the differences between frequencies of occurrence of responses in the SC2 and SC3 super-categories were highly significant ($p = 0.000$) for all four subgroups. Overall, students in all four subgroups were less inclined to motivate

their choice of confidence judgement ratings based on perceived possession of procedural knowledge (SC2) in the posttest. It seems that with exposure to teaching, students were more aware of what they did and did not know and were thus not too quick to claim the possession of procedural knowledge. However, mixed trends were found for SC3. Students who stayed or became realistic in their performance evaluation (R-R, OC-R) were less inclined to motivate their choice of confidence judgement ratings in the posttest in terms of vague, subjective feelings of certainty or uncertainty (SC3) observed in statements such as *"don't feel confident"*, *"it's right"* or *"just know"*. An opposite trend was observed for the students who remained or became overconfident in their judgements of performance (OC-OC, R-OC). The largest increase in the prevalence of SC3 responses was recorded for the R-OC subgroup. This increase was almost twice that of the increase observed for OC-OC. This is significant, especially in the light of the poor performance of the R-OC subgroup.

Two features identified in the posttest responses of the OC-R subgroup revealed metacognitive processes unique to this subgroup of students. The first unique feature was their inclination to admit when they felt unprepared in the posttest (SC4, 6.1%), which represented the highest incidence of SC4 responses recorded for any of the subgroups. This was evident in statements like *"Never really understood in class"* or *"We did not do this section in detail yet"*. Secondly, their inclination to make judgements based on external factors (SC5) was high in the pretest and it almost doubled in the posttest. Several examples are included for this super-category in order to demonstrate the nuances of these statements: *"Question not clearly understandable"*, *"Because we have done such questions in the small group lecture and I understood"*, *"Did almost the same question this morning and I got it right"* and *"Quite sure about my answer, learnt it in high school as well as here"*. Considering that the students in this subgroup showed the highest learning gain, the type of SC4 and SC5 responses provided by this subgroup suggest that they acquired a more differentiated knowledge of the topic during teaching. The increase in SC4 and SC5 responses may indicate that after teaching the students in the OC-R subgroup improved in their ability to recognise and reveal when they were or were not familiar with the question, when they did or did not understand what was being asked or when the question was easy to solve. This awareness could have enabled them to apply more effective metacognitive monitoring during the teaching and learning of stoichiometry, resulting in a mastery of the content that was superior to that of all of the other subgroups.

DISCUSSION

The interpretation of qualitative results provided another basis for comparing students in the subgroups and it confirmed the finding that we were dealing with four distinct groups with unique properties. The qualitative results assisted in the identification of the factors, whether rational or subjective, that students relied on in making confidence judgements. Observation and analysis of shifts in the prevalence of response categories across all subgroups between the pre- and posttests provided valuable information on changes in metacognitive monitoring after the chemistry content had been taught. These findings complement those derived from the analysis of quantitative data by providing insight into the metacognitive processes associated with performance evaluation, especially those that are associated with higher learning gain.

With the help of relevant literature we were able to understand that responses such as *"I believe that I am correct"* may be an indication of judgements made by students on the basis of feelings of knowing characterised by an intuitive feeling or a hunch which requires no justification. Feelings of not knowing observed in statements like *"I don't know"*, may be observed when little or no information could be retrieved. The danger here is that people may be quick to make an *"I don't know"* judgement when they find the information in the question unfamiliar. This may be the instance when a student would state that he/she does not know the answer to a question because he/she has never come across a stoichiometry question with a pictorial presentation of atoms or molecules as part of the problem statement and then make a statement like *"I am not familiar with the reactions in diagrams"*. Affective feeling-based judgements were observed in responses such as *"this is how I feel"* or *"I am not confident at all 'cause I have never done this. I don't even know how to calculate the answer"*.

Koriat (2000) labels judgements made on the basis of information in metamemory as information-based metacognitive judgements. Judgements made on the basis of experience or how an individual feels at the time of making the judgement he labels experience-based metacognitive judgements. Therefore subjective responses given as motivations for the confidence judgements made in the test, i.e. responses listed under SC3 labelled "Global evaluation of answer", and SC4 - Feeling of unpreparedness, indicate that those judgements were more experience- or feelings-based than information-based. Responses listed in the Declarative knowledge and Procedural knowledge super-categories (SC1 and SC2) imply that the confidence judgements were informed by content information retrieved from metamemory. Responses listed under SC5 labelled "External factors" consisted of a mixture of judgements which were experience and information-based. According to the free-response explanations of the students in the pre- and posttests the frequency with which students in the OC-R and R-R subgroups motivated their choice of confidence indicators based on subjective judgements (SC3) decreased while that of the students in the OC-OC and R-OC subgroups increased. This is a significant finding in the sense that the teaching

and learning experience was most productive for the R-R and OC-R subgroups. They were not only able to acknowledge that they did not know what they did not know but they based their confidence or lack thereof on objective factors. The danger with the tendency of the OC-OC and R-OC group to base their confidence or lack thereof on subjective external factors is that, that which needs to be corrected in order to improve performance is beyond their control.

It seems that during their construction of a confidence judgement students try to search and retrieve information which they can use as evidence that a task will be successfully executed or a question accurately answered. However, in the absence of such information, students may resort to feelings as reference. The trends observed in our sample showed that students who, in spite of teaching, still over-estimated their performance, relied more heavily on feelings whilst answering the test questions rather than on their possession or lack of information to guide their choice of confidence judgement ratings. This probably explains why they remained (OC-OC subgroup) or became (R-OC subgroup) biased in their performance evaluation even after teaching.

Implications for the Teaching and Learning of Chemistry

Intuitively a science educator would expect students to apply logical reasoning when required to motivate their confidence in the correctness of an answer in a science test. Solving stoichiometry problems usually requires predominantly procedural knowledge. Stoichiometry has a very minimal component of memory and recall problems. It is for this specific topic necessary to go beyond recall of declarative knowledge. Solving a typical stoichiometry problem requires both formal reasoning and the use of multistep mathematical operations. When stoichiometry problems include pictorial representations at the atomic or molecular level in their problem statements, students would also have to demonstrate the successful manipulation of the sub-micro and symbolic levels of thinking as described by Johnstone (1991). Solving such stoichiometry problems will require students to demonstrate representational competence and would challenge them to reveal their conceptual understanding. Furthermore it was expected that the nature of the topic, and hence the type of questions on the topic would elicit rational rather than subjective responses. We expected that students would base their choice of confidence indicators on the possession or lack of formal reasoning, mathematical skills, conceptual understanding, declarative knowledge, or the procedural knowledge that may be required to solve the problem. Instead, our results revealed that even in a science test in the specific format that we used, students believed an answer to be correct based on feelings rather than on rationally motivated judgements. A super-category which clearly reflects a subjective judgement, SC3, (Global evaluation of answer) constituted a substantial 30 to 40% of all the confidence judgements in the pre- and posttest responses for all the subgroups and this was unexpected (OC-OC - Pre: 33.8%, Post: 38.7%; OC-R - Pre: 35.6, Post: 31.6%; R-OC - Pre: 34.4%, Post: 43.9%; R-R - Pre: 38.3%, Post: 32.1%).

Our findings suggest that increasing content knowledge alone may not be enough to raise the metacognitive ability of students. Chemistry educators should be aware that students often make vague subjective judgements of performance even on a topic like stoichiometry, which requires predominantly procedural knowledge and formal reasoning. Our study has shown that this deficiency, when associated with poor accuracy of self-evaluation, may hamper learning gain. The pattern observed in the quantitative and qualitative results of the OC-R subgroup seem to suggest that the ability to metacognitively think about one's performance may be a critical ingredient for successful learning. It seems that an attempt to increase metacognition among university students instruction should amongst other things focus on the teaching of self-regulatory skills such as specific monitoring and regulatory strategies that students can use in different academic tasks like during test-taking. Such an approach may prevent the damage caused by failure and preserve the positive contribution of confidence, albeit excessively positive.

REFERENCES

- Adedokun, O. A., & Burgess, W. D. (2012). Analysis of paired dichotomous data: a gentle introduction to the McNemar test in SPSS. *Journal of MultiDisciplinary Evaluation*, 8(17), 125-131.
- Ben-Zvi, R., Eylon, B., & Silberstein, J. (1988). Theories, principles and laws. *Education in Chemistry*, 89-92.
- Beyer, S. (1999). Gender differences in the accuracy of grade expectancies and evaluations. *Sex Roles*, 41(314), 279-296. <https://doi.org/10.1023/A:1018810430105>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3, 77-101. <https://doi.org/10.1191/1478088706qp063oa>
- Dunlosky, J., Serra, M. J., Matvey, G., & Rawson, K. A. (2005). Second-order judgements about judgements of learning. *The Journal of General Psychology*, 132(4), 335-346. <https://doi.org/10.3200/GENP.132.4.335-346>
- Ehrlinger, J. (2008). Skill level, self-views and self-theories as sources of error in performance evaluation. *Social and Personality Psychology Compass*, 2(1), 382-398. <https://doi.org/10.1111/j.1751-9004.2007.00047.x>

- Fernandez-Duque, D., & Black, S. E. (2007). Metacognitive judgment and denial of deficit: Evidence from frontotemporal dementia. *Judgment and Decision Making*, 2(5), 359–370.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new era of cognitive-developmental inquiry. *American Psychologist*, 34, 906–911. <https://doi.org/10.1037/0003-066X.34.10.906>
- Fleiss, J. L. (1981). *Statistical Methods for Rates and Proportions* (2nd Ed.). New York: Wiley.
- Hacker, D. J., Bol, L., & Bahbahani, K. (2008). Explaining calibration accuracy in classroom contexts: the effects of incentives, reflection, and explanatory style. *Metacognition Learning*, 3, 101–121. <https://doi.org/10.1007/s11409-008-9021-5>
- Hake, R. R. (1998). Interactive-engagement vs traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66, 64–74. <https://doi.org/10.1119/1.18809>
- Harrits, G. S. (2011). More than method? : A discussion of paradigm differences within mixed methods research. *Journal of Mixed Methods Research*, 5(2), 150–166. <https://doi.org/10.1177/1558689811402506>
- Huddle, P. A., & Pillay, A. E. (1996). An in-depth study of misconceptions in stoichiometry and chemical equilibrium at a South African university. *Journal of Research in Science Teaching*, 33(1), 65–77. [https://doi.org/10.1002/\(SICI\)1098-2736\(199601\)33:1<65::AID-TEA4>3.0.CO;2-N](https://doi.org/10.1002/(SICI)1098-2736(199601)33:1<65::AID-TEA4>3.0.CO;2-N)
- Jing, L., Kazuhisa, N., & Yuejia, L. (2003). Neural correlates of “feeling-of-not-knowing”: evidence from functional MRI. *Chinese Science Bulletin*, 48(2), 144–147. <https://doi.org/10.1360/03tb9029>
- Johnstone, A. H. (1991). Why is science difficult to learn? Things are seldom what they seem. *Journal of Computer assisted learning*, 7(2), 75–83. <https://doi.org/10.1111/j.1365-2729.1991.tb00230.x>
- Koriat, A. (1993). How do we know what we know? The accessibility model of the feeling of knowing. *Psychological Review*, 100, 609–639. <https://doi.org/10.1037/0033-295X.100.4.609>
- Koriat, A. (2000). The feeling of knowing: Some metatheoretical implications for consciousness and control. *Consciousness and Cognition*, 9, 149–171. <https://doi.org/10.1006/ccog.2000.0433>
- Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one’s knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(2), 187–194. <https://doi.org/10.1037/0278-7393.31.2.187>
- Koriat, A., & Bjork, R. A. (2006). Mending metacognitive illusions: a comparison mnemonic-based and theory-based procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(5), 1133–1145. <https://doi.org/10.1037/0278-7393.32.5.1133>
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one’s own incompetence lead to inflated performance evaluations. *Journal of Personality and Social Psychology*, 77(6), 1121–1134. <https://doi.org/10.1037/0022-3514.77.6.1121>
- Leech, N. L., & Onwuegbuzie, A. J. (2011). Mixed research in counselling: Trends in the literature. *Measurement and Evaluation in Counselling and Development*, 44(3), 169–180. <https://doi.org/10.1177/0748175611409848>
- Mathabathe, K.C., & Potgieter, M. (2014). Metacognitive monitoring and learning gain in foundation chemistry. *Chemistry Education Research and Practice*, 15, 94–104.
- Maxwell, J. A., & Loomis, D. M. (2003). Mixed methods design: An alternative approach. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioural research*. India: Sage publications.
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation: advances in research and theory*. San Diego, California: Academic Press, Inc.
- Nowell, C., & Alston, M. R. (2007). I thought I got an A! Overconfidence across the economics curriculum. *Journal of Economic Education*, 131–142. <https://doi.org/10.3200/JECE.38.2.131-142>
- Nussinson, R., & Koriat, A. (2008). Correcting experience-based judgements: the perseverance of subjective experience in the face of the correction of judgement. *Metacognition Learning*, 3, 159–174. <https://doi.org/10.1007/s11409-008-9024-2>
- Potgieter, M., Davidowitz B., & Mathabatha S. (2007). Do they know that they don’t know? The relationship between confidence and performance of first year chemistry students at three tertiary institutions in South Africa. *Presented at the 38th annual conference of the Australasian Science Education Research Association (ASERA) in Fremantle, WA.*
- Ridley, D. S., Schutz, P. A., Glanz, R. S., & Weinstein, C. E. (1992). Self-regulated learning: the interactive influence of metacognitive awareness and goal-setting. *Journal of Experimental Education*, 60(4), 293–306. <https://doi.org/10.1080/00220973.1992.9943867>

- Rosenthal, D. M. (2000). Consciousness, Content, and Metacognitive Judgments. *Consciousness and Cognition*, 9, 203–214. <https://doi.org/10.1006/ccog.2000.0437>
- Schaefer, P. S., Williams, C. C., Goodie, A. S., & Campbell, W. K. (2004). Overconfidence and the Big Five. *Journal of Research in Personality*, 38, 473–480. <https://doi.org/10.1016/j.jrp.2003.09.010>
- Schwartz, B. L., & Perfect, T. J. (2002). Introduction: toward an applied metacognition. In B. L. Schwartz & T. J. Perfect (Eds.), *Applied Metacognition*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511489976.002>
- Winnie, P. H., & Nesbit, J. C. (2010). The psychology of academic achievement. *Annual Review of Psychology*, 61, 653–678. <https://doi.org/10.1146/annurev.psych.093008.100348>

<http://www.ejmste.com>