Paper ID #26839

Evaluation of Improvements in Visualization Test Scores Using Predictive Analytics

Dr. Jorge Rodriguez P.E., Western Michigan University

Professor in the Department of Engineering Design, Manufacturing, and Management Systems (EDMMS) at Western Michigan University's (WMU). Co-Director of the Center for Integrated Design (CID), and currently the college representative to the President's University-wide Sustainability Committee at WMU. Received his Ph.D. in Mechanical Engineering-Design from University of Wisconsin-Madison and received an MBA from Rutgers University. His B.S. degree was in Mechanical and Electrical Engineering at Monterrey Tech (ITESM-Monterrey Campus). Teaches courses in CAD/CAE, Mechanical Design, Finite Element Method and Optimization. His interest are in the area of product development, topology optimization, additive manufacturing, sustainable design, and biomechanics.

Dr. Diana Bairaktarova, Virginia Tech

Dr. Diana Bairaktarova is an Assistant Professor in the Department of Engineering Education at Virginia Tech. Through real-world engineering applications, Dr. Bairaktarova's experiential learning research spans from engineering to psychology to learning sciences, as she uncovers how individual performance is influenced by aptitudes, spatial skills, personal interests and direct manipulation of mechanical objects.

Evaluation of Improvements in Visualization Test Scores Using Predictive Analytics

Abstract

Spatial visualization skills have been long identified as critical competence for success in STEM disciplines, particularly in engineering and technology fields. Several initiatives to improve these skills have been implemented at various academic institutions. This study aims to apply data analytics (DA) to generate a predictive model for improvement of scores in a commonly used spatial visualization test. This model is based on pre- and post- scores by first-year engineering students, and the objective is to identify the factors that have the largest influence on the improvement of the scores. The generated predictive model provides information on dominant factors, i.e., specific questions in the test or demographics, that will help in establishing pedagogical activities aimed at improving spatial skills of students.

The dataset used in this study is from a college of engineering's incoming class, who are required to take the visualization test, and then are offered a one-credit course. Initial analyses are for initial validation of the generated model by comparing the results to previously generated ones. Three different definitions of improvement are used in this study, i.e., raw score, percentage, and tier, given that particular objectives might be different. Results from this study are in line with one observed in previous reports, with an overall test performance improvement, and more involved test question being more influential factors. Similarly, some of the results involving demographic factors follow in a limited fashion previously observed trends. This study shows that DA is a useful tool that will help in the search for specific objective information regarding the value of activities aimed at improving spatial visualization skills.

Introduction

There has been an increasing trend at universities in the USA regarding the use of spatial visualization skills indicators as predictors of successful performance by students in technical programs. Several initiatives to improve such skills have been implemented at various academic institutions, with some positive overall results. A main goal pursued with those initiatives is to improve the students' spatial visualization skills, as measured by their scores in a standardized visualization test, by comparing the pre-initiative and the post-initiative performance of the students. Thus having as well an indication of the effectiveness of the implemented initiative.

There are several tests that have been applied to measure spatial visualization skills of students [1, 2], and there are numerous studies that have collected and analyzed information regarding demographics, spatial visualization skills, and academic performance [3, 4]. Of interest are studies where spatial visualization skills have been linked to abilities to do engineering and technology work, and subsequent studies that have provided a relationship between those skills by students and their performance in engineering courses, particularly for engineering graphics and design courses [5]. Additionally, there are reports that indicate the importance of improving spatial visualization skills when looking at students' performance in technology and engineering courses [6]. Other reports indicate the value of improving such skills as the complexity of the problem increases [7], which is one of the reason to take a closer look at pre- and post- scores in a standardized test such as Purdue Spatial Visualization Test with Rotations (PSVT:R) [8]. The PSVT:R consists of 30 questions with increasing degree of difficulty in terms of number and sequence of spatial rotations that need to be applied to a 3D object in order to end with a desired configuration.

Predictive analytics techniques are being applied in order to extract any potential trends in the dataset being utilized in this study. This modeling will help in the identification of factors, question number or demographic, that have significant impact in the prediction of test score improvement. Predictive analytics is a topic in Data Analytics (DA), which is a generic term used to refer to a set of quantitative and qualitative approaches that are applied to provide the basis for some decision making [9]. Typical objectives that are usually pursued when performing modeling with DA techniques are identification of options/factors to increase productivity, boost business profit, or accomplish a given behavior or performance [10]. Predictive analytics is extensively used in business environments, particularly consumer sciences and where service/ product customization is pursued. It is as well an approach that has gained acceptance in its application to engineering and technical problems. There have been some applications in academic settings, but its application on pedagogical approaches is something novel with high potential.

The software package used in this study is RapidMiner, a commercially available DA software that offers different approaches for the analysis and visualization of datasets, thus allowing comparison of the results, and some optimization. Data analytics techniques help in the identification of dominant factors in a dataset that result in the prediction of a specific performance or behavior. In this study it means identification of dominant questions in the standardized spatial visualization test and/or demographic parameters, that have a direct positive impact in test score improvement by students. The DA technique being utilized in this study is Decision-Tree, which has been identified as a good general purpose algorithm, with acceptable reliability in predictions models. This approach allows for graphical output that is very helpful in envisioning the predictive model that is developed [11]. A decision-tree is a collection of nodes in a root-branch sequence that defines selection paths based on specific class or numerical value of selected parameter (e.g., final test score). Each node represents a splitting rule for one specific attribute (e.g., answer to a test question). This analytic tool has as well the option to reduce predictive errors by searching for an optimal decision-tree development, according to a specified criterion [12].

The objective in this study is to search for dominant factors that predict positive test score improvement when comparing pre-intervention to post-intervention evaluation of students' spatial visualization skills. Another goal is to identify influential test question(s) and/or demographic factors that will move the predictive modeling efforts into a broader identification and grouping of attributes (i.e., subsets of questions). Results from this study will serve in the definition of pedagogical approaches to follow in the initiatives aimed at improving spatial visualization skills. Of interest as well is that results from the predictive models obtained for each subset of data (i.e., pre- and post-) will be compared to the results of the predictive model for score improvement, with the expectation that possible relationships can be established, thus having a more robust pedagogical intervention.

Methodology and Results

The DA approach followed in this study follows a decision tree algorithm, which is a recursive partitioning scheme that does data mining by identifying decisions that segment the dataset, thus identifying influential factors [13 – 15]. This approach is similar to the ones utilized in previous studies where predictive models had been generated to identify dominant parameters for test score estimation. The PSVT:R test was utilized, and it was administered to first-year students coming into the College of Engineering at Virginia Tech. The dataset consists of demographic information and answers to PSVT:R questions by students in the class, a data subset is from the administration of the test before the start of the improvement initiative (pre-), and another data subset is from the administration of the test at the end of the initiative (post-). The initiative in this case is a semester long one-credit course offered to first-year engineering students to improve their spatial visualization skills. The test was administered online during a lecture session, in both occasions.

Table I. General Information for Dataset Utilized.

Number of students		185	
Gender			
	Female	89	48.1%
	Male	96	51.9%
Age			
	17	21	11.4%
	18	152	82.2%
	19	9	4.9%
	Other	3	1.6%
Ethnicity			
	African American	13	7.0%
	Asian	55	29.7%
	Caucasian	95	51.4%
	Hispanic	10	5.4%
	Other	12	6.5%
Pre-score			
	Average		16.394
	S.D.		4.632
Post-score			
	Average		18.319
	S.D.		4.341

The complete dataset consists of 185 valid cases, meaning students that have taken the pre- and the post- tests and received a score. The demographic information used in this study is gender, age, and ethnicity. Table I provides some basic descriptive statistics for the students in the dataset.

Pre- and Post-subsets

The first step was to generate a predictive model for the Pre- and the Post- results with the intention of doing some validation by comparing to the trends observed in previous studies where predictive analytics was applied to identify specific question(s) that predict a good score in the test (i.e., 'top score'). This model was based on the identification of test question(s) that are good predictor for top scores. In the previous pilot studies, a test score greater than 24 was identified as a top-score, which was as well evaluated in the range 21-30, with similar trends in most cases.

Results for Pre- and Post-scores

To have a better idea of the general performance by the participant students in this study, Figure 1 shows the Pre and the Post scores for each student, sorted out from smaller to larger on their Pre- score. It can be seen that there are students that performed better, students that performed worse, and some students that had the same score in both tests, with the corresponding average and standard deviation given in Table I.

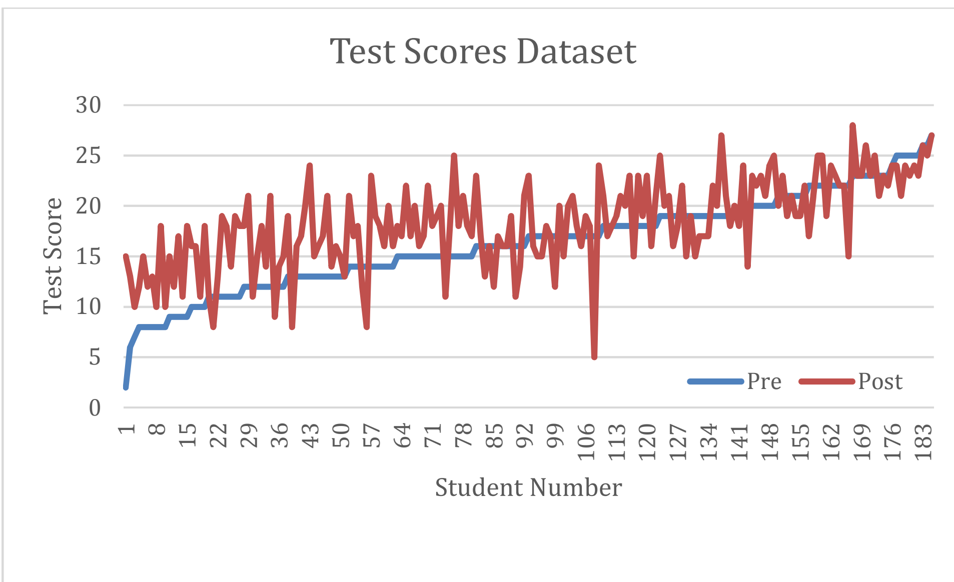
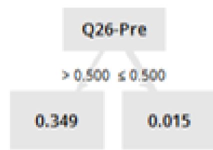
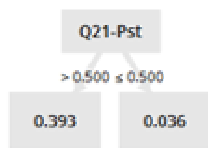


Figure 1. Pre- and Post- test scores for applicable dataset.

For the Pre-data, question 26 (Q26) was identified by the analytical tool, given its highest probability of prediction of a top-score, as the main influencer, and for the Post-data, question 21 (Q21) was identified as the main influencer (Figure 2). These results are in line with previously generated models where questions involving rotations about two or more axes are identified as the ones that dictate higher test performance. These tree models indicate that if, for example Q26 is answered correctly (>0.500 – a score of 1) then the model predicts an overall score close to the average for all students (0.349); otherwise (i.e., if Q26 is answered incorrectly (<0.500 – a score of 0) then the predicted value is significantly off mark.



Model for Pre-scores – Q26 most influential factor.



Model for Post-scores – Q21 most influential factor.

Figure 2. Predictive models are one-factor for Pre- and Post-scores.

Improvement Measurement

For the evaluation of improvement in performance, basically the difference between Pre- score and Post- score for each student is used. However, there were three different ways of identifying such difference as:

a) raw score increase (decrease)
b) percentage improvement
c) tier indicator of becoming top-scorer.

Each one of these measurements have value per se, and can be used in different situations to measure the improvement shown by the students. The more direct measurement is the first option, raw score, which is basically the Post-score minus the Pre-score; this is a valid indicator however it might misrepresent the actual improvement since a student with low score in the Pre-test has more room to get a high increase, which does not imply automatically that it is at the level of top-score. The second option is a popular technique that tries to minimize the effect of raw numbers, percentage improvement, however it might have some bias for the low-scorers since they might show huge percentage of improvement but not indicating that the new score is a top-score.

The third option was defined with basis on the ultimate objective of having improved visualization skills in order to have higher possibilities of doing a technical career. Therefore, it tries to capture if the Post- score is good enough to become a top-score. This indicator is the difference between the 'tier' were the Post-score is, compared to the 'tier' were the Pre-score was. Four tiers were defined in this calculation: Tier 1 – score higher than one Standard Deviation (SD) above average grade for the group; Tier 2 – score between average group score and one SD above; Tier 3 – score between average group score and one SD below; and Tier 4 –

score more than one SD below the average group score. Top scores are in Tier 1, and Table II summarizes the different values that give a better overall picture of the numbers being used under these measurements.

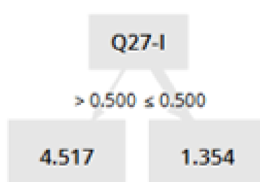
Descriptive statistical parameters for the three measurement of improvement are provided in Table II, illustrating the existence of students that improved, student that did not improve, and students that showed no variation in their test performance.

Table II. Summary of Improvement Measurements.

	Raw Score	Percentage	Tier
Average	1.924	19.460	-0.027
Standard Deviation	3.826	55.294	0.810
Minimum value	-12.0	-70.6	-2.0
Maximum Value	13.0	650.0	2.0

Results for Improvement Model

Predictive models were generated for each one of the improvement measurements in order to identify the most influential question(s). It yielded single factor (question) for two of the measurements: raw score and tier improvement (Figure 3). For raw score increase, question 27 (Q27) was identified as the most influential factor, and for tier improvement question 25 (Q25) is the best predictive factor. For the second measurement, percentage increase, there was no question that showed significant influence, i.e., multiple factors are needed for prediction. This might be due to the significant range of percentages in the dataset, i.e., one case of 650% increase. This particular case can be considered an outlier in the dataset and considered as invalid input, thus further investigation will be taking place in the future. As previously indicated, these results are in line with other generated models where question(s) involving rotations about two or more axes are identified as the one(s) that dictate test performance.



Model for Improvement Raw Score – Q27 most influential factor.

Figure 3. Predictive model is a one-factor for Improvement Raw Score.

Demographic Factors

Predictive analytics is applied to generate models where demographics factors are taking into account together with the students' responses in the standardized test. Three demographic factors were included: gender, age, and ethnicity. In each case, a model was generated that takes into account only one factor at a time, and then all three factors are included in the generation of another predictive model.

Results for Demographic Factors

Table III below shows the results that were obtained in each one of the established situations. The table includes results for models based on two improvement measurements: Raw Score and Tier Jump, given that the other improvement measurement - Percentage Improvement - did not yield valid models. In the table the symbol (~) indicates that the corresponding factor has an influence, but is not a primary or significant one. It is important to note that in reality all factors have some influence on the prediction model, with some of them having greater influence than other, thus resulting on the one-factor models given in aforementioned cases.

Table III. Effect of Demographic Data on Models.

	Improvement Type	
	Raw Score – Q27	**Tier – Q25**
Gender	No influence	~Positive (M) ~Negative (F)
Age	No influence	No influence
Ethnicity	No influence	~Positive (C) ~Negative (A) ~Negative (H)

This situation of being a secondary influential factor is illustrated with the graph in Figure 4, where the most influential factors are given for one of the cases in Table II, gender influence on Tier Improvement measurement model, where it has been mentioned that there is a one-factor model (Q25), as it is shown in Figure 4. Other factor (secondary ones), including gender do have some influence, but not a significant one to become predictive factor.

In this particular case as well there is a negative effect of gender female on the Tier Improvement measurement as indicated for the red label "Contradicts Prediction" (it is a positive effect – green label for Male gender) indicating that the gender factor does support an improvement in Tier for male students, but does not support it for female students.

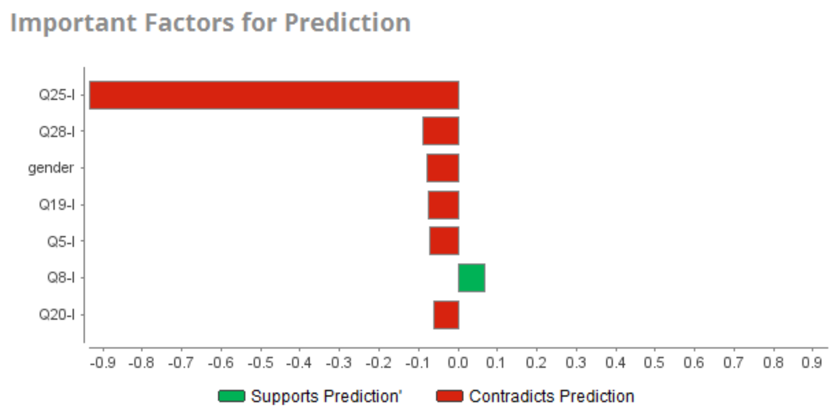


Figure 4. List of influential factor for predictive model with gender data.

Comments and Conclusions

The results of this study provide some interesting information on what should be expected from academic activities geared towards the improvement of spatial visualization skills in students, and therefore on what can be expected in terms of their performance in technical curricula. In general, the results are in line with those observed for situation where an intervention has been implemented: there is an overall improvement of scores, with some specific factors (questions) being more influential than others. Particularly, single factors were identified for two of the measurement of improvement used: Raw and Tier, but multiple factors are needed for a predictive model using Percentage as improvement indicator. It is of interest as well that the one-factor models identify question that belong to the same subset, questions that require rotation about at least two axes in order to have a correct answer. Thus reflecting the importance of such exercises for spatial visualization skills.

In terms of demographic parameters, the results indicate that they do not have a significant influence in the improvement of test scores. Important to note is that the implication is that multiple factors are needed to have a predictive model. Additionally, it is important to note that such result is not an effect because of the use of data analytics. The application of statistical analysis results in similar conclusions that the ones showing up with DA. The results from the improvement predictive model indicate that there is no significant effect by demographic factors, and performing ANOVA on the dataset produces values of probability that indicate invalid hypothesis (i.e., $p \geq 0.05$). Specifically, two-way ANOVA resulted in a $p = 0.694$ when gender is considered, $p = 0.350$ for age, and $p = 0.665$ for ethnicity. is considered, which certainly is a factor but having non-significant effect in the generation of the prediction models.

It is important to note that there are couple of instances when the dataset utilized had less than 185 cases. The first instance was during initial modeling when two subsets of data are used, one to generate the supervised model, and the second one to test (and adapt) the actual model that was generated. The other instance was when incomplete information for a student was available, e.g., not all demographic information was provided, which resulted in either assigning a value of 'other' and/or eliminating the record from the dataset. There are DA schemes that can be applied to obtain, perhaps, better models, e.g., identification and removal of outliers in the data, but it is something that needs to be properly evaluated to avoid influencing the models.

The work presented shows that DA is a useful tool that helps in the search for objective information regarding what is producing acceptable results, and what is not, which goes towards the ultimate objective of defining specific pedagogical interventions that will have the most impact on spatial visualization skills.

References

[1] S. Strong and R. Smith, "Spatial Visualization: Fundamentals and Trends in Engineering Graphics", *Journal of Industrial Technology*, vol. 18, no. 1, 2002.

[2] J. Yue, "Spatial Visualization by Realistic 3D Views", *Engineering Design Graphics Journal*, vol. 72, no. 1, winter 2008.

[3] G. Prieto and A. D. Velasco, "Does spatial visualization ability improve after studying technical drawing?", *Quality and Quantity*, Research Note, Springer, 2010.

[4] S. A. Sorby, "Developing 3-D Spatial Visualization Skills", *Engineering Design Graphics Journal*, 63(1), 21-32, 1999.

[5] S. A. Sorby, "Assessment of a New and Improved Course for the Development of 3-D Spatial Skills", *Engineering Design Graphics Journal*, 69(3), 6-13, 2005.

[6] M. Kozhevnikov and R. Thornton, "Real-Time Data Display, Spatial Visualization Ability, and Learning Force and Motion Concepts", *Journal of Science Education and Technology*, vol. 15, no. 1, Springer Science & Business Media, 2006.

[7] S. Titus and E. Horsman, "Characterizing and Improving Spatial Visualization Skills", *Journal of Geoscience Education*, vol. 57, no. 4, 2009.

[8] R. Guay, Purdue Spatial Visualization Test – Visualization of Rotations. West Lafayette, IN, Purdue Research Foundation, 1977.

[9] Big Data Analytics, "What it is and why it matters", https://www.sas.com/en_us/insights/analytics/big-data-analytics.html, 2017.

[10] Data Analytics, "Make Better Business Decisions", https://www.qlik.com/us/lp/ppc/qlik-sense-desktop/business-analytics?sourceID1=google&Campaign_Type=Non-Brand&KW=data%20% 2620analytics&k_clickid=24c8bad7-5950-446c-9aa6-68027642b4fd&gclid=EAIaIQobChMIlraIe6V2QIVT7nACh39LAGpEAAYAiAAEgJR4_D_B wE, 2017.

[11] Best Data Analysis Tools, https://www.softwareadvice.com/bi/data-analysis-comparison, 2017.

[12] RapidMiner, User's Manual, https://docs.rapidminer.com/downloads/RapidMiner-v6-user-manual.pdf, 2017.

[13] J. Rodriguez and L. G. Rodriguez-Velazquez, "Comparison of Spatial Visualization Skills in Two Approaches to Entry-Level Graphic Courses", Proceedings of ASEE Annual Conference, 2016.

[14] J. Rodriguez and L. G. Rodriguez, “Comparison of Spatial Visualization Skills in Courses with Either Graphics or Solid Modeling Content”, Proceedings of ASEE-EDGD Mid-year Conference, 2016.

[15] J. Rodriguez and L. G. Rodriguez-Velazquez, “Application of Data Analytics Approach to Spatial Visualization Test Results”, Proceedings of ASEE Annual Conference, 2018.