

Inaccuracy of self-evaluation as additional variable for prediction of students at risk of failing first-year chemistry

Marietjie Potgieter^a, Mia Ackermann^a and Lizelle Fletcher^b

Received 2nd April 2009, Accepted 3rd December 2009

DOI: 10.1039/C001042C

Early identification of students at risk of failing first-year chemistry allows timely intervention. Cognitive factors alone are insufficient predictors for success; however, non-cognitive factors are usually difficult to measure. We have explored the use of demographic and performance variables, as well as the accuracy of self-evaluation as an indicator of metacognitive ability, as possible indicators for students at risk of failing the first semester course in General Chemistry (CMY 117) at the University of Pretoria. Variables with a strong correlation with performance in CMY 117 were used to develop a prediction model based on logistic regression. Three variables, *i.e.* prior performance in mathematics and in physical science, and the extent of overconfidence expressed as the ratio between expected and actual performance in a chemistry pre-test written at the start of the semester, were shown to be significant predictors for risk of failing. The highest overall accuracy of prediction (76%) was obtained for a subset of students with a C or D grade for their high school leaving examination in mathematics when high risk students were defined as those with a final mark for CMY 117 as 51% or lower. The prediction model, based on the model building data set, had a sensitivity of 92% and a specificity of 46%; whilst the sensitivity and specificity using the validation data set were 88% and 38% respectively.

Keywords: chemistry; prediction; prediction model; students at risk of failing; self-evaluation, overconfidence; logistic regression; predictor variables

Introduction

There is a need to predict student success in first-year chemistry in order to sustain acceptable throughput of science and engineering graduates. Early identification of students at risk of failing allows timely intervention in an attempt to prevent failure. Well-researched factors such as mathematical ability, chemistry ability and demographic factors are often used as indicators of success or failure in tertiary chemistry (McFate and Olmsted, 1999; Legg *et al.*, 2001; Wagner *et al.*, 2002). The predictive capability of these factors is weakened by the unique social and psychological profile of each student, which cannot be measured directly (Angel and LaLonde, 1998; McFate and Olmsted, 1999; Tai *et al.*, 2005). This study attempts to improve the predictive capability of existing statistical models by indirectly incorporating psychological factors.

The basic assumption is that the student's ability to accurately assess his test performance as indication of mastery of chemistry is a necessary prerequisite for success at tertiary level. The inability to do so results in overconfidence, which was shown to be related to risk of failure (Ochse, 2003). Overconfidence as a construct in the context of this study refers exclusively to an overoptimistic judgment of one's own performance. This judgment is made immediately after the task has been completed (postdiction). An incorrect judgment

is either positively or negatively biased, thereby reflecting overconfidence or under-confidence, respectively.

Moderate pass rates (ca. 60%) of the first semester course in General Chemistry (CMY 117) at the University of Pretoria (UP) are a concern for the institution, because of large student enrolment in the module (ca. 950). The course content for CMY 117 is based on international textbooks for similar courses offered elsewhere in first world countries (*e.g.* Silberberg, 2006; Brown *et al.*, 2008), and the proficiencies of students upon entry to the course, as demonstrated by their performance on the Chemical Concepts Inventory (Mulford and Robinson, 2002) closely resembles that reported for first year students at Purdue University, USA (Supplement to Mulford and Robinson, 2002). Annually, almost 100 students achieve a final mark between 40% and 49% for this module. If these students could be identified early enough, it may be possible to offer remedial support to prevent failure. This project aims to develop a regression model for the early prediction of risk of failing CMY 117, hence identifying students at risk within the first few weeks after lectures have commenced.

Literature review

Placement tests such as the Toledo Exam, California Chemistry Diagnostic Test, SAT and ACT have been developed to predict student success (referenced in Lewis and Lewis, 2007). Most of these tests have reasonably good overall predictive accuracy, but their predictive accuracy for students at risk of failing is poor (McFate and Olmsted, 1999;

^a Department of Chemistry, University of Pretoria, South Africa, E-mail: marietjie.potgieter@up.ac.za

^b Department of Statistics, University of Pretoria, South Africa.

Wagner *et al.*, 2002; Lewis and Lewis, 2007). Students' prior academic performances in mathematics and chemistry have a moderately strong correlation with performance in first-year chemistry (McFate and Olmsted, 1999; Evans, 2000; McKenzie and Schweitzer, 2001; Wagner *et al.*, 2002; Thai *et al.*, 2005). Similarly, the cognitive variable, logical reasoning ability is as successful for the prediction of success in tertiary chemistry as SAT scores (Bunce and Hutchinson, 1993). In a recent article in this journal, Lewis and Lewis (2007) showed that formal thought (as measured by the Test of Logical Thinking, TOLT) and general achievement (as measured by SAT) represent separate and distinct prediction factors for students at risk of failing first-year General Chemistry. This result implies that each of these proficiencies represent a separate hindrance to success in chemistry.

There is ample evidence, however, that academic ability alone is not a sufficient predictor of first-year success. Demographic, psychological, sociological and institutional factors have also been investigated for their possible predictive power (reviewed by Evans, 2000). Demographic factors such as gender, linguistic skills and socio-economic status (SES) have been shown to influence tertiary success (North, 1985; Evans, 2000), but the results obtained for gender as a predictor variable are inconclusive. Maturity, rather than age, might be negatively correlated with students' risk of failing. Maturity in this context refers to students who delayed commencing tertiary studies after secondary school (Evans, 2000). Psychological factors include academic preparedness, learning strategies, locus of control, stress and anxiety (Evans, 2000; Tchen *et al.*, 2001). A student's academic self-concept and attribution style also influence his success at tertiary level (Evans, 2000; McKenzie and Schweitzer, 2001). Moderate to weak, but significant correlations between chemistry and mathematics self-concepts and achievement in first-year chemistry were found in studies by House (1995) and Bauer (2005). Sociological factors such as family and peer support determine whether a student will persist in tertiary studies, but do not necessarily predict first-year success (Evans, 2000). Institutional factors that determine academic performance are academic and social integration into the university, and educational support from lecturers (Evans, 2000; Tchen *et al.*, 2001). We have included several demographic factors in our study, as well as an expression of metacognitive ability, which may be categorised as a psychological factor; however, sociological and institutional factors are beyond the scope of this project.

The literature provides further insight into metacognition and its relevance to academic success. Metacognition is a multi-faceted construct described by Gourgey (2001) as "... awareness of how one learns; awareness of when one does or 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." Tobias and Everson (2002) have found that the ability to differentiate between what is known (learned) and unknown (unlearned) is an important ingredient for success in all academic settings. Students with

high metacognitive ability know when they have mastered or not mastered the required academic tasks and can adjust their learning accordingly. Students' ability to monitor their learning is a key component of self-regulated learning, which, in turn, is an essential requirement for success at tertiary level (Schunk and Zimmerman, 1994).

Metacognitive skills are domain specific. In the area of chemistry the development of metacognitive skills is required for in-depth learning, enhanced conceptual understanding and successful problem solving (Rickey and Stacy, 2000). In a recent article in this journal, Cooper *et al.* (2008) have reported the development of an across-method-and-time instrument to assess the use of metacognition and to guide the implementation of interventions to promote its use in chemistry problem solving.

A number of studies have been reported where metacognitive ability of students was assessed and correlated with test performance by means of confidence judgments where respondents indicated the likelihood that the answers provided to each multiple-choice question was correct (*e.g.* Sinkavich, 1995; Carvalho, 2007). Confidence judgments represent one of a range of measures commonly used by cognitive psychologists to probe metacognitive activity. Researchers have consistently found that people are not very good at accurately assessing their knowledge or competence (Bol and Hacker, 2008 and references cited therein). Mabe and West (1982) surveyed 55 calibration studies with a combined population of 14,811 subjects across a variety of domains and found the average correlation between self-estimates and actual performance to be only 0.29. While performance estimates have been shown to be generally unrealistic, overestimation of performance was particularly prevalent (recently reviewed by Dunning, 2005). Numerous studies reported a relationship between ability level and accuracy of calibration (*e.g.* Sinkavich, 1995; Hacker *et al.*, 2000; Larres *et al.*, 2003; Isaacson and Fujita, 2006). The weakest performance was associated with the largest error in calibration, whereas best performing students judged their performance most accurately. Isaacson and Fujita (2006) showed that low achieving students are less able to predict their performance after writing a test, rely more on time spent on studying than on mastery of concepts to decide their confidence for success, are less likely to adjust their self-efficacy depending on feedback received from taking a test, and show the largest discrepancy between their actual performance and their expected performance.

A large error in self-evaluation is, however, not merely an indication of lower cognitive ability. The accuracy of confidence judgments is related to both intelligence and personality (Pallier *et al.*, 2002). Schaefer *et al.* (2004) have found overconfidence to be positively correlated with two personality measures, *i.e.* extraversion and openness to experience. Gramzov and coworkers (2003) have argued that exaggerated self-reports are motivated by either the need for self-enhancement or self-protection, and showed that this underlying motivation has either a positive or negative relationship to subsequent performance. It is evident that the accuracy of performance self-evaluation as expressed by a

confidence statement is a reflection of both cognitive and non-cognitive factors which influence the quality of judgment that is made.

Our study contributes to a large collection of predictor papers, but it has the specific focus of prediction of risk of failing first semester General Chemistry for which only a handful of previous studies have been reported (Legg *et al.*, 2001; Wagner *et al.*, 2002; Lewis and Lewis, 2007). Furthermore, we have explored the predictive power of accuracy of self-evaluation as a measure of metacognitive ability, for risk of failing, which has not been reported before.

Methodology

Data sources

As part of a bigger project, we have developed an assessment instrument to monitor baseline conceptual understanding of students on first entry to tertiary chemistry (Potgieter *et al.*, 2008). This instrument was designed to serve the dual purpose of diagnosis and prediction. It was used as a pre-test within the first week of instruction to collect performance and confidence data from all first-time entrants who enrolled for CMY 117 in 2005 (513 students) and 2007 (581 students). The pre-test consisted of 10 general items and 37 chemistry items, each of which was followed by a self-reported certainty of response indication, where students reported their confidence in the correctness of their answers (Table 1). Self-reported data on home language, type of school, chemistry self-concept and motivation for passing CMY 117 was collected by means of the 10 general items. Data obtained from the UP bureau of student statistics included students' mathematics and science results for their school leaving examinations, language proficiency, study program, race, gender and year of matriculation.

Students' actual performance on the pretest was calculated by assigning a value of 1 for a correct and 0 (zero) for an incorrect answer to the chemistry test items. Students' expected pretest performance was calculated from the confidence data collected with the same instrument. The item was scored zero if the student selected confidence indicators A or B and one if the student selected C or D (Table 1), based on the assumption that the student expected to be wrong or correct, respectively.

Compiling the data set

The 2005 and 2007 data sets were combined (1094 records) to minimize instructor or year specific effects. Records which were incomplete due to student attrition (139 records) were removed. The combined data set (955 students) was then divided into categories based on performance in CMY 117. The data set consisted of the following variables: Race; Gender; Language proficiency; Home language; Type of school; Mathematics and Science results obtained in the final school leaving examinations; Year of high school graduation; Delay (years since high school graduation); CMY 117 final results for the course; Actual performance in pretest and Expected performance in pretest.

Guidelines with respect to the minimum observation-to

Table 1 Self-reported confidence index for each chemistry item in the pre-test

Totally guessed answer	Almost a guess	Almost certain	Certain
A	B	C	D

-predictor ratio applicable to logistic regression recommend a minimum ratio of 10 to 1, taking the categories of the qualitative variables into account (Peng *et al.*, 2002). In view of the exploratory nature of this cohort, combined with the large number of categorical variables in this data set, several with 5 categories, it was decided to reserve a large portion of the data set for model building purposes. Hence, 80% of the total number of students was randomly selected from each performance category for the model building data set (*i.e.* 764 students). The remaining 20% (191 students) formed the model validation data set. The two student groups generated in this manner were checked to ensure that they reflected the pass/fail ratio typically obtained for this module over the past four years.

Variable definition

Dependent variable *Risk_51*. The dependent variable *Risk_51* is derived from the final mark that a student obtained for the module CMY 117, labeled *CMY117*. At the very minimum, high risk students (at risk of failing the course CMY 117) are those who achieved a final mark of 49% or lower for the module. However, upper limits of 50% and 51% for CMY 117 performance were also explored for the definition of high risk students. An upper limit of 50% includes all candidates who passed the module after writing a supplementary examination and 51% would include those marginal candidates who could have failed, given a different exam paper. The intention of prediction of risk of failure was to offer remedial support to students who are likely to fail without it. A conservative approach was therefore adopted for the development of the prediction model by defining high risk students as those who achieved a final mark of 51% or lower for the module CMY 117. The binary dependent variable, *Risk_51*, is defined as the probability for a student to obtain a final mark for CMY 117 at or below 51% (high risk) or above 51% (low risk). This decision about the cut-off performance level was evaluated and verified during the model development phase (Ackerman, 2008). Therefore, only the results for the 51% cut-off are reported here.

Predictor variables. During this study the strength of twelve potential predictors for risk of failing CMY 117 were explored; six demographic variables (Race, Gender, Home language, Language proficiency, Type of school and Delay); three performance variables (Mathematics and Science results obtained in final school leaving examinations and Actual performance in pre-test) and three variables for accuracy of self-evaluation (Expected performance in pre-test, and two expressions of miscalibration calculated from the difference between Actual and Expected performance: *Overconfidence* and *Ratio*). For the purpose of the report and discussion below

a definition of the following eight variables are provided:

- **Math:** Grade 12 mathematics results (Final school leaving examination)
- **Science:** Grade 12 science results (Final school leaving examination)
- **Delay:** Time lapsed (in years) between high school graduation and commencement of tertiary studies
- **Pre:** Students' actual performance in the pre-test (%)
- **Expected:** Students' expected performance in the pre-test (%)
- **Overconfidence:** Calculated as the difference between the expected and the actual test performance expressed as a percentage of the number of questions in the test.
- **Ratio:** The extent of miscalibration, calculated as the expected test performance divided by the actual test performance
- **MathSym:** High school mathematics results were converted to a symbol to determine risk distribution within categories: A (≥ 80), B ($70 \leq x \leq 79$), C ($60 \leq x \leq 69$), D + E ($40 \leq x \leq 59$)

Data Exploration

The original model building data set consisted of 742 students who had complete records. The descriptive statistics in Table 2 show that the mean of *Expected* is 20% higher than that of *Pre*. The large difference between the two means reflects the extent of miscalibration that was observed among the students. Similarly, the mean of *Ratio* (1.47) implies that students, on average, overestimated their true performance by 47%.

Correlation of the predictor variables with the dependent variable *Risk_51*. The potential predictor variables belong to three categories: demographic characteristics, prior performance and metacognitive ability.

Demographic characteristics. None of the demographic variables considered as possible predictor variables had a significant association with *Risk_51*. Approximately the same percentage of males and females were classified as high risk and low risk. This lack of a significant correlation between gender and the dependent variable agrees with results obtained by Evans (2000). The large bias in the data set in terms of race (86% white students), home language (Afrikaans 52%, English 37%, 11% African languages and others) and language proficiency (91% proficient) does not allow any inferences to be made regarding the reported influence of linguistic skills and cultural background on performance (North, 1985; Evans, 2000). Socio-economic status (SES), as evidenced by the type of school attended, has no predictive ability for performance in our data set: this is probably the case because only 4% of students attended rural or township schools. *Delay* (years since high school graduation) was used as an indication of student maturity. Eighty eight percent of students commenced their studies directly after high school and 10% started a year later. The 68 students who had the single gap year are equally split between High risk and Low risk. No conclusions can therefore be drawn regarding student performance and maturity.

Table 2 Descriptive statistics

	Minimum	Maximum	Mean	Std. Deviation
<i>CMY117</i>	19	94	54.75	14.058
<i>Math</i>	40	100	70.43	12.136
<i>Science</i>	44	98	71.94	10.602
<i>Delay</i>	1	8	1.18	0.621
<i>Pre</i>	15	97	52.67	14.832
<i>Expected</i>	3	100	73.32	16.192
<i>Overconfidence</i>	-41	70	20.66	16.577
<i>Ratio</i>	0.06*	3.8	1.4741	0.46775

*Low value for *Ratio* because of difference between expected performance (3%) and actual performance (43%). No statistical reason to exclude from data set.

Table 3 Frequency table for Grade 12 mathematics performance

Mathematics performance	Frequency	Percent
A	212	28.6
B	182	24.5
C	190	25.6
D	145	19.5
E	13	1.8

Table 4 Distribution of students in risk categories according to prior mathematics performance

<i>MathSym</i>		<i>Risk_51</i>		Total
		Low Risk	High Risk	
A	Count	189	23	212
	% within <i>MathSym</i>	89.2%	10.8%	100.0%
B	Count	109	73	182
	% within <i>MathSym</i>	59.9%	40.1%	100.0%
C	Count	80	110	190
	% within <i>MathSym</i>	42.1%	57.9%	100.0%
D + E	Count	34	124	158
	% within <i>MathSym</i>	21.5%	78.5%	100.0%
Total	Count	412	330	742
	% within <i>MathSym</i>	55.5%	44.5%	100.0%

Student prior performance. The data set was divided into five categories based on Grade 12 mathematics results. Students were fairly equally distributed across the first four categories as shown in Table 3.

Table 4 presents results of prior performance in mathematics, cross-tabulated with the dependent variable, *Risk_51*. For statistical purposes the students with E symbols for mathematics were grouped with students with D symbols.

As expected, the percentage of students in the high risk category increased with weaker performance in mathematics. Using the Chi-Square test, this association was found to be statistically significant ($\chi^2(3) = 186.318$, p -value < 0.001). Grade 12 performance in science was used as indication of prior performance in chemistry. A picture almost identical to that for mathematics emerged for science (not shown): An increase in the proportion of high risk students with weaker science performance. This association is also statistically

Table 5 Correlation table between *Math*, *Science*, *Pre* and *CMY117*

	<i>Math</i>	<i>Science</i>	<i>Pre</i>
<i>CMY117</i>	0.643*	0.718*	0.405*
<i>Math</i>	1	0.752*	0.325*
<i>Science</i>		1	0.390*

* All correlations are significant at the 0.01 level (2-tailed).

Table 6 Correlation between *CMY117*, *Ratio*, *Expected*, and *Overconfidence*

	<i>Expected</i>	<i>Overconfidence</i>	<i>Ratio</i>
<i>CMY117</i>	0.168	-0.198	-0.278
<i>Overconfidence</i>	0.59	1	0.922

significant ($\chi^2(3) = 231.790$, p -value < 0.001).

The correlations between the different performance variables *Math* (Grade 12 mathematics), *Science* (Grade 12 Science), *Pre* (pre-test performance) and *CMY117* (final results for introductory chemistry) is reported in Table 5.

A linear relationship between *Math* and *CMY117* and *Science* and *CMY117* was found, with the latter having a stronger relationship with *CMY117* than *Math*. Student performance in the pre-test, *Pre*, is not so strongly correlated with *CMY117*. Even though the correlation coefficient is significant at a 1% level, *Pre* only explains 16% of the variation in *CMY117* where *Math* explains 41% and *Science* 52%. As to be expected the correlation between Grade 12 mathematics and science is also statistically significant. The Pearson's correlation coefficient of $r=0.752$, however, does not indicate collinearity, hence both can be used as predictors in the regression model.

Accuracy of self-evaluation. We have used students' confidence judgments as a measure of accuracy of self-evaluation. Various ways of incorporating students' accuracy of self-evaluation into the model were explored. Even though all the confidence variables (*Expected*, *Overconfidence* and *Ratio*) are significantly correlated with *CMY117* at a 1% level, the correlations are very weak (Table 6). *Ratio* (Expected performance/Actual performance) has the strongest correlation. The negative sign indicates an inverse relationship, which reflects the fact that a stronger positive bias in self-evaluation (overconfidence) is associated with weaker performance in the course CMY 117. As expected the association between *CMY117* and *Overconfidence* is also negative.

Ratio has a stronger correlation with *CMY117* than *Overconfidence*, which made *Ratio* the variable of choice to consider in the prediction model. Accuracy of self-evaluation as reflected by the *Ratio* variable across performance levels is explored next. The pre-test consisted mainly of conceptual questions that could not be answered from memory alone. It was written at the start of the academic year after a long summer vacation. Under these circumstances a miscalibration for three to five test items out of a total of 37 (10–15%) was considered acceptable. It was therefore heuristically decided to define overconfident students as those who overestimated

Table 7 Distribution of students' CMY 117 performance across the confidence categories

Performance in CMY 117	N	Overconfident (%) ^a	Confident (%) ^a	Under-confident (%) ^a
≤ 30 %	34	91	3	6
31 – 40 %	95	81	15	4
41 – 50 %	180	84	9	7
51 – 60 %	183	77	16	7
61 – 70 %	151	73	22	5
71 – 80 %	70	64	33	3
≥ 81 %	29	52	45	3

^a Percentage of students in each performance category.

their true performance by more than 15% (*Ratio* > 1.15). Because of the skewness of *Ratio* data (Table 2) it was decided that the margin for acceptable error should also be slightly skewed towards overoptimism. Under-confident students were defined as those who underestimated their actual performance by more than 10% (*Ratio* < 0.9) and realists (labelled confident) as those who estimated their performance between these margins ($0.9 \leq \text{Ratio} \leq 1.15$). Table 7 displays the distribution of students over the three categories of confidence relative to *CMY117*.

In general, the majority of students were overconfident, but the prevalence of overconfidence increases substantially with weaker performance in CMY 117, which suggests overconfidence as a risk factor for success in CMY 117.

Model development

For the purpose of this study students were to be classified as either high risk or low risk. The use of a logistic regression model is appropriate for the prediction of a dichotomous dependent variable, which was *Risk_51* in our case (Legg *et al.*, 2001; Hair *et al.*, 2006). Logistic regression models readily accommodate categorical predictor variables, and do not require that predictor variables follow a multivariate normal distribution with an equal covariance matrix, hence are preferable to a discriminant function analysis, which could also be used to classify observations (Peng *et al.*, 2002).

Performance in high school mathematics constitutes the primary criterion for admission to the Faculty of Natural Sciences at UP. From Table 4 it follows that 71% of all high risk students obtained a C or lower symbol for Grade 12 mathematics (234 of the 330 students). This subgroup (students with a C symbol or lower for Grade 12 mathematics) was hence chosen for model building, as the focus of this study is on the identification of students at risk of failing. The resulting data set has 348 observations, justifying the original decision to allocate a large proportion of the complete data set (80%) to the model building set.

Unlike previously mentioned placement tests, this model aims to predict students at risk of failing more accurately, rather than accurately predicting students with the ability to pass. Since this model will not be used to determine placement, wrongly classifying a student as high risk will only result in additional help that was not needed, and will have no

Table 8 Sensitivity, specificity and overall prediction accuracy of the significant predictor variables

<i>Risk_51</i> n= 348 (100%)	High risk 234 (67%)	Low risk 114 (33%)	
	Sensitivity	Specificity	Overall
<i>Math</i>	230 (98%)	13 (11%)	70%
<i>Science</i>	209 (89%)	41 (36%)	72%
<i>Ratio</i>	231 (99%)	1 (1%)	67%
<i>Math and Science</i>	209 (89%)	39 (34%)	71%
<i>Math, Science and Ratio</i>	214 (92%)	52 (46%)	76%

implication on the student financially, or on the course of the student's studies.

A stepwise procedure was employed to identify base variables for the model. The model was then refined by the introduction of single additional variables to explore whether further improvement of prediction accuracy could be achieved. It was established that *Math*, *Science* and *Ratio* were the best predictors for the model building data set at a 5% level of significance.

Results and discussion

The adequacy of the final logistic regression model was assessed using the Goodness-of-Fit test of Hosmer and Lemeshow (1991) using SPSS Statistics 17.0. The non-significant Chi-square statistic confirms that the selected model is appropriate ($\chi^2(8) = 10.361$, p -value = 0.241). The prediction accuracy of a statistical model is typically reported using three terms: sensitivity, specificity and overall prediction accuracy. In the context of this study, sensitivity refers to the ability of the model to correctly classify a student at risk of failing CMY 117 as high risk and specificity is the ability of the model to correctly classify a student not at risk of failing CMY 117 as low risk. The overall prediction accuracy of a model is the weighted average of these two prediction percentages.

The effect of the three significant predictor variables, both individually and in combination, on the prediction accuracy of the model is reported in Table 8.

The prediction accuracy of the various models in Table 8 should be judged against a reference model based purely on chance. If the only information available is the distribution of the dependent variable, we could predict all students as high risk in which case we would be correct 67% of the time with a sensitivity of 100%. However, the specificity of this reference model would be 0%. The prediction accuracy of the model including only *Math*, for example, improves by merely 3% to 70%, but the specificity increases to 11% with a loss of only 2% in sensitivity. *Math* by itself is an insufficient predictor variable as it tends to classify most students as high risk. Its low specificity however warrants its use in combination with *Science*. *Science* is a more accurate single predictor variable than *Math* and the model diagnostics improve slightly when *Science* is used in combination with *Math* rather than alone. As mentioned, the University of Pretoria admits students to B.Sc. according to their mathematics results. Based on what is

Table 9 Summary of prediction accuracies and variable impact for the model building data set

	<i>Risk_51</i>
Overall Accuracy	76%
Specificity	46%
Sensitivity	92%
Variable Impact*	
<i>Math</i>	0.955
<i>Science</i>	0.898
<i>Ratio</i>	2.142

*Estimated regression coefficients Exp(B).

observed in Table 8, it might be more accurate to admit students according to their science results, since *Science* is a better predictor of a students' risk of failure. *Ratio* as a single predictor variable also classifies almost all students as high risk, but in combination the three variables have the highest overall predictive accuracy. The prediction accuracy of this final model, when judged against the reference model, improved by 9% with a moderate loss of 8% in sensitivity and a considerable gain of 46% in specificity.

From a cost effective point of view, including only *Math* and *Science* as prediction variables eliminates the need for a pretest for the determination of the *Ratio* variable. However, they lack in prediction accuracy of low risk students. If *Ratio* is included the specificity of the model increases significantly, thereby reducing the number of students wrongly classified as high risk. The cost of pretest administration must be offset against the cost of remedial support, because with only *Math* and *Science* as predictor variables more students will be identified as in need of support. Table 9 provides a summary of the variable impact of the model.

The large value for the impact of the *Ratio* variable in all the models confirms that the metacognitive ability of accuracy of self-evaluation is indeed a requirement for academic success. This is especially valuable for the students in the CD mathematics performance bracket as *Ratio* has the ability to distinguish between high risk and low risk students with similar high school results for mathematics and science. Almost half of the cohort is represented in this category (47%, Table 4). The success of the CD model is especially rewarding in the light of the large size of this group and the high prevalence of high risk as well as overconfidence within this group (Tables 4 and 7). These students suffer the double curse of being unskilled and unable to recognise it (Kruger and Dunning, 1999).

The exclusion of students who achieved A or B symbols for Grade 12 mathematics from our model for the identification of students who should be included in remedial activities can be justified as follows: Stronger students are on average better at evaluating their own performance (Table 7). They should have a stronger academic foundation to draw on in order to recover in the case of looming failure and are therefore less vulnerable. Furthermore, it may be exceptionally difficult to persuade students who performed well in high school to participate in remedial activities.

For the purpose of the validation of the model with *Math*, *Science* and *Ratio* as predictor variables the subset of 84 students with complete records who obtained C and D

symbols for Grade 12 mathematics were selected from the validation data set (191 students). As stated earlier, the validation data set consisted of 20% of candidates randomly selected according to performance in the module CMY 117. The values for overall prediction accuracy (73%), sensitivity (88%) and specificity (38%) compare favourably to those obtained for the model building data set in spite of the smaller sample size. Judged against the reference model based on chance, this regression model is worthy of consideration as the 38% gain in specificity more than compensates for the modest 6% improvement in prediction accuracy and the 12% loss in sensitivity.

Conclusion

In this study we have demonstrated how confidence judgments can be used to strengthen statistical models for the prediction of students at risk of failing first semester General Chemistry at the University of Pretoria. The sensitivity of our model for the prediction of high risk students is unmatched in the literature (Wagner *et al.*, 2002). One of the limitations of this study is the fact that the independent variable, performance in first semester General Chemistry, is specific to this institution in that it depends on the nature of teaching, learning and assessment as conducted at this institution. We do believe, however, that the findings of the study will be confirmed at other institutions as well for the following reasons: together with only four other universities on the African continent, the University of Pretoria was rated as a first class tertiary institution by international standards (Academic Ranking of World Universities, 2007). As a result, the context and the cohort chosen for the study have characteristics comparable to many others in first world countries. Another limitation is the fact that a fairly comprehensive pre-test consisting of 37 chemistry items was used to determine the accuracy of self-evaluation. The length of this instrument was required for diagnostic purposes, but it presents a complication when the instrument is used for prediction. An extension of the present investigation is planned to determine whether a subset of test items can be identified which would provide the same quality of information on accuracy of self-evaluation for future inclusion in the prediction model.

We have demonstrated that high risk students can be accurately identified for early intervention within the first few weeks of the semester. High risk students were defined as those who achieved a final mark of 51% or lower for the chemistry module CMY 117. No conclusions regarding the influence of demographic factors on student performance could be made. Prior performance in mathematics and science showed significant correlation with first semester performance. Accuracy of self-evaluation as reflected in the variable *Ratio*, which is a measure of metacognitive ability, did not correlate very strongly with performance in CMY 117, but was shown to be a significant predictor variable in the model. The model's overall prediction accuracy compares well to those reported in literature, and its sensitivity of 92% towards high risk students is unmatched (Wagner *et al.*, 2002). The model's low specificity can be tolerated since it

will not be used for placement purposes. Instead, it will classify a larger number of marginal students for remedial support, which will be to their benefit. This project has demonstrated that confidence judgments can be used successfully to strengthen statistical models for the prediction of students at risk of failing first semester General Chemistry.

References

- Academic Ranking of World Universities, Graduate School of Education, Shanghai Jiao Tong University, August 2007, Retrieved 18 February, 2009, from <http://www.arwu.org/rank/2007/ranking2007.htm>.
- Ackermann M., (2008), Department of Chemistry, University of Pretoria, Internal report for Honours B.Sc. research project, available from the corresponding author.
- Angel S. A. and LaLonde D. E., (1998), Science success strategies: an interdisciplinary course for improving science and mathematics education, *J. Chem. Educ.*, **75**, 1437-1441.
- Bauer C. F., (2005), Beyond "student attitudes": chemistry self-concept inventory for assessment of the affective component of student learning, *J. Chem. Educ.*, **82**, 1864-1870.
- Bol L. and Hacker D. J., in *Big theories revisited*, edited by D. M. McInerney and S. van Etten, (2004), Connecticut: Information Age Publishing.
- Bond T. G. and Fox C. M., (2007), *Applying the Rasch model*. New Jersey: Lawrence Erlbaum Associates.
- Brown T. L., LeMay H. E., Bursten B. E. and Murphy C. J., (2008), *Chemistry: the central science* (11th ed.), Upper Saddle River, NJ: Pearson/Prentice-Hall.
- Bunce D. M. and Hutchinson K. D., (1993), The use of the GALT (Group Assessment of Logical Thinking) as a predictor of academic success in college chemistry, *J. Chem. Educ.*, **70**, 183-187.
- Carvalho M. K. F., (2007), Confidence judgments in real classroom settings: monitoring performance in different types of tests, *Int. J. Psychol.*, **42**, 1-16.
- Cooper M. M., Sandi-Urena S. and Stevens R., (2008), Reliable multi method assessment of metacognition use in chemistry problem solving, *Chem. Educ. Res. Pract.*, **9**, 18-24.
- Dunning D., (2005), *Self-Insight: roadblocks and detours on the path of knowing thyself*, New York: Psychology Press.
- Evans M., (2000), Planning for the transition to tertiary study: a literature review, *J. Inst. Res.*, **9**, 1-13.
- Gourgey A. F., (2001), Developing students' cognitive 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.
- Gramzov R. H., Elliot A. J., Asher E. and McGregor H. A., (2003), Self-evaluation bias and academic performance: some ways and some reasons why, *J. Res. Pers.*, **37**, 41-61.
- Hacker D. J., Bol L., Horgan D., and Rokow E., (2000), Test prediction and performance in a classroom context, *J. Educ. Psychol.*, **92**, 160-170.
- Hair J. F., Black W., Babin B., Anderson R. E., and Tatham R. L., (2006), *Multivariate Data Analysis*, 6th ed., Prentice Hall.
- Hosmer W. D. and Lemeshow S., (1991), The importance of assessing the fit of logistic regression models: a case study, *Am. J. Public Health.*, **81**, 1630-1635.
- House J. D., (1995), Noncognitive predictors of achievement in introductory college chemistry, *Res. High. Educ.*, **36**, 473-490.
- Isaacson R. M. and Fujita F., (2006), Metacognitive knowledge monitoring and self-regulated learning: Academic success and reflections on learning, *J. Scholarship Teach. Learn.*, **6**, 39-55.
- Kruger J. and Dunning D., (1999), Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments, *J. Pers. Soc. Psychol.*, **77**(6), 1121-1134.
- Larres P. M., Ballantine J. A. and Whittington M., (2003), Evaluating the validity of self-assessment: measuring computer literacy among entry-level undergraduates within accounting degree programmes at two UK universities, *Account. Educ.*, **12**, 97-112.

- Legg M. J., Legg J. C. and Greenbowe T. J., (2001), Analysis of success in general chemistry based on diagnostic testing using logistic regression, *J. Chem. Educ.*, **78**, 1117-1121.
- Lewis S. E. and Lewis J. E., (2007), Predicting at-risk students in general chemistry: comparing formal thought to a general achievement measure, *Chem. Educ. Res. Pract.*, **8**, 32-51.
- Mabe P. A. III and West S. G., (1982), Validity of self-evaluation of ability: a review and meta-analysis, *J. Appl. Psychol.*, **67**, 280-296.
- McFate C. and Olmsted J., (1999), Assessing student preparation through placement tests, *J. Chem. Educ.*, **76**, 562-565.
- McKenzie K. and Schweitzer R., (2001), Who succeeds at university? Factors predicting academic performance in first-year Australian university students, *High. Educ. Res. Dev.* **22**, 21-33.
- Mulford D. R. and Robinson W. R., (2002), An inventory for alternate conceptions among first-semester general chemistry students, *J. Chem. Educ.*, **79**, 739-744.
- North S., (1985), Language proficiency and first-year university achievements: A case study at the University of Dar es Salaam. *Journal of the Faculty of Arts and Social Sciences*, VII, 27-41.
- Ochse C., (2003), Are positive self-perceptions and expectancies really beneficial in an academic context? *South African J. High. Educ.*, **17**, 67-73.
- Pallier G., Wilkinson R., Danthiir V., Kleitman S., Knezevic G., Stankov L. and Roberts R. D., (2002), The role of individual differences in the accuracy of confidence judgments, *J. Gen. Psychol.*, **129**, 257-299.
- Peng C. J., Lee K. L. and Ingersoll G. M., (2002), An introduction to logistic regression analysis and reporting. *J. Educ. Res.*, **96**, 3-14.
- Potgieter M., Davidowitz B. and Venter E., (2008), Assessment of preparedness of first-year chemistry students: development and application of an instrument for diagnostic and placement purposes, *African Journal of Research in Mathematics, Science and Technology Education, Special edition*, **12**, 1-18. Web version: <https://www.up.ac.za/dspace/handle/2263/10179>, accessed on 24 July 2009. The instrument can be obtained from the first author of this article.
- Rickey D. and Stacy A. M., (2000), The role of metacognition in learning chemistry, *J. Chem. Educ.*, **77**, 915-920.
- Schaefer P. S., Williams C. C., Goodie A. S. and Campbell W. K., (2004), Overconfidence and the Big Five, *J. Res. Pers.*, **38**, 473-480.
- Schunk D. H. and Zimmerman B. J., (1994), Self-regulation of learning and performance: Issues and educational applications. Mahwah, NJ: Erlbaum.
- Silberberg M. S., (2006), Chemistry, the molecular nature of matter and change (4th ed.). New York: McGraw-Hill.
- Sinkavich F. J., (1995), Performance and metamemory: do students know what they don't know? *J. Inst. Psychol.* **22**, 77-87.
- Tchen G., Carter A., Gibbons P. and McLaughlin P., (2001), What is the relationship between indicators of stress and academic performance in first-year chemistry? A prospective study, *J. Inst. Res.*, **10**, 1-12.
- Thai R. H., Sadler P. M. and Loehr J. F., (2005), Factors influencing success in introductory college chemistry, *J. Res. Sci. Teach.*, **42**, 987-1012.
- Tobias S. and Everson H., (2002), Knowing what you know and what you don't: further research on metacognitive knowledge monitoring, College Board Report No. 2002-3. College Board, NY.
- Wagner E. P., Sasser H. and DiBiase W. J., (2002), Predicting students at risk in general chemistry using pre-semester assessments and demographic information, *J. Chem. Educ.*, **79**, 749-755.