DECISION SCIENCES
JOURNAL OF INNOVATIVE EDUCATION

Decision Sciences Journal of Innovative Education
Volume 13 Number 4
October 2015
Printed in the U.S.A.
© 2015 Decision Sciences Institute

EMPIRICAL RESEARCH

Required Collaborative Work in Online Courses: A Predictive Modeling Approach

Marlene A. Smith
University of Colorado Denver, Campus Box 165, PO Box 173364, Denver, CO 80217-3364, e-mail: ma.smith@ucdenver.edu

Deborah L. Kellogg†
University of Colorado Denver, Campus Box 165, PO Box 173364, Denver, CO 80217-3364, e-mail: Deborah.Kellogg@ucdenver.edu

ABSTRACT

This article describes a predictive model that assesses whether a student will have greater perceived learning in group assignments or in individual work. The model produces correct classifications 87.5% of the time. The research is notable in that it is the first in the education literature to adopt a predictive modeling methodology using data collected via a designed experiment. All subjects experienced both a collaborative and an individual assignment, thus mitigating uncontrolled external factors in the measurement of differences in perceived learning. The exploratory nature of the work prompted the use of Partial Least Squares Regression for estimation. The work serves as an illustration of how predictive modeling might enlighten those in educational and academic settings.

Subject Areas: Collaborative teaching, Curriculum design, Online education, and Pedagogy.

INTRODUCTION

The purpose of this article is to further our understanding of the nature of online collaborative work in the context of required group assignments in an adult learning population. The subject is worthy of scrutiny because evidence of the efficacy of collaborative work in online education is inconclusive. This leaves a gap in our understanding of how best to educate online students and design courses for them. Proponents of online group work cite enhanced critical thinking skills, higher level learning, increased socialization, enhanced reasoning and constructive learning, and increased student satisfaction as some of the benefits to collaboration (e.g., An & Kim, 2009; An, Kim, & Kim, 2008; Bernard, Rojo de Rubalcava, & St. Pierre, 2000; Cameron, Morgan, Williams, & Kostelecky, 2009; Ding, Chen, Knudson, & Braun, 2011; Lukman & Krajnc, 2012; Resta & Laferriere, 2007; Swan, 2001).

†Corresponding Author.

Moreover, collaboration is a reality of today's workplace, so requiring group work in online coursework helps prepare students for that eventuality.

Yet there are contradictory views. The meta-analysis of Lou, Abrami, and D'Apollonia (2001) found no consistent learning effects arising from group versus individual learning environments. Arbaugh (2005) studied both perceived learning and satisfaction with group assignments in online environments and found no association. Lukman & Krajnc, (2012) reported that virtual collaborative learners believed they learned no more from group assignments than from other more traditional learning methods, yet invested more time in collaborative assignments.

Other authors warn that any benefits that might accrue from online collaborations are not without cost (Kellogg & Smith, 2009). An et al. (2008) found that students often encountered difficulties with collaborative online work arising from cognitive problems, individual differences, grading system difficulties, scheduling troubles, and the communication challenges encountered in online learning environments. Cornelius, Gordon, and Ackland (2009) discuss the paradox of adult learners who desire individual flexibility and self-direction yet may also require social participation for maximum learning. The online environment can make collaborative work difficult (Bernard et al., 2000), thus competing with the flexibility needs of online learners (Kellogg & Smith, 2009).

Some of the contrariness of the academic findings can certainly be traced to the populations studied, methods used, and outcomes measured. Yet what are a developer and teacher of online courses to make of this miasma? Should student collaboration be made an integral component of online courses or not? Perhaps, to practitioners, that is the wrong question to be asking. Instead, perhaps we should seek to learn *which students* are likely to benefit from online collaboration and which are not. The distinction is important. To date, academic research involving the efficacy of online student collaboration has, rightly, focused on theory development and hypothesis testing using descriptive or explanatory methods. Marks, Sibley, and Arbaugh (2005), for example, used the sign and significance of a coefficient in a structural equation model to find in favor of a theorized positive relationship between the number of group projects and students' perceived learning. Importantly, they found that student/instructor interactions were stronger predictors of perceived learning than student to student interactions. That finding speaks to students in the aggregate. It is however not unreasonable to suspect that some students in the study perceived a benefit from group assignments while others did not. That distinction is best made using predictive models.

Predictive modeling is only now making inroads in educational and academic settings. Dziuban, Moskal, Cavanagh, and Watts (2012), for example, used logistic regression and classification and regression trees methods to estimate the probabilities of students withdrawing from, or receiving low grades at, the University of Central Florida. Arnold (2010) described Purdue University's Signals project that studied at-risk students at the course level. The multicampus PAR project (Ice et al., 2012) identified which among 640,000 students were likely to remain enrolled or graduate from their university. Campbell, DeBlois, and Oblinger (2007) described Baylor University's initiative that identified prospective enrollees using student inquiry information and another at the University of Alabama that detected students who were unlikely to return for their sophomore year. The common feature of these initiatives is the focus on students as individuals and not students in the aggregate.

The current research examines whether a predictive model might be developed to identify which students would benefit from online group work. It is the first in the distance learning literature to use a predictive modeling approach to address online collaborative learning. The focus is not on developing an interpretable model on which to conduct inference. Instead, a predictive modeling approach is used. Shmueli (2010) examined the myriad ways in which predictive modeling differs from explanatory or descriptive modeling. Selection of independent variables, for instance, proceeds quite differently. Predictor variables whose values are only known ex post cannot be used in predictive modeling, regardless of their theoretical relevance. In addition, "... it is sometimes the case that removing inputs with small coefficients, *even if they are statistically significant*, results in improved prediction accuracy" (Shmueli, 2010, p. 300). The approach taken here and the outcomes produced are thus distinctly different from those commonly found in the distance education literature that uses explanatory methods.

The data were collected via an experimental design. Unlike other observational studies in which different students faced different learning protocols, each student was subjected to two different pedagogical activities, a group assignment and another in which students were required to work independently. In this way, the data provide a sharper measure of differences in perceptions of group and individual work that is free of uncontrolled external factors.

This work is necessarily exploratory in nature. Because the application of predictive methods in educational research is uncharted, no guidance can come from the literature about independent variables that have produced successful predictive models. Even so, a candidate pool was assembled via a literature review of earlier descriptive and explanatory work, the topic of the next section of the article. The article then details the variables used in the study and describes the chosen statistical methodology. A reporting of the results is followed by discussion, limitations, and future research.

LITERATURE REVIEW

Many variables have been included in explanatory/descriptive models of online learning, serving as either direct or moderating effects. Given the present focus on predictive models, the literature review focuses exclusively on those variables that can be measured prior to assigning collaborative work without regard to their reported statistical significance.

Demographic data are often collected in learning research. Gender differences have been shown to be significant in several studies, with females performing better than their male counterparts in online learning environments (e.g., Chu, 2010; Rovai & Baker, 2005). Gender, age, and employment were cited by Park (2007) as factors that explain why students drop out of online programs. Ke and Xie (2009) considered age in their study of adult online learners and found age to be significant only with regard to the memorization of new facts and not for more integrative learning. Others have found both age and gender to be important to learning outcomes (e.g., Anstine & Skidmore, 2005). In their study of factors which might influence the motivation of adult online students to learn, Lim and Kim (2003) believed that marital status would have an effect; however it was not significant. The employment status of online students was considered by Park (2007) as employed adult learners have multiple roles which may conflict with their learning. More generally, Anstine and Skidmore (2005) argued that demographic information should serve as needed control variables in multivariate studies of learning outcomes.

User experience and perceived expertise with technology are often included in studies of online instruction. For example, computer skills and use are related to positive attitudes toward Internet-aided instruction, and this relationship is moderated by gender and age (Koroghlanian & Brinkerhoff, 2007). Volery and Lord (2000) found that computer experience had an interaction affect with gender in understanding online engagement. Kuo, Walker, Belland, and Schroder (2013) found that Internet expertise had a strong impact on satisfaction with online courses.

A student's individual learning style has also been linked to educational outcomes. Eom, Wen, and Ashill (2006) found that in online environments, *visual* learners had increased satisfaction and improved outcomes. By definition and in general, *active* learners enjoy working in groups while *reflective* learners prefer working alone (Felder & Spurlin, 2005). Reflective learners were also found to be more likely to enroll in online courses and complete them, and were more successful (Battalio, 2009; Doherty & Maddux, 2002). Sabry and Baldwin (2003) found that *global* learners preferred asynchronous online discussions whereas *sequential* learners preferred interacting with things rather than people. *Verbal* learners like to hear information and engage in discussion (Hawk & Shah, 2007), thus preferring discussions, particularly ones that include voice exchange (e.g., face-to-face meetings, telephone, or online conferencing).

Personality traits measured using the "Big Five" classification system have been found to have predictive value (Duff, Boyle, Dunleavy, & Ferguson, 2004; O'Connor, & Paunonen, 2007; Schnienderjans & Kim, 2005; Whittingham, 2006;). The "Big Five" personality dimensions include o*penness* (intellectual, independent-minded), *conscientiousness* (orderly, responsible), *extraversion* (talkative, assertive), *agreeableness* (good-natured, cooperative), and *neuroticism* (easily upset, nervous) (John & Srivastava, 1999). The use of personality traits is generally suggested to help form higher performing teams or groups (Bradley & Hebert, 1997; Thomas, Moore, & Scott, 1996). Correa, Hinsley, and Zuniga (2010) found that extraversion impacted Internet interactions for younger people whereas openness was important for older users.

The existing research thus suggests four broad categories of inputs for this study: demographics, technology proficiency, learning styles, and personality traits. The next section provides details about the measurements used for these and other variables. The data collection methodology is also described.

MEASURES

This study uses one outcome variable and 34 input variables. The Appendix provides details and summary information for all variables.

The Outcome Variable

Three measures of learning outcomes are widely found in the literature: actual learning (measured by grades), perceived learning, and satisfaction (e.g., Alavi, 1994; Arbaugh, 2005; Chitkushev, Vodenska, & Zlateva, 2014; Jiang & Ting, 2000). In this study, the dependent variable is the difference between learning outcomes from an independent assignment and a group assignment. Actual grades could not be used because differences in grades would measure the difference between a group grade, applied to all students in the group, and an individual grade earned a single student. Satisfaction follows and is correlated with perceived learning (Eom et al., 2006; Swan, 2001). Because each subject in the study completed five long surveys over the course of a 16-week term, reducing survey size was a serious consideration. Perceived learning was chosen as the outcome variable.

Perceived learning was measured using the seven-point Likert scale (1 = strongly disagree, 7 = strongly agree) developed by Alavi (1994) as reported in Marks et al. (2005). Items were edited only to reflect whether or not a given assignment involved group work. For example, "I improved my ability to integrate facts and develop generalizations from the course material" became either "*Working alone* improved my ability to integrate facts . . . " or "*Being in a group* improved my ability to integrate facts . . . " All seven items loaded on a single factor and the value of alpha was .975.

The dependent variable subtracts the centered and scaled perceived learning on the independent assignment from that of the group assignment. Negative values indicate that a student reported more learning from the independent assignment, while positive values indicate that more learning arose from the collaborative assignment. The values of this variable ranged from −5.14 to 5.29, with a mean of −.75. On average, students reported learning marginally more from independent work than from group work. The frequency distribution of scores is presented in Figure 1.

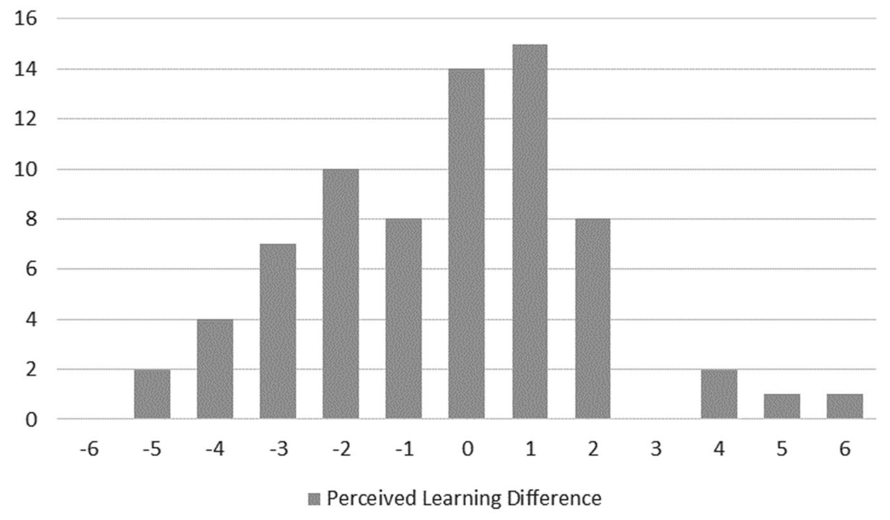
Input Variables

The literature review guided us toward four broad categories of potential independent variables. The authors' collective, but anecdotal, knowledge over two decades of online teaching experience contributed a fifth "other" group.

Learning Styles

Although recent literature questions the "learning style hypothesis" (e.g., Pashler, McDaniel, Rohrer, & Bjork, 2008) that holds that when students are paired with teaching methods that match their learning styles they will learn more or more efficiently, learning styles are included to elucidate preferences for different ways of interacting with other students or the material in an online environment. The Felder-Solomon Index of Learning Styles (ILS) described in Felder (1988) was selected to evaluate student learning styles. Papp (2001) found that this index had better explanatory value than other learning style scales. Students completed this questionnaire online at http://www.engr.ncsu.edu/learningstyles/ilsweb.html. Because only overall scores for each of the learning styles are reported for participants, it was not possible to compute validation measures. However, this scale

Figure 1: Frequency distribution—perceived learning difference.



is widely used and is well-validated (e.g., Felder & Spurlin, 2005; Litzinger, Lee, & Felder, 2007; Zywno, 2003). The ILS is scored on a continuum along four paired dimensions, Sensory/Intuitive, Visual/Verbal, Active/Passive, and Sequential/Global that represent learning style preferences. Scores can range from -11 to $+11$, and those between -3 and $+3$ indicate no preference (Felder & Spurlin, 2005). While the full range of preferences was found in this sample, there was no evidence of systematic bias for any one learning style; the means were all between ± 3.

Personality Traits

The Goldberg 50 question version of the International Personality Inventory Pool (IPIP) was used to assess the "Big Five" personality dimensions (http://ipip.ori.org/). This inventory is also widely used and well-validated (e.g., Gow, Whiteman, Pattie, & Deary, 2005; Lim & Ployhart, 2006; Socha, Cooper, & McCord, 2010). The Goldberg version of the IPIP is scored on a five-point Likert scale. Factor analysis was performed, and as a result, five survey items were removed as they did not load on any of the defined dimensions. The final alpha values, found in the Appendix, are consistent with others found in the literature. Our sample of students was slightly more open, conscientious, extraverted, and agreeable, and slightly less neurotic but not to a degree that would indicate bias.

Demographics

Standard demographic information was obtained including age, gender, marital status, number of dependents, family income, and whether a student was employed. In our sample, 57.9% of students were at least 30 years old, 56.9% were female, almost all (91.7%) were employed, 59.7% were married, and 40.3% had salaries no higher than $75K. These results are consistent with those of other surveys conducted of the students in the program and reflect the nature of working adult students in a graduate business program.

Technology Proficiency

Proficiency with technology was assessed using 13 self-reported measures and one computed measure. The items and anchors used in this study have been used in existing research (e.g., Arbaugh, 2000; Campbell & Williams, 1990; Cassidy & Eachus, 2002; Martin, 2008; van Braak, 2004). Six questions asked subjects to rate, from never-used to advanced, their proficiency with, for example, the use of computers, the Internet, and online communications. Six additional questions were included to capture the frequency of use of various technologies such as the hours spent per day using a computer for work and the number of online courses taken. The number of hours worked on a computer per day for work, school, and personal tasks were added together. Students reported using a computer for an average of 11 hours per day and spending an average of 24 hours per week using the Internet. All of them rated their proficiency with the Internet as moderate to advanced.

Other

The data set also included measures of other obligations faced by the students (work hours, travel, and other courses they were taking) and undergraduate and current graduate grade point average (GPA). On average, the students in the sample reported working 38.9 hours per week while traveling 2.3 times per semester for work.

METHODS

Subjects, Assignments, and the Experimental Design

A long-running, online MBA course in operations management was selected for the study. This course is typically taken near the end of a student's program. Students were contacted by a professor (not the course instructor) before the course began requesting participation. Student volunteers were informed that the research entailed a semester-long project about teaching and learning effectiveness and that they would be required to complete several surveys throughout the semester. They were not specifically informed that the experiment involved collaborative learning and were not compensated for their participation although they did receive results about their learning styles and personality traits. One hundred and ninety two students were solicited and 72 students completed all parts of the experiment for a completion rate of 37.5%. All surveys were administered online. Students completed their assessments of each assignment before grades and feedback were provided. The course instructor was never informed about who was participating and who was not.

In the first week of the course, before any of the assignments were given, participating students completed several online surveys that provided data for the input variables. The repeated measures design that generated the dependent variable unfolded as follows. Three assignments were chosen for the experiment;

Table 1: Experimental design.

Semester	Group	Independent	Sample Size
One	Assignment A	Assignment B	21
Two	Assignment B	Assignment C	8
Three	Assignment C	Assignment A	30
Four	Assignment B	Assignment C	13

Note: Design for Semester Two was repeated in Semester Four due to small sample size.
Assignment A occurred in week seven of the 16-week semester.
Assignment B occurred in week 11 of the 16-week semester.
Assignment C occurred in week 14 of the 16-week semester.

due dates were in weeks 7, 11, and 14 of the 16-week semester. These weeks were chosen to allow students to gain individual expertise prior to being asked to work in groups while also mitigating the potential for end-of-course effects such as those due to finals and workload issues. The first assignment was a profit maximization blending problem. The context was a process planning decision. The second assignment was a cost minimization network problem; it required students to address supply chain management issues in their analysis. The third assignment introduced binary variables to control for logical decisions (either/or) which had to be linearized. The context for the analysis was a location and deployment decision.

Over the course of four semesters, whether an assignment was completed in a group or independently was rotated using the experimental design found in Table 1, thereby mitigating the effects of increased expertise gained throughout a term. Each assignment either required group collaboration or was required to be completed independently without any assistance from other students, colleagues, or the instructor. Each student had both an assignment that required group work and one that was to be completed individually. The three assignments were not strictly identical over the course of the four terms. They were edited so that different numerical answers were obtained and the underlying situation appeared different, yet the modeling and analysis skills required between terms were identical. Following completion of each experimental assignment, students completed an online survey in which they were asked to assess perceived learning.

Group membership was assigned randomly by the instructor. However, the instructor did permit private e-mail requests in the event of problematic groupings; that is, students could confidentially ask that they not be grouped with a particular student. No requests for this consideration were received. Group size was limited to three or four members since smaller groups tend to be more conducive to improved outcomes (Lou et al., 2001).

Statistical Methods

Partial Least Squares Regression (PLS Regression) was used to analyze the data set. The PLS Regression process, derived from Wold's work (Carrascal, Galvan, & Gordo, 2009; Ferrer, Aguado, Vidal-Puig, Prats, & Zarzo, 2008; Wold, Sjöström & Eriksson, 2001;), is conducive to exploratory projects such as this one. While the PLS estimation process has been used for path modeling via structural equation techniques, PLS Regression is different in that no underlying theoretical model is introduced or tested. The estimation process finds linear combinations of the predictor variables; interpretation of individual variables is impossible and no attempt is made to describe the underlying factors. No assumptions are made about the distributions of residuals. The technique is designed to mitigate multicollinearity and it has been shown to perform well in samples with small numbers of observations (Carrascal et al., 2009; Garthwaite, 1994). Although widely used in chemical engineering, only a few examples of PLS Regression models have been seen in social science research (e.g., Laitinen, 2008; Su & Zheng, 2008; Lee, Lee, Cho, Im, & Kim, 2011; Tenenhaus, Pagès, Ambroisine, & Guinot, 2005).

The data were analyzed with the JMP 11 PRO Partial Least Squares Platform using the NIPALS algorithm. K-Fold was used to mitigate overfitting. The stepwise process outlined in Anderson and Bro (2010) was used for variable reduction. Variables with the lowest Variance Importance Projection (VIP) were removed one at a time. At each removal step, the Root Mean PRESS, percentage variation explained, and factor plots were examined to determine variable relevance and to assess the number of factors in the model (Anderson & Bro, 2010; Anzanello, Albin, & Chaovalitwongse, 2012; Denham, 2000; Forina, Casolino, & Pizarro Millan, 1999). Several variables with values of VIP less than 1.0 remained in the model as they provided explanatory power that was lost when they were removed (Anderson & Bro, 2010; Chong & Jun, 2005; Mehmood, Liland, Snipen, & Sæbø, 2012). After the final model was derived, the classification / misclassification ratio and plots were analyzed as a final check of predictive power.

RESULTS

The final PLS regression model was a significant four-factor model that explained 53.18% of the variability in perceived learning differences. It included the variables shown in Table 2. Variables from all categories were in the final model: demographic, personality traits, learning styles, technology proficiencies, and other. The number of hours worked appears in the model. This controls for the opportunity cost effects of working adult students.

The predictive ability of the final model is depicted in Figure 2. When the model predicted that a student's perceived learning favors collaboration yet actual perceived learning favored independent work, a misclassification has occurred; these reside in quadrant II (upper left) of Figure 2. When a student reported learning more with group work but the model predicted perceived learning would be greater with independent assignments, another type of misclassification has occurred (quadrant IV, lower right). Consequently, nine of 72 students were misclassified and 87.5% of observations were classified correctly. The misclassification rate of 12.5% compares favorably with those of other studies to have used these methods in social science settings; e.g., 11.6%–26.5% reported by Laitinen (2008) and 55% in Lee et al. (2011). The correlation between actual and predicted values in Figure 2 is .854. Equally notable is the small magnitude of the misclassified observations. Although differences between actual and model-predicted perceived learning run from about ± 5 standard units, all misclassified observations were between ± 2 with five of them less than $|1|$.

Table 2: Variables included in the PLS regression.

Variable	VIP
Visual/Verbal	1.2648
Neuroticism	1.2470
Computer proficiency	1.1996
Business school GPA	1.0925
Chat room proficiency	1.0049
Total daily computer hours	.9775
Daily computer hours for school work	.9652
Threaded discussion proficiency	.9438
Streaming media proficiency	.9263
Age	.9221
Community experience	.8799
Sensing/Intuitive	.8739
Weekly work hours	.8591
Marital status	.8578
Conscientiousness	.8395

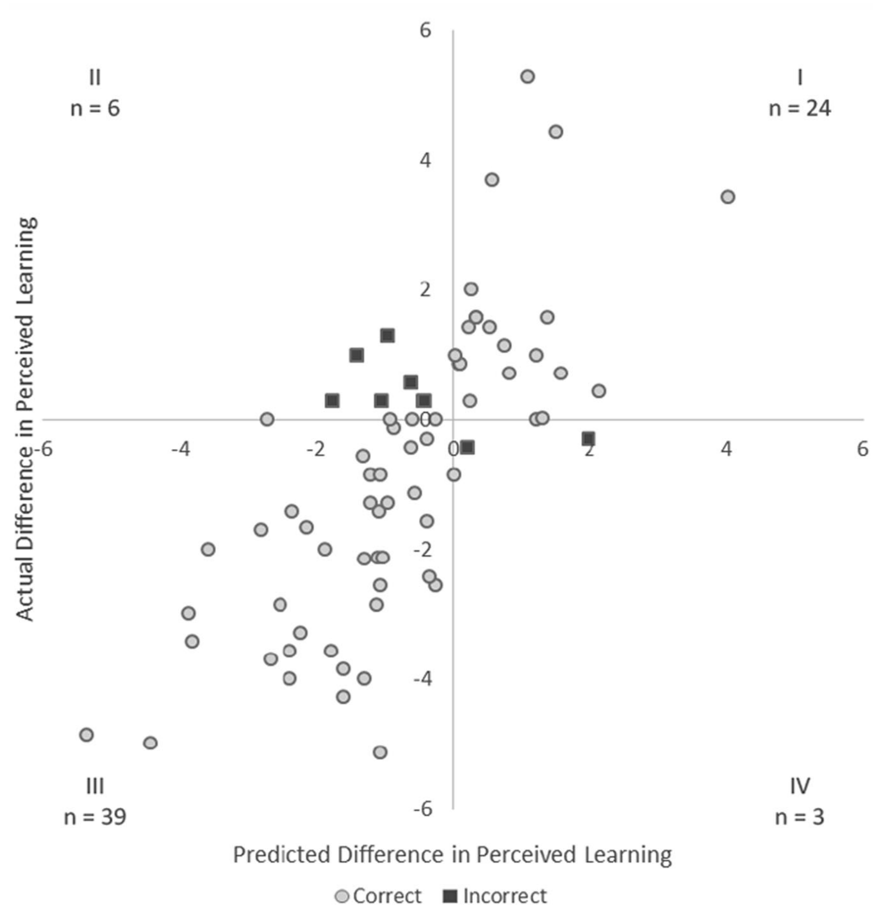
DISCUSSION

Using a predictive model like the one developed here, an instructor could advise or assign work using the prediction. Students with positive predictions would be advised to join groups and those with negative predictions would be advised to work independently. Those near zero could use their best judgment. Based on the model, it is expected that if students followed the advice, about 12.5% of them would be guided toward the wrong learning environment, but none of them would be expected to be harmed seriously.

Without a predictive model, many students' learning might be compromised, some substantially. Consider, for example, the instructor who buys into the argument that group work is essential to good online learning and thereby requires collaborative work of all students. According to Figure 2, the model predicts that 62.5% (the 45 independent-learning students in quadrants II and III) of students will have been forced into the least beneficial learning modality for them. Alternatively, those instructors who, for whatever reason, prohibit group work are predicted to disenfranchise 37.5% of their students.

Instructors are the final arbiters of what is best for their students. Some instructors may require group work (or not) regardless of the costs and benefits to students. For example, courses that teach the skills and methods of online communication or the management of individuals and teams are natural outlets for required collaborative work. Other instructors may prohibit group work based on equally valid arguments such as a call from future employers to measure individual performance, or the headaches of managing free rider problems. Among those who wish to learn how their choices impact individual students, predictive modeling might provide the appropriate insights.

Figure 2: Prediction graph for PLS regression model.



LIMITATIONS AND FUTURE DIRECTIONS

This study speaks to perceived learning. By necessity, more objective measures of learning, such as grades, could not be used because the instructor could not assess the performance of an individual student when the output was based on a group report. This situation is analogous to the "levels mismatch" problem described by Benbunan-Fich (2010) and "levels incongruence" in Gallivan and Benbunan-Fich (2005). One solution might have been to use peer reviews to apportion the group grade to individuals. This approach was rejected because the group assignment grade would then have encompassed both an objective (instructor) assessment and the subjective assessment of peers; it would not have been clear what was being measured once the difference between the group-based grade and the grade on the individual assignment was calculated. Instead, the cleaner measure of each student's self-assessment of both the individual and group assignments was used.

Recent research however suggests that perceived learning is a viable measure for the purpose of this research. Benbunan-Fich (2010, p. 325) suggested that self-reported learning more closely measures how much one has learned, while objective measures such as grades reflect how much one knows, stating ". . . learning perception scales typically intend to assess the *process* whereby the amount of currently held knowledge has been changed." The present study intended to do just that, to measure whether students report more learning via individual or collaborative work.

The meta-analysis of Sitzmann, Ely, Brown, and Bauer (2010) suggested that self-assessments reflect affective, more than cognitive, components of learning. To date, this and other research into self-reported learning have used methodologies that speak to the aggregate rather than the individual. This article presents a different perspective. A predictive modeler might argue that self-assessments more closely measure cognitive learning for some students but affective learning for others, or may be some mix of both. Although it is beyond the scope of this work to delve more closely into the meaning of self-assessed learning at the individual student level, the authors concur with Benbunan-Fich (2010), Sitzmann et al. (2010), and others that more research is needed before we can fully understand the constructs behind self-reported learning.

Earlier research regarding the efficacy of collaborative work in online education might be classified into three broad groups. One body of work would suggest that practitioners include collaboration in their online courses because collaboration produces various cognitive, affective, or other benefits for students (e.g., An et al., 2008; Bernard et al. 2000; Cameron et al. 2009; Ding et al., 2011; Resta & Laferriere, 2007; Swan, 2001). A contrarian view would warn that practitioners who use collaboration might not see the advertised benefits because there is no statistical link between collaboration and various learning outcomes (e.g., Arbaugh, 2005; Kellogg & Smith, 2009; Lou et al., 2001; Lukman & Krajnc, 2012). A third perspective argues that instructors should weigh the benefits against the costs of collaboration (e.g., An & Kim, 2009; Bernard et al., 2000; Cornelius et al., 2009; Kellogg & Smith, 2009; Lukman & Krajnc, 2012). This article provides a bridge between these three perspectives. A predictive modeler would argue that the benefits of collaboration accrue to some students and not others, and that a prediction model might identify which is which while also quantifying the costs to students of a suboptimal (for them) learning environment. These ideas are illustrated via a sample of working adult learners in an online MBA program. This model and its results are not expected to generalize to different populations, for example, to K-12 or undergraduate students, nonworking adults, or those learning face-to-face. Indeed, among the challenges and opportunities of predictive modeling is identifying inputs relevant to the population of interest.

Predictive models are not a substitute for explanatory models and theory-building; each has its own place in the evolution of knowledge. Explanatory models are not usually evaluated for their predictive potential, while predictive models do not provide explanatory insights. For educators, a predictive modeling approach seems to be a viable addition in determining how to develop differentiated instruction.

REFERENCES

Alavi, M. (1994). Computer-mediated collaborative learning: An empirical evaluation. *MIS Quarterly*, *18*(2), 159–174.

An, H., & Kim, S. (2009). The benefits and limitations of online group work in a teacher education program. *Journal of Technology and Teacher Education*, *17*(2), 2465–2472.

An, H., Kim, S., & Kim, B. (2008). Teacher perspectives on online collaborative learning: Factors perceived as facilitating and impeding successful online group work. *Contemporary Issues in Technology and Teacher Education*, *8*(1), 65–83.

Anderson, C. M., & Bro, R. (2010). Variable selection in regression—A Tutorial. *Journal of Chemometrics*, *24*, 728–737.

Anstine, J., & Skidmore, M. (2005). A small sample study of traditional and online courses with sample selection adjustment. *The Journal of Economic Education*, *36*(2), 107–127.

Anzanello, M. J., Albin, S. L., & Chaovalitwongse, W. A. (2012). Multicriteria variable selection for classification of production batches. *European Journal of Operations Research*, *218*, 97–105.

Arbaugh, J. B. (2000). Virtual classroom characteristics and student satisfaction with internet-based MBA courses. *Journal of Management Education*, *24*(1), 32–54.

Arbaugh, J. B. (2005). Is there an optimal design for on-line MBA courses? *Academy of Management Learning & Education*, *4*(2), 135–149.

Arnold, K. (2010) Signals: Applying academic analytics. *Educause Quarterly*, *33*(1).

Battalio, J. (2009). Success in distance education: Do learning styles and multiple formats matter? *The American Journal of Distance Education*, *23*(2), 71–87.

Benbunan-Fich, R. (2010). Is self-reported learning a proxy metric for learning? Perspectives from the information systems literature. *Academy of Management Learning & Education*, *9*(2), 321–328.

Bernard, R. M., Rojo de Rubalcava, B., & St. Pierre, D. (2000). Collaborative online distance learning: Issues for future practice and research. *Distance Education*, *21*(2), 260–277.

Bradley, J. H., & Hebert, F. J. (1997). The effect of personality type on team performance. *Journal of Management Development*, *16*(5), 337–353.

Cameron, B. A., Morgan, K., Williams, K. C., & Kostelecky, K. L. (2009). Group projects: Student perceptions of the relationship between social tasks and a sense of community in online group work. *The American Journal of Distance Education*, *23*(1), 20–33.

Campbell, J. P., DeBlois, P. B., & Oblinger, D. G. (2007). Academic analytics: A new tool for a new era. *Educause Review*, July/August, 41–57.

Campbell, N., & Williams, J. E. (1990). Relation of computer attitudes and computer attributions to enrollment in high school computer courses and self-perceived computer proficiency. *Journal of Research on Computing in Education*, *22*(3), 276–289.

Carrascal, L. M., Galvan, I., & Gordo, O. (2009). Partial least squares regression as an alternative to current regression methods used in ecology. *Oikos*, *118*, 681–690.

Cassidy, S., & Eachus, P. (2002). Developing the computer user self-efficacy (CUSE) scale: Investigating the relationship between computer self-efficacy, gender and experience with computers. *Journal of Educational Computing Research*, *26*(2), 133–153.

Chitkushev, L., Vodenska, I., & Zlateva, T. (2014). Digital learning impact factors: Student satisfaction and performance in online courses. *International Journal of Information & Education Technology*, *4*(4), 356–359.

Chong, I. G., & Jun, C. H. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems*, *78*(1), 103–112.

Chu, R. J. C. (2010). How family support and Internet self-efficacy influence the effects of e-learning among higher aged adults—Analyses of gender and age differences. *Computers & Education*, *55*(1), 255–264.

Cornelius, S., Gordon, C., & Ackland, A. (2009). Towards flexible learning for adult learners in professional contexts: An activity-focused course design. *Interactive Learning Environments*, *19*(4), 381–393.

Correa, T., Hinsley, A. W., & Zuniga, H. (2010). Who interacts on the Web?: The intersection of users' personality and social media use. *Computers in Human Behavior*, *26*(2), 247–253.

Denham, M. C. (2000). Choosing the number of factors in partial least squares regression: Estimating and minimizing the mean squared error of prediction. *Journal of Chemometrics*, *14*, 351–361.

Ding, R., Chen, J., Knudson, R., & Braun, F. (2011). *What matters? An investigation of student collaborative e-learning*. Paper presented at the 2011 International Conference on E-Business and E-Government (ICEE), Shanghai, China.

Doherty, W. A., & Maddux, C. D. (2002). An investigation of methods of instruction and student learning styles in internet-based community college courses. *Computers in the Schools*, *19*(3–4), 23–32.

Duff, A., Boyle, E., Dunleavy, K., & Ferguson, J. (2004). The relationship between personality, approach to learning and academic performance. *Personality and Individual Differences*, *36*(8), 1907–1920.

Dziuban, C., Moskal, P., Cavanagh, T., & Watts, A. (2012). Analytics that inform the university: Using data you already have. *Journal of Asynchronous Learning Networks*, *16*(3), 21–38.

Eom, S. B., Wen, H. J., & Ashill, N. (2006). The determinants of students' perceived learning outcomes and satisfaction in university online education: An empirical investigation. *Decision Sciences Journal of Innovative Education*, *4*(2), 215–235.

Felder, R. M. (1988). Learning and teaching styles in engineering education. *Engineering Education*, *78*(8), 674–681.

Felder, R. M., & Spurlin, J. (2005). Applications, reliability, and validity of the Index of Learning Styles. *International Journal of Engineering Education*, *21*(1), 103–112.

Ferrer, A., Aguado, D., Vidal-Puig, S., Prats, J. M., & Zarzo, M. (2008). PLS: A versatile tool for industrial process improvement and optimization. *Applied Stochastic Models in Business and Industry*, *24*(6), 551–567.

Forina, M., Casolino, C., & Pizarro Millan, C. (1999). Iterative predictor weighting (IPW) PLS: A technique for the elimination of useless predictors in regression problems. *Journal of Chemometrics*, *13*, 165–184.

Gallivan, M. J., & Benbunan-Fich, R. (2005). A framework for analyzing levels of analysis: Issues in studies of e-collaboration. *IEEE Transactions on Professional Communication*, *48*(1), 87–104.

Garthwaite, P. H. (1994). An interpretation of partial least squares. *Journal of the American Statistical Association*, *89*(425), 122–127.

Gow, A. J., Whiteman, M. C., Pattie, A., & Deary, I. J. (2005). Goldberg's "IPIP" Big-Five factor markers: Internal consistency and concurrent validation in Scotland. *Personality and Individual Differences*, *39*(2), 317–329.

Hawk, T. F., & Shah, A. J. (2007). Using learning style instruments to enhance student learning. *Decision Sciences Journal of Innovative Education*, *5*(1), 1–19.

Ice, P., Diaz, S., Swan, K., Burgess, M., Sharkey, M., Sherrill, J., Huston, D., & Okimoto, H. (2012). The PAR framework proof of concept: Initial findings from a multi-institutional analysis of federated postsecondary data. *Journal of Asynchronous Learning Networks*, *16*(3), 63–86.

Jiang, M., & Ting, E. (2000). A study of factors influencing student's perceived learning in a web-based course environment. *International Journal of Educational Telecommunications*, *6*(4), 317–338.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of Personality: Theory and Research*, *2*(1999), 102–138.

Ke, F., & Xie, K. (2009). Toward deep learning for adult students in online courses. *The Internet and Higher Education*, *12*(3), 136–145.

Kellogg, D. L., & Smith, M. A. (2009). Student-to-student interaction revisited: A case study of working adult business students in online courses. *Decision Sciences Journal of Innovative Education*, *7*(2), 433–456.

Koroghlanian, C. M., & Brinkerhoff, J. (2007). Online students' technology skills and attitudes toward online instruction. *Journal of Educational Technology Systems*, *36*(2), 219–244.

Kuo, Y. C., Walker, A. E., Belland, B. R., & Schroder, K. E. (2013). A predictive study of student satisfaction in online education programs. *The International Review of Research in Open and Distance Learning*, *14*(1), 16–39.

Laitinen, E. K. (2008). Data system for assessing probability of failure in SME reorganization. *Industrial Management & Data Systems*, *108*(7), 849–866.

Lee, H., Lee, Y., Cho, H., Im, K., & Kim, Y. S. (2011). Mining churning behaviors and developing retention strategies based on a partial least squares (PLS) model. *Decision Support Systems*, *52*(1), 207–216.

Lim, B. C. & Ployhart, R. E. (2006). Assessing the convergent and discriminant validity of Goldberg's International Personality Item Pool. *Organizational Research Methods*, *9*(1), 29–54.

Lim, D. H., & Kim, H. (2003). Motivation and learner characteristics affecting online learning and learning application. *Journal of Educational Technology Systems*, *31*(4), 423–439.

Litzinger, T. A., Lee, S. H., & Felder, R. M. (2007). A psychometric study of the Index of Learning Styles. *Journal of Engineering Education*, *96*(4), 309–319.

Lou, Y., Abrami, P. C., & D'Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis, *Review of Educational Research*, *71*(3) 449–521.

Lukman, R., & Krajnc, M. (2012). Exploring non-traditional learning methods in virtual and real-world environments. *Educational Technology & Society*, *15*(1), 237–247.

Marks, R. B., Sibley, S. D., & Arbaugh, J. B. (2005). A structural equation model for effective online learning. *Journal of Management Education*, *29*(4), 531–563.

Martin, F. (2008). Blackboard as the learning management system of a computer literacy course. *Journal of Online Learning and Teaching*, *4*(2), 138–145.

Mehmood, T., Liland, K. H., Snipen, L., & Sæbø, S. (2012). A review of variable selection methods in partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, *118*, 62–69.

O'Connor, M. C., & Paunonen, S. V. (2007). Big Five personality predictors of post-secondary academic performance. *Personality and Individual Differences*, *43*(5), 971–990.

Papp, R. (2001). Student learning styles and distance learning, *Proceedings of the 16th Annual Conference of the International Academy for Information Management*, http://eric.ed.gov/?id = ED474077 (accessed December 20, 2014).

Park, J. H. (2007). Factors related to learner dropout in online learning. *Online Submission*. http://eric.ed.gov/?id = ED504556 (accessed December 16, 2014).

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles concepts and evidence. *Psychological Science in the Public Interest*, *9*(3), 105–119.

Resta, P., & Laferriere, T. (2007). Technology in support of collaborative learning, *Educational Psychology Review*, *19*(1), 65–83.

Rovai, A. P., & Baker, J. D. (2005). Gender differences in online learning: Sense of community, perceived learning, and interpersonal interactions. *Quarterly Review of Distance Education*, *6*(1), 31–44.

Sabry, K., & Baldwin, L. (2003). Web-based learning interaction and learning styles. *British Journal of Educational Technology*, *34*(4), 443–454.

Schnienderjans, M. J., & Kim, E. B. (2005). Relationship of student undergraduate achievement and personality characteristics in a total web-based environment: An empirical study. *Decision Sciences Journal of Innovative Education*, *3*(2), 205–221.

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, *25*(3), 289–310.

Sitzmann, T., Ely, K. E., Brown, K. G., & Bauer, K. N. (2010). Self-assessment of knowledge: A cognitive learning or affective measure? *Academy of Management Learning & Education*, *9*(2), 169–191.

Socha, A., Cooper, C. A., & McCord, D. M. (2010). Confirmatory factor analysis of the M5-50: An implementation of the International Personality Item Pool Set. *Psychological Assessment*, *22*(1), 43–49.

Su, H., & Zheng, G. (2008). A partial least squares regression-based fusion model for predicting the trend in drowsiness. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, *38*(5), 1085–1092.

Swan, K. (2001). Virtual interaction: Design factors affecting student satisfaction and perceived learning in asynchronous online courses. *Distance Education*, *22*(2), 306–331.

Tenenhaus, M., Pagès, J., Ambroisine, L., & Guinot, C. (2005). PLS methodology to study relationships between hedonic judgments and product characteristics. *Food Quality and Preference*, *16*(4), 315–325.

Thomas, P., Moore, K. S., & Scott, K. S. (1996). The relationship between self-efficacy for participating in self-managed work groups and the Big Five personality dimensions. *Journal of Organizational Behavior*, *17*(4), 349–362.

vanBraak, J. P. (2004). Domains and determinants of university students' self-perceived computer competence. *Computers & Education*, *43*(3), 299–312.

Volery, T., & Lord, D. (2000). Critical success factors in online education. *International Journal of Educational Management*, *14*(5), 216–223.

Whittingham, K. L. (2006). Impact of personality on academic performance of MBA students: Qualitative versus quantitative courses. *Decision Sciences Journal of Innovative Education*, *4*(2), 175–190.

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, *58*(2), 109–130.

Zywno, M. S. (2003). A contribution to validation of score meaning for Felder-Soloman's index of learning styles. *Proceedings of the 2003 American Society for Engineering Education Annual Conference & Exposition*, Vol. *119*, 1–5.

APPENDIX: VARIABLES AND SUMMARY STATISTICS

Perceived Learning		
	Mean	St. Dev.
Perceived learning—Group Assignment	4.50	1.73
Perceived learning—Independent Assignment	5.25	1.58
Perceived learning—Difference	−.75	2.21

Demographic Predictors (number in sample)		
Age	≤30	>30
	43	29
Gender	Male	Female
	31	41
Employed	Yes	No
	66	6

Income	≤$25K	>$25 K to ≤50K	>$50K to ≤$75	>$75K
	3	15	11	43
Marital Status	Married	Single	Divorced	Committed Relationship
	43	12	2	15

	Mean	St. Dev.
How many children or dependents do you have?	.6	.99

Learning Styles (ILS) Predictors		
	Mean	St. Dev.
Sensory/Intuitive	−1.9	5.36
Visual/Verbal	2.5	5.59
Active/Reflective	−.7	4.34
Sequential/Global	−.4	5.21

Personality (IPIP) Predictors			
	Mean	St. Dev.	Alpha
Openness	3.6	1.00	.836
Conscientiousness	3.9	0.59	.760
Extraversion	3.4	0.73	.833
Agreeableness	3.7	0.60	.796
Neuroticism	2.3	0.73	.788

Continued

Computer / Internet Proficiency Predictors (number in sample)				
How would you rate your proficiency with:	Never Used	Beginner	Moderate	Advanced
Computers	N/A	0	26	46
The Internet	N/A	0	18	54
Using chat rooms	9	16	27	20
Using threaded discussions	2	3	35	32
Using streaming video	8	17	31	16
Using online communities	15	18	28	11

	Mean	St. Dev.
For how many years have you been using a computer	16.0	4.64
For how many years have you been using the Internet	10.4	2.50
How many hours per week do you use the Internet	24.0	18.62
In a typical day, how many hours per day do you spend using a computer for:		
Workplace/business tasks	6.9	2.33
Graduate school course work	3.0	5.03
Home/personal tasks	1.2	.91
Sum of previous three for total computer hours per day	11.1	4.60
How many other online courses have you taken	3.5	2.75

Other Predictors		
	Mean	St. Dev.
Considering only your employment hours and in a typical week, how many hours per week do you spend doing your job	38.9	15.12
This semester, how many times will be you out of town for work-related travel	2.3	3.19
Not including this course, how many other courses are you taking	1.6	.85
What was your undergraduate GPA	3.33	.36
What is your current Business School GPA	3.56	.30

Dr. Marlene Smith is on the faculty at the University of Colorado Denver's Business Analytics program. She teaches predictive analytics, forecasting, and introductory statistics courses and does research in statistical modeling.

Dr. Deborah Kellogg is an Associate Professor at the University of Colorado Denver in the Business Analytics program. She teaches operations management and VBA for modelers and has won several teaching awards including Online Innovative Teacher of the Year. Her research interests are in service operations and online education.