

# A Genetically Optimized Predictive System for Success in General Chemistry Using a Diagnostic Algebra Test

Cameron I. Cooper · Paul T. Pearson

Published online: 7 June 2011  
© Springer Science+Business Media, LLC 2011

**Abstract** In higher education, many high-enrollment introductory courses have evolved into “gatekeeper” courses due to their high failure rates. These courses prevent many students from attaining their educational goals and often become graduation roadblocks. At the authors’ home institution, general chemistry has become a gatekeeper course in which approximately 25% of students do not pass. This failure rate in chemistry is common, and often higher, at many other institutions of higher education, and mathematical deficiencies are perceived to be a large contributing factor. This paper details the development of a highly accurate predictive system that identifies students at the beginning of the semester who are “at-risk” for earning a grade of C- or below in chemistry. The predictive accuracy of this system is maximized by using a genetically optimized neural network to analyze the results of a diagnostic algebra test designed for a specific population. Once at-risk students have been identified, they can be helped to improve their chances of success using techniques such as concurrent support courses, online tutorials, “just-in-time” instructional aides, study skills, motivational interviewing, and/or peer mentoring.

**Keywords** General chemistry · Neural network · Student success · Gatekeeper course · Diagnostic algebra test

## Introduction

Due to the hierarchical nature of science, general chemistry is a gateway course that many students must pass before they can enroll in more advanced science courses. This is especially problematic at the authors’ home institution where approximately thirteen degree programs/tracks depend upon successful completion of general chemistry. There are many treatments designed to improve student success and retain science majors, such as concurrent support courses, online tutorials, “just-in-time” instructional aides, study skills training, motivational interviewing, and/or peer mentoring. In order for these treatments to be successful, however, they must be provided to students at risk for failure as early in the semester as possible.

Many institutions of higher education experience failure rates of 25% or more in general chemistry (Campbell 2011). Identifying students at risk for failure in general chemistry at the beginning of the semester is a complex task. No single measure will predict success in general chemistry all of the time. Many studies, detailed later in the Related Studies section, have identified algebra skills as a significant variable for predicting student success in general chemistry. Anecdotally, the chemistry faculty at the authors’ home institution have found that students who fail chemistry typically lack the mathematical skills necessary to pass the course. The goal of this research is to create a predictive system which accurately identifies students at risk for failure in general chemistry due to mathematical deficiencies.

This study examines a diagnostic algebra test (the “DAT”) created by the authors for the purpose of identifying students at the beginning of the semester who are at risk for failure in general chemistry. Nine statistically significant variables for predicting student success or

---

C. I. Cooper (✉) · P. T. Pearson  
Fort Lewis College, Durango, CO, USA  
e-mail: cooper\_c@fortlewis.edu

P. T. Pearson  
e-mail: pearson\_p@fortlewis.edu

failure in chemistry were identified: whether a student is declared white, cumulative GPA, math ACT score, five individual questions from the DAT, and the total score on the DAT. Using neural networks, a function was constructed that gives the probability a student will pass general chemistry with a grade of C or better based on these nine predictor variables. The accuracy of the results was maximized by using an adaptive learning algorithm, called backpropagation, to find optimal weights for each edge of the neural network.

Using all nine predictors for success in general chemistry, the neural network correctly predicted 76% of the students whose course grade was a C- or worse as being at risk for failure at the beginning of the semester. When the total score on the DAT is eliminated as a predictor, the accuracy for identifying at-risk students drops down to 60%, which is a larger drop than when cumulative GPA or math ACT scores are individually eliminated as predictors. This suggests that the DAT is not only valid, but also a powerful predictor for identifying at-risk students. Cumulative GPA is historically one of the best single predictors of college success (Cooper 2007, 2009), and the results of the DAT are as good as for cumulative GPA. One advantage of the DAT over cumulative GPA is that scores on the DAT are easy to obtain, while cumulative GPA for a student may be difficult to obtain, especially for new students or transfer students.

## Background

The college at which this research was conducted is a comprehensive, 4 year public liberal arts college with approximately 3,700 students. The student population is multi-culturally diverse, with approximately 20% Native American students, and is split nearly equally by gender. In a typical semester, approximately 190 students enroll in three sections of general chemistry, and approximately 25% of them do not pass.

At the authors' home institution, the chemistry faculty is interested in identifying students at risk for failure due to mathematical deficiencies. General chemistry requires that students have taken or are concurrently enrolled in College Algebra or Algebra for Calculus. Due to time restrictions and curricular requirements, most chemistry professors do not have time to review presumed algebra skills during class. Instead, chemistry faculty would like to enroll these at-risk students in a concurrent one-credit chemistry problem solving course designed to help students pass general chemistry, and/or give them "think aheads" that will help them with the mathematical skills they need "just in time" before they use them in their chemistry class.

## Related Studies

Over the past 50 years there has been increasing interest in diagnostic tests that predict student success and failure in general chemistry. Such tests include the Toledo Chemistry Placement Exam (Hovey and Krohn 1963; Niedzielski and Walmsley 1982), the Group Assessment of Logical Thinking (GALT) and the Test of Logical Thinking (TOLT) exams (Bunce and Hutchinson 1993; Burke et al. 1999; Bird 2010; Lewis et al. 2010), the California Chemistry Diagnostic Test (Russell 1994; Karpp 1995), the Fullerton Test (McFate and Olmsted 1999), the CU-Denver Placement Exam (Kimbrough and Weaver 1999), the Student Pre-Semester Assessment (SPSA) (Wagner et al. 2002), the University of Iowa Placement Exam (Pienta 2003), the University of Minnesota Mathematics Assessment for Second-Semester Chemistry (UnMASSC) test (Leopold and Edgar 2008), and the math SAT (Pickering 1975; Ozsohomonyan and Loftus 1979; Spencer 1996). The GALT and TOLT exams measure formal reasoning abilities, while the other exams measure ability in chemistry and/or mathematics. The studies that examine mathematical ability all offer evidence to support the general conclusion that success in general chemistry correlates positively with mathematical ability. In fact, one study found that mathematics background was the most strongly predictive of outcome in the course, followed by chemistry background (Wagner et al. 2002).

Predicting whether students will pass or fail general chemistry has been studied thoroughly using the Student Pre-Semester Assessment (SPSA), the math SAT, and the Toledo Exam, sometimes in combination with data about math background, chemistry background, and age (Wagner et al. 2002). Both the SPSA and Toledo Exams measure ability in mathematics in conjunction with chemistry. In contrast, the diagnostic algebra test (DAT) developed by the authors measures only algebra ability. Some algebra questions on the DAT are presented in the context of chemistry, but they do not require knowledge of chemistry. This study constructs a predictive system based on a neural network analysis of the results of the DAT, math ACT scores, cumulative GPA, and student demographics. A comparison of results for the neural network with other predictive systems from Wagner is given in Table 1. Most significantly, the neural network correctly identified 75.6% of students who will fail general chemistry, which is much better than the next best result at 40.8%.

## Methods

Neural networks have been found to be effective means of predicting student success in a variety of settings. They

**Table 1** Comparison of predictive systems

Predictive system	Sensitivity (%) <sup>a</sup>	Specificity (%) <sup>b</sup>	Predictive accuracy (%) <sup>c</sup>
Neural network (DAT, GPA, math ACT, demographics)	75.6	85.2	82.9
SPSA+ (SPSA, math & chem background, age)	40.8	90.6	76.3
SPSA	40.8	87.7	74.1
Math SAT+ (Math SAT, math & chem background, age)	29.6	89.1	72.8
Toledo exam	27.7	87.0	75.1

Results for neural network define passing as C- or better, all other predictive systems define passing as D or better

<sup>a</sup> Percentage of students correctly predicted to fail general chemistry

<sup>b</sup> Percentage of students correctly predicted to pass general chemistry

<sup>c</sup> Overall percentage of students correctly predicted to fail or pass general chemistry

have been shown to give accurate predictions within the fields of mathematics (Cooper 2007), actuarial science (Schumacher et al. 2010), first-year education (Cooper 2008), business administration (Naik and Ragothaman 2004), and graduate studies (Hardgrave and Wilson 1994). Neural networks as a means to predict success in general chemistry would appear to be a viable method given its successful application to a wide variety of fields.

The predictive system developed in this research follows a genetically optimized version of creating a decision support system for predicting student success in a developmental mathematics course developed by the first author in his doctoral studies (Cooper 2007). This research demonstrates the applicability of the methodology to other fields of study outside mathematics (i.e., general chemistry). This project included the following four stages in the creation of the neural network-based predictive system: instrument creation & data collection, input analysis, and neural network creation & evaluation.

#### Instrument Creation and Data Collection

The diagnostic algebra test (DAT) utilized in this research was designed to assess basic math skills and to identify students at risk for failing general chemistry. The DAT consists of twenty multiple choice questions, each of which has four choices and one correct answer. Each question is worth one point, and there is no penalty for incorrect answers. The seven topics and twenty questions on the DAT are given in Table 2.

The DAT is a 20 min paper-and-pencil test given in class during the first week. Students are not allowed to use calculators or notes. The test was announced, but the material on the test was not discussed before it was administered. The tests are scored by machine so that professors can promptly advise students of outcomes and their at-risk classification.

The questions on the DAT were chosen by two mathematicians and one chemist. Some questions on the DAT are

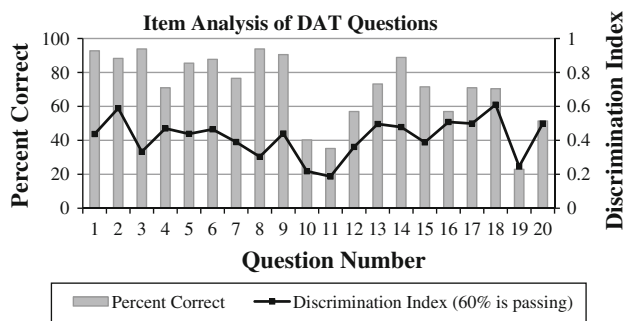
similar to questions on the Moorpark College Chemistry Placement Exam (Moorpark 2010). Three criteria were important for selecting questions. First, only those math skills used most often in general chemistry were tested to ensure the test has high validity. Second, a variety of basic algebra skills were tested to make the instrument general enough to identify students at risk for failure, and also to increase the spread of scores, which should also increase the reliability of the results. Third, each question should require only a small number of algebra skills so that the results for individual questions are specific enough to minimize the false identification of students who will pass the course despite poor performance on the DAT. Individual questions were designed so that the arithmetic would not require a calculator and so that the choices included distracters indicative of common mathematical errors. The multiple choice DAT was designed to have no easily discernible patterns in the answers, and to have no bias toward any particular letter for correct answers. Resultant reliability and validity are discussed later in the paper in the Results section.

An item analysis of the questions on the DAT is given in Fig. 1. The discrimination index is a number between  $-1$  and  $1$  that measures success on an item for students who passed the DAT relative to those who did not pass. A discrimination index above  $0.20$  is considered adequate, while  $0.30$ – $0.50$  is considered good (Wagner et al. 2002). The discrimination index for a question is defined to be  $\frac{c_p}{p} - \frac{c_f}{f}$ , where  $c_p$  is the number of students who answered the question correctly and passed the test,  $c_f$  is the number of students who answered the question correctly and failed the test, and  $p$  and  $f$  are the number of students who passed and failed the test, respectively.

The mean score on the DAT was  $70.9\%$  with standard deviation  $13.7\%$ . When the minimal passing score for the DAT is set at  $60\%$ , the mean discrimination index for all twenty questions is  $0.418$ , which suggests that overall the DAT offers good discrimination. Individual discrimination index results are shown in Fig. 1.

**Table 2** The diagnostic algebra test

Topics	Diagnostic algebra test questions
1. Basic arithmetic and fractions	1. Evaluate $-4 - (-3)$ 2. Evaluate $\frac{-6-3}{4-1}$ 3. Evaluate $2^3 + 3^2$ 4. Write $\frac{m}{2} + \frac{m}{3}$ as a single fraction
2. Linear equations, slopes, and graphs	5. Solve $2x - (4 + 5x) = 14$ for $x$ 6. Solve $\frac{x-7}{3} = 20$ for $x$ 7. What is the slope of the line segment through the points $(0, -2)$ and $(4, 1)$ ? (A graph was given.)
3. Unit conversion, the metric system, and significant digits	8. There are 100 cm in 1 m. If a bug travels 352 cm, how many meters did the bug travel? 9. There are 2.54 cm in 1 inch. A stack of books is 10 inches high. How high is the stack of books in centimeters? 10. There are 1,000 m in 1 km, 60 min in 1 h, and 60 s in 1 min. Convert 72 km/hour to meters/second 11. There are 100 cm in 1 m, and 10 mm in 1 cm. Evaluate $12.50 \text{ cm} - 25 \text{ mm} + 0.300 \text{ m}$ and determine your final answer in centimeters with the proper number of significant figures
4. Scientific notation and powers of 10	12. Simplify $\frac{(8 \times 10^{-4})(3 \times 10^9)}{(4 \times 10^2)}$ 13. Express 3,600 s in scientific notation 14. Express $4.2 \times 10^{-3}$ in decimal form
5. Percentages in the context of chemistry	15. A penny is made from 2% copper and 98% zinc. If the penny has total mass 3 g, what is the mass of the copper in the penny? 16. When popcorn pops, it loses water explosively. If a kernel of corn weighing 0.25 g before popping weighs 0.15 g afterward, what percentage of its mass did it lose on popping?
6. Exponent rules and fractional exponents	17. Simplify $x^4 x^3$ 18. Assuming that $x \geq 0$ , simplify $\sqrt{16x^2}$
7. Evaluating and solving formulas in the context of chemistry	19. A certain glucose solution with mass 100 g has a density of $1.25 \text{ g/cm}^3$ . Determine the volume of this solution in $\text{cm}^3$ . Use that density is mass divided by volume: $d = m/v$ 20. Charles' Law for gases says that the volume, $V$ , of a gas (in liters) is directly proportional to the temperature, $T$ , of the gas (in degrees Kelvin) with constant of proportionality $c$ : $V = cT$ . If the volume of a gas is 600 L and its temperature is 200°K, find $c$



**Fig. 1** Item analysis of DAT questions

**Input Analysis**

Predicting which students are at risk for failure is a complex task that should use measures of current knowledge, background knowledge, and demographics that correlate positively or negatively with academic success. Demographic and academic data from the institution’s student information database, specific questions on the DAT, and

total score on the DAT were considered as possible predictors. The non-parametric statistic Spearman Rho, although only linear, was chosen as a first measure of association for potential ordinal and categorical predictors for course success. Of the 32 possible inputs considered, 9 were found to be statistically significant. Four of these were found to be statistically significant at  $p \leq 0.01$  level (2-tailed)\*\*, and five were found to be statistically significant at  $p \leq 0.05$  level (2-tailed)\*. Collectively, these nine variables provided the baseline neural network input space. A summary of the inputs and their corresponding Spearman Rho correlation coefficients is shown in Table 3.

**Neural Network Creation and Evaluation**

Using Neurosolutions, a comprehensive neural network software package, the authors created a multilayered, feedforward, backpropagation neural network, also referred to as a multilayered perceptron. Neurosolutions utilizes a breadboard approach to neural network creation. The

**Table 3** Statistically significant inputs

Predictor	Spearman's Rho
Declared white	0.200**
Math ACT	0.167*
Cumulative GPA	0.173*
Question 3 on the DAT	0.155*
Question 4 on the DAT	0.217**
Question 5 on the DAT	0.202**
Question 7 on the DAT	0.186*
Question 14 on the DAT	0.214**
Total score on the DAT	0.186*

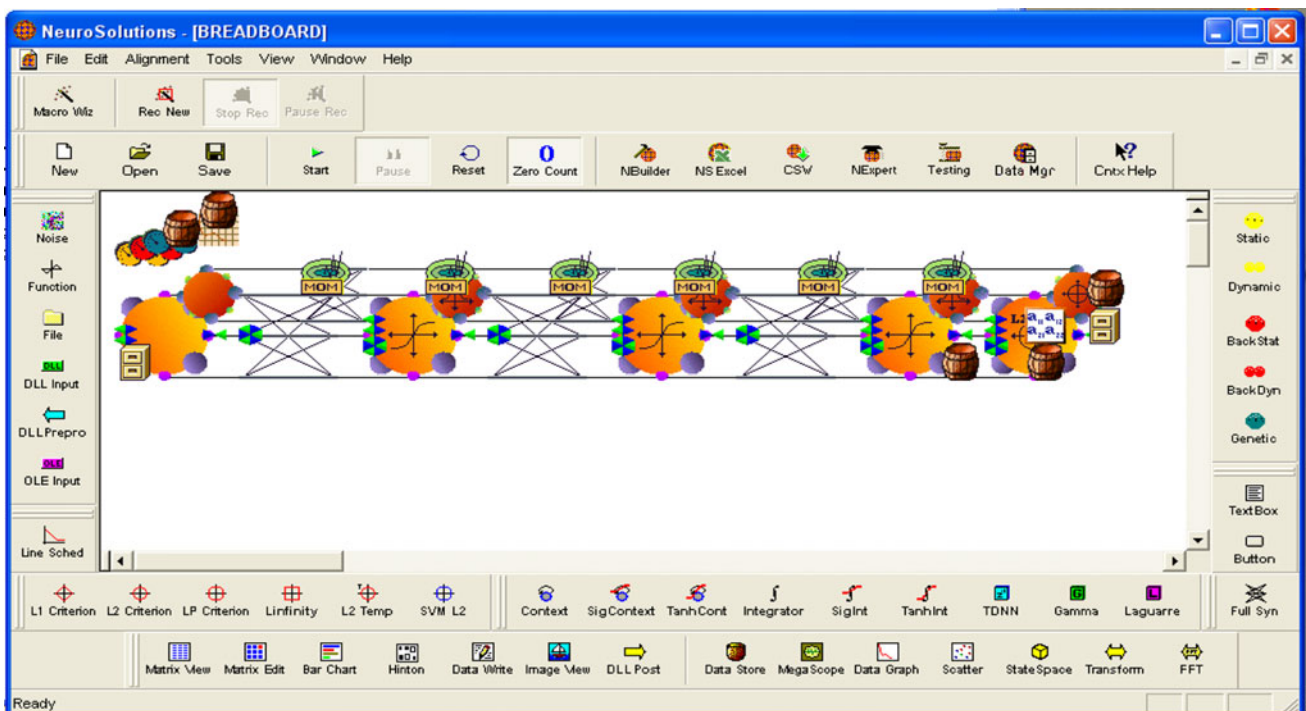
\* Correlation is significant at the 0.05 level (2-tailed)

\*\* Correlation is significant at the 0.01 level (2-tailed)

breadboard approach stems from the company's software engineers' background in electrical engineering. In regards to neural network development, most of the networks components can be added via a "drag-and-drop" fashion in a fairly easy-to-use interface. With a Microsoft Excel add-in for data entry and manipulation, researchers interested in utilizing and testing neural networks as a data mining technique can minimize the computational complexities associated with neural network creation and create an accurate predictive system relatively quickly. A screenshot of Neurosolutions development platform is shown in Fig. 2.

Multilayered perceptrons are the most commonly used neural network architecture in data mining (Turban et al. 2006). The popularity of this neural network approach is due in large part to the mathematical proof that multilayer backpropagation networks are universal approximators (Hornik et al. 1989). A universal approximator is a numerical technique capable of finding a mapping between inputs and outputs if such a mapping exists.

Neurosolutions also offers two algorithmic features to optimize the parameters necessary to create a neural network and to maximize the use of all the data collected for both training and testing. First, one of the algorithmic features of Neurosolutions is the capability for genetic algorithm optimizations. Genetic algorithmic optimization provides researchers the ability to determine best the values of the numerous parameters/weights that are involved in creating a network. A processing element is a simple mathematical model of a biological neuron. Each processing element creates a weighted sum of inputs and individually becomes activated or not via a threshold function. The training process in neural network creation takes predetermined outcome data and algorithmically determines these weights via the backpropagation algorithm. A set of processing elements creates a layer within a neural network. Neural network creation began with a network with an overabundance of processing elements in each of its layers (i.e., 15 in the first hidden layer and 12 in second hidden layer). This allowed for genetic optimization



**Fig. 2** Neurosolutions user interface

to reduce the number of processing elements. The genetic optimization setup used to create the best network in this research is shown in Fig. 3.

The neural network accuracy across the number of generations is given in Fig. 4. This figure shows that genetic algorithms helped optimize the network parameters, dramatically decreased the mean squared error (MSE), and improved overall predictive accuracy. A steepest descent algorithm with momentum was used to avoid local minima and saddle points that might prevent the minimization of error.

In addition to genetic optimization, the second algorithmic feature Neurosolutions offers is the Leave-*N*-Out algorithm. Leave-*N*-Out allows researchers to utilize all of their data for both training and testing via a sliding cross validation window. This feature is especially useful when data is a scarce commodity, which is the case in this research. For the predictive network created in this research, *N* = 1 was used. Selecting *N* = 1 allows nearly all of the data, except for one, to be used for training from the very beginning of the training cycle. Again, training a neural network determines the weights within each of the processing elements and thereby gives the network its predictive power. The weights resulting from the training cycle create the predictive capability of the network. Using the neural network parameters found via the genetic

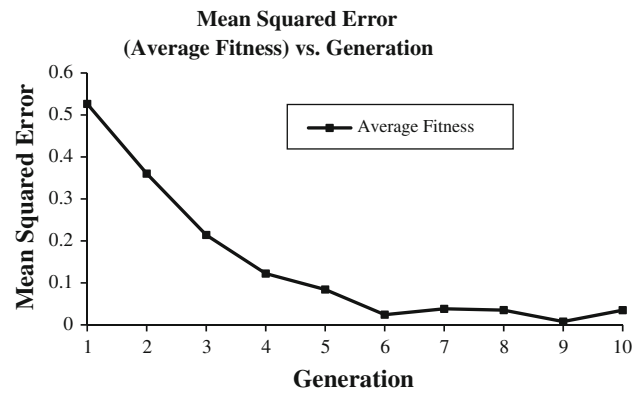


Fig. 4 Mean squared error (average fitness) versus generation

algorithm optimization process (e.g., eight processing elements in the first hidden layer and nine in the second hidden layer), Leave-*N*-Out with *N* = 1 was applied. Figure 5 depicts the mean squared error versus the epoch number.

From the learning curve depicted in Fig. 5, learning appears to have levelled off towards the end of the learning process. This suggests the network has fully learned from the data it was given, and has not over-modelled the specific data presented to it during training. This keeps the network generalizable and able to provide accurate predictions for out-of-sample data (i.e., data not used during training).

An input space of nine variables is not an unreasonable number in order to create a generalizable neural network which can perform in a production setting. In practice, however, the number of inputs should be minimized in order to have the most generalizable results (i.e., accurate predictions when presented with new data). Hence, backwards elimination was chosen as a means to possibly prune the number of inputs for the neural network. Backwards elimination entails removing the least statistically

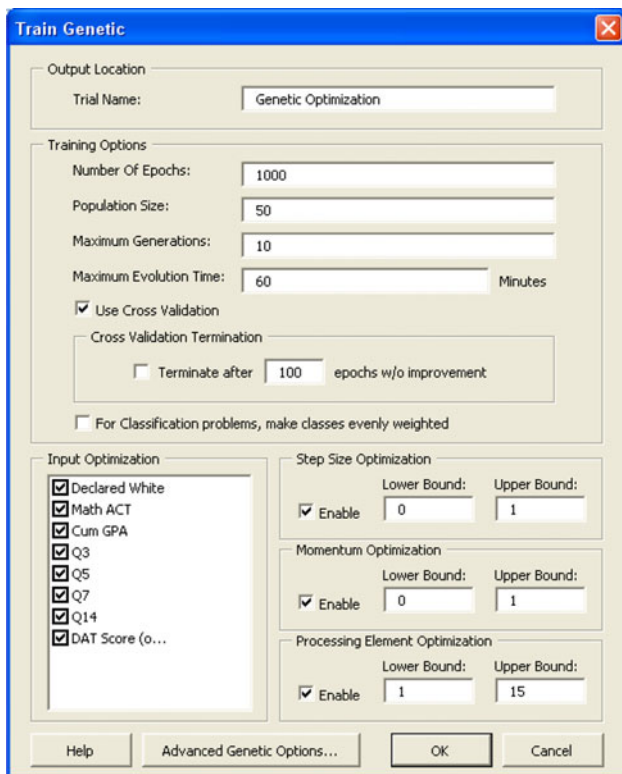


Fig. 3 Genetic algorithm optimization

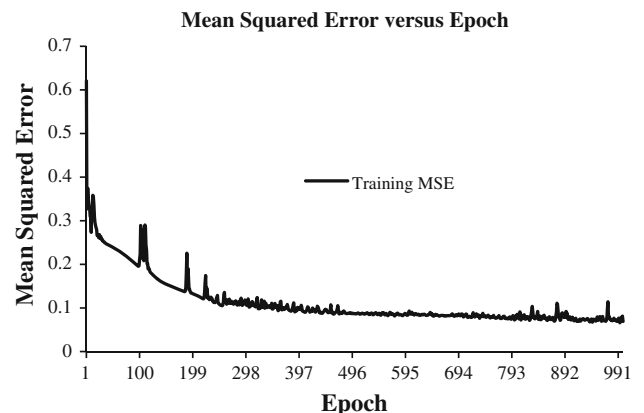


Fig. 5 Mean squared error versus epoch

significant variable, recreating the neural network, and determining the new accuracy. If the accuracy improves, the input is eliminated as a predictor. Otherwise, the input is kept.

## Results

Before using backwards elimination, a baseline accuracy using all nine of the statistically significant variables from Table 3 needs to be found. Using all nine input variables, the system correctly predicted  $76\% = \frac{34}{45}$  of the at-risk students who did not pass the course,  $85\% = \frac{121}{142}$  of the not at-risk students who passed the course, and correctly categorized approximately  $83\% = \frac{34+121}{187}$  of all chemistry students. These findings are summarized in Table 4, also referred to as a confusion matrix. In reference to sensitivity, the predictive system identified 45 students to be at risk, while 11 of these students were actually not at risk. In other words, a Type I error occurred  $24\% = \frac{11}{45}$  of the time. In reference to specificity, 142 of the participants were categorized by the system as being not at risk, while 21 were actually at risk. In other words, a Type II error occurred  $15\% = \frac{21}{142}$  of the time (Table 4).

## Discussion

The ability to categorize correctly approximately 83% of general chemistry students at the very beginning of the semester would be a great asset to chemistry professors. It is the authors' opinion that all nine predictors be kept as inputs to the neural network. The overall predictive accuracy remained near 80% regardless of the eliminated input. However, the accuracy for identifying at-risk students all dipped below 70% when an input was eliminated, thereby substantially increasing the chance of a Type I error—the error of greatest importance given the desire of this research to identify at-risk students and thereby increase student success (Table 5). In fact, eliminating either the total score on the DAT or cumulative GPA reduced the predictive accuracy for at-risk students from 76 to 60%. This suggests that the total score on the DAT is as good a predictor as cumulative GPA, which is significant because cumulative GPA is historically one of the best predictors (Cooper 2007, 2009). Creating an effective treatment for the purposes of mitigating a classification of being at risk could possibly be done via an item analysis of the nine predictive variables, as in Table 6.

The validity and reliability of the instrument and overall predictive system need to be addressed. First, in regards to

**Table 4** Confusion matrix showing accuracy and error for the DAT

	At-risk students (C- or below)	Not at-risk students (C or above)	All students	Accuracy of prediction	Error of prediction
Predicted to be at risk	34	11	45	$76\% = \frac{34}{45}$	$24\% = \frac{11}{45}$
Predicted not to be at risk	21	121	142	$85\% = \frac{121}{142}$	$15\% = \frac{21}{142}$

**Table 5** Changes in predictive accuracy of the DAT when one predictor is eliminated during backwards elimination

	At-risk students (%)	Not at-risk students (%)	All students (%)
All nine predictors (baseline accuracy)	76	85	83
Eliminated predictor	New predictive accuracy (%)	New predictive accuracy (%)	New predictive accuracy (%)
Declared white	62	90	83
Math ACT score	67	89	83
Cumulative GPA	60	80	75
Question 3 on the DAT	69	90	85
Question 4 on the DAT	55	85	78
Question 5 on the DAT	62	80	76
Question 7 on the DAT	62	82	78
Question 14 on the DAT	69	83	80
Total score on DAT	60	85	79

**Table 6** Possible at-risk mitigation approaches for a given predictor

Predictor	Possible treatment
Declared white	Additional support/funding for tutoring in the Native American Center (NAC) on campus. Recruitment and training of well-qualified tutors for the NAC. Develop and support TRIO programs, Ronald McNair programs, Upward Bound, and other programs for individuals from disadvantaged backgrounds
Math ACT	Raise college admissions standards
Cumulative GPA	Offer a concurrent support class that covers basic college study skills such as note taking, time management, and motivation
Questions on the DAT	Develop a concurrent support course that covers algebra topics in a “just-in-time” fashion. Develop practice algebra questions called “think aheads” that are distributed online and used by all students taking general chemistry. Coordinate timelines for math skills taught in algebra classes with math skills used in general chemistry classes

validity, due to the “out-of-sample” treatment of data from the Leave-*N*-Out algorithm, the found predictive accuracies are believed to be similar to what would be found when actually performing classifications of actual general chemistry students. This provides this research a high level of criterion-related validity. Second, in regards to reliability, using all nine statistically significant inputs, a Cronbach’s Alpha of 0.70 was computed using SPSS, which is an acceptable level of reliability. All nine inputs collectively appear to be measuring a similar characteristic, the at-risk nature of a student in general chemistry.

## Conclusions

The predictive system created in this research correctly identified 76% of at-risk students who did not pass general chemistry, a result which is approximately 35% better than other reported results (see Table 1). Given the behavioral nature of the data and context, the diagnostic algebra test created within this research coupled with a neural-network based predictive system offers faculty and others interested in increasing student success rates in chemistry a viable means to identify at-risk students. The Type I and Type II errors, reliability, and validity all seem to be within acceptable ranges. The metric of greatest concern might be the 24% likelihood of a Type I error, in which students are identified as being at risk who are actually not at risk. However, these misidentified students will not be harmed by participating in any academic success treatments deemed appropriate by their chemistry faculty members. In fact, these students are likely to be on the cusp of success, and their participation will only help them advance academically. Students who are identified as at-risk might perceive themselves as a negatively stereotyped and underperform as a result. Treatments for this could include exercises in self-affirmation of values and attitude-changing strategies that frame adversity as common and transient (Cohen et al. 2009; Miyake et al. 2010; Walton and Cohen 2011).

With reasonable Type I and Type II error categories, the only other concerns regarding use of this predictive system might be validity and reliability. In regards to validity, the relatively high accuracies seen via the Leave-*N*-Out algorithm gives this system a high level of criterion related validity. Data used via Leave-*N*-Out is considered to be out-of-sample, thereby making the accuracies indicative of what is to be found in a production situation. By using Leave-*N*-Out, it is the authors’ opinion that a predictive system has been created that accurately identifies at-risk students. Thus, given the acceptability of Type I error, Type II error, reliability, and validity, the predictive system presented in this research offers a viable and potentially worthwhile means to improve student success in chemistry.

The success of the predictive system developed in this study for the purpose of correctly identifying students at risk for failure in general chemistry is a result of several factors. First, the predictive system uses a genetically optimized neural network to determine weights for the inputs that give the highest predictive accuracy instead of leaving it to humans to make poorly informed guesses at appropriate weights and threshold values. Second, the DAT was created jointly by professors in math and chemistry who drew on their extensive experience to choose questions that not only appear often in chemistry, but also target common mistakes they have seen made by students in the past who tended to fail. Third, the questions on the DAT that were most predictive were also among the most basic algebra questions, which might help explain why the neural network predictive system identified at-risk students who will eventually fail better than other predictive systems that do not include such basic algebra questions.

The authors believe this research sets the foundation for a potential pilot of a production setting. The predictive system developed here could still be improved by replacing some questions on the DAT by ones that are statistically significant, have high discrimination indices, and high predictive accuracy as measured by a neural network analysis. With these improvements, it is possible that the number of inputs to the neural network could be reduced.



Reducing the number of inputs would simplify data collection, reduce reliance on data that might not be easily obtained, and, most importantly, should increase the predictive power of the system. The logical outcome of this research is to create and implement treatments for students at risk for failing general chemistry.

**Acknowledgments** This work was partially supported by a U.S. Department of Education Title III Strengthening Institutions Program grant #P031A080188. The authors would like to thank Professors Amy Getz and Robert Milofsky of Fort Lewis College for their help creating the diagnostic algebra test. Also, the authors would like to acknowledge some of the initial cleaning and coding data work performed by Fort Lewis College computer science student, Skylar Bolton. Skylar's work was supported by the Undergraduate Student Research, Scholarly and Creative Activities Grant Program at Fort Lewis College.

## References

- Bird L (2010) Logical reasoning ability and student performance in general chemistry. *J Chem Educ* 87(5):541–546
- Bunce DM, Hutchinson KD (1993) The use of the GALT (Group Assessment of Logical Thinking) as a predictor of academic success in college chemistry. *J Chem Educ* 70(3):183–187
- Burke BA, Bowen RJ, Casalnuovo J, Walton ED, Hiemenz PC, Millner CN (1999) Using assessment in general chemistry curriculum redesign. *Assess Update* 11(2):1–2 & 12–13
- Campbell J (2011) Examples of Information on chemistry course failure rates. <http://www.a-s.clayton.edu/campbell/DeptInfo/deptmeetings/chemfailure%20rates.doc>. Accessed 31 Mar 2011
- Cohen GL, Garcia J, Purdie-Vaughns V, Apfel N, Brzustoski P (2009) Recursive processes in self-affirmation: intervening to close the minority achievement gap. *Science* 324(5925):400–403
- Cooper CI (2007) A neural network-based decision support system for identifying and remediating at-risk students in developmental mathematics. Ph.D. dissertation, Walden University, United States—Minnesota. Retrieved 25 Feb 2010, from Dissertations & Theses: Full Text. (Publication No. AAT 3263504)
- Cooper CI (2008) Predicting persistence of college freshmen using neural networks. In: Presentation at the international joint conferences on computer, information, and systems sciences, and engineering
- Cooper C (2009) Predicting first-year retention of college freshmen using neural networks. In: Proceedings of the 2009 American Society for Engineering Education Northeast conference
- Hardgrave BC, Wilson RL (1994) Predicting graduate student success: a comparison of neural networks and traditional techniques. *Comput Oper Res* 21(3):249–264
- Hornik K, Stinchcombe M, White H (1989) Multilayer feedforward networks are universal approximators. *Neural Netw* 2(5):359–366
- Hovey NW, Krohn A (1963) An evaluation of the Toledo chemistry placement exam. *J Chem Educ* 40(7):370–372
- Karpp ER (1995) Paths to success, volume III. Validating the California chemistry diagnostic test for local use; ERIC document reproduction service no. 384 402. Glendale Community College, CA, Aug 1995
- Kimbrough DR, Weaver GC (1999) Improving the background knowledge of non-traditional students. *Innovative Higher Educ* 23(3):197–219
- Leopold DG, Edgar B (2008) Degree of mathematics fluency and success in second-semester introductory chemistry. *J Chem Educ* 85(5):724–731
- Lewis JE, Jiang B, Xu X, Garcia A (2010) Comparing two tests of formal reasoning in a college chemistry context. *J Chem Educ* 87(12):1430–1437
- McFate C, Olmsted J (1999) Assessing student preparation through placement tests. *J Chem Educ* 76(4):562–565
- Miyake A, Kost-Smith LE, Finkelstein ND, Pollock SJ, Cohen GL, Ito TA (2010) Reducing the gender achievement gap in college science: a classroom study of values affirmation. *Science* 330(6008):1234–1237
- Moorpark College Chemistry Department Placement Exam (2010) [http://sunny.moorparkcollege.edu/\\_chemistry/pretest/diagnostic\\_exam.pdf](http://sunny.moorparkcollege.edu/_chemistry/pretest/diagnostic_exam.pdf). Accessed 11 June 2010
- Naik B, Ragothaman S (2004) Using neural networks to predict MBA student success. *Coll Stud J* 38(1):143–150
- Neural Solutions (2010) Neural networks. <http://www.neurosolutions.com/>. Accessed 31 Mar 2011
- Niedzielski RJ, Walmsley F (1982) What do incoming freshmen remember from high school chemistry? *J Chem Educ* 59(2):149–151
- Ozsgomonyan A, Loftus D (1979) Predictors of general chemistry grades. *J Chem Educ* 56(3):173–175
- Pickering M (1975) Helping the high risk freshman chemist. *J Chem Educ* 52(8):512–514
- Pienta NJ (2003) A placement examination and mathematics tutorial for general chemistry. *J Chem Educ* 80(11):1244–1246
- Russell AA (1994) A rationally designed chemistry diagnostic test. *J Chem Educ* 71(4):314–317
- Schumacher P, Olinsky A, Quinn J, Smith R (2010) A comparison of logistic regression, neural networks, and classification trees predicting success of actuarial students. *J Educ Bus* 85(5):213–245
- Spencer HE (1996) Mathematical SAT test scores and college chemistry grades. *J Chem Educ* 73(12):1150–1153
- Turban E, Aronson JE, Liang TP (2006) Decision support systems and intelligent systems, 8th edn. Prentice Hall, Upper Saddle River
- Wagner EP, Sasser H, DiBiase WJ (2002) Predicting students at risk in general chemistry using pre-semester assessments and demographic information. *J Chem Educ* 79(6):749–755
- Walton GM, Cohen GL (2011) A brief social-belonging intervention improves academic and health outcomes of minority students. *Science* 331(6023):1447–1451