Evidence Based Library and Information Practice

Article

Using Analytic Tools with California School Library Survey Data

Dr. Lesley Farmer
Professor of Librarianship
California State University
Long Beach, California, United States of America
Email: Lesley.Farmer@csulb.edu

Dr. Alan Safer
California State University
Long Beach, California, United States of America
Email: Alan.Safer@csulb.edu

Joanna Leack
California State University
Long Beach, California, United States of America
Email: Joannaleack@gmail.com

Received: 2 Dec. 2014 **Accepted**: 4 Feb. 2015

Abstract

Objective — California school libraries have new state standards, which can serve to guide their programs. Based on pre-standard and post-standard library survey data, this research compares California school library programs to determine the variables that can potentially help a school library reach the state standards, and to develop a predictive model of those variables.

Methods – Variations of decision trees and logistic regression statistical techniques were applied to the library survey data in order to create the best-fit model.

Results – Best models were chosen within each technique, and then compared, concluding that the decision tree using the CART algorithm had the most accurate results. Numerous variables

came up as important across different models, including: funding sources, collection size, and access to online subscriptions.

Conclusion – School library metrics can help both librarians and the educational community analyze school library programs closely and determine effective ways to maximize the school library's impact on student learning. More generally, library resources and services can be measured as data points, and then modeling statistics can be applied in order to optimize library operations.

Introduction

From preschools to college, every school's mission is to provide their students with the very best education possible. To do this, schools have to provide many things, such as curriculum, instruction, resources, and an effective learning environment.

Within this framework, school libraries have as their mission: "to ensure that students and staff are effective users of ideas and information" (AASL, 2007, p. 8). This mission involves both physical and intellectual access, and requires considering preconditions, such as providing as much material in as many different formats as possible, or being open during commonly accessible hours for students. Having proper staff, enough funding, and a quality collection potentially positively impact the school community, and therefore can have a positive effect on student learning outcomes.

In order to keep libraries striving to provide the best program of resources and services possible, some states set standards for those conditions that benefit student learning. In 2011 the California Department of Education put into effect statewide Model School Library Standards, which address many aspects of a library including resources, staff, services, and budget. The standards included both student performance standards and research-based standards for the school library programs themselves (Farmer & Safer, 2010).

This notion of standards transcends school libraries. Academic, public, and special libraries also need to provide the resources and services to meet their communities' information needs. Library systems may have baseline standards, stating a minimum number of volumes, subscriptions, equipment, staff, and required services. At the least, libraries often compare their resources and services to those of their counterparts, so that normative measures emerge. By identifying key factors that impact the library's operational effectiveness, and by developing a predictive model, libraries can optimize funding decisions and develop evidence-based standards and guidelines.

This research used several data analytic techniques to determine which aspects of a California school library program affect its ability to meet these statewide standards. These statistical methods can be applied to many other library settings.

Literature Review

In this age of added accountability and value-added impact, school librarians need to show how they contribute to the school's mission. Furthermore, in tough economic times, school librarians have needed to make their case in order to continue their programs.

Hundreds of research studies have found significant positive correlations between aspects of school library programs and student achievement. Mansfield University's literature review (Kachel, 2013), Scholastic's *What Works!* (2008), Farmer's 2003 synthesis and the Library Research Service website (http://www.lrs.org/data-tools/school-

libraries/impact-studies/) provide compendia of school library impact studies.

Traditionally, school library programs have based their worth on their input and processes, that is, their resources and services (Hatry, 2006). Those program elements, however, have to be used in order to have an impact on student success, so usage figures are also kept. In the final analysis, student work, test scores, grades, retention and graduation rates serve as more useful data points of impact, although the library's contribution is generally harder to measure (except for analysis of research projects). Often librarians resort to perception-based assessment methods such as anecdotal observations, surveys, and interviews or focus groups, which may be more subjective compared to data-driven analyses (Loertscher, 2008; Mardis, 2011).

These data help determine the baseline quantity and quality of resources and services needed in order to provide satisfactory library programs: in other words, standards. In 2009 the American Association of School Librarians (AASL) developed guidelines for school library programs, based on their 2007 standards for 21st century learners, membership surveys, and focus groups. Kentucky and Missouri have formally adopted these national standards, and California's 2011 standards (see Appendix A) were informed by the AASL's work. Three-quarters of states have state-based school library standards, which tend to focus on staffing and resource quantitative measures and do not reflect the AASL's 2009 guidelines (Council of State School Library Consultants, 2014). Only Montana's state standards appeared to be research-based (Bartow, 2009). Only Texas investigated the relationship between their state standards and student achievement as measured by state standardized achievement tests (Smith, 2001), which informed their 2005 revision (Texas State Library and Archives Commission, 2005); their conclusions were based on Pearson Correlation statistics.

As the California Model School Library Standards were being developed, the state Library Consultant saw the need to underpin the standards with research. To that end, she asked Dr Farmer to review the literature about school library standards and program factors that significantly impact student success. Updating her 2003 literature review, and drawing upon other existing compendia, as noted above, Dr Farmer identified contributing variables that appeared consistently in the literature:

- staffing (*full-time credentialed school librarian*, *full-time paraprofessional*)
- access (*flexible access to the library throughout the day for groups and individuals*)
- services (*instruction, collaboration, reading guidance and promotion*, reference, interlibrary loan)
- resources (large current diverse and relevant materials that are well organized)
- technology (*Internet connectivity, online databases, online library catalogue, library web portal*)
- management variables (budget, *administrative support, documented policies, and procedures, strategic plan with assessment*).

The presence of the specific variables (shown in italics) became the basis for the California school library baseline standards. The variables that are quantitative in nature (e.g., budget size, currency of collection) were calculated to determine adequate levels of support, which also constituted part of the baseline library standards (California Department of Education, 2011).

As the California school library program standards were being approved, Farmer and Safer (2010) wanted to determine if a significant difference existed between those school libraries that met the standards and those who did not. Using the state's most recent school library data

set (2007-2008), the researchers applied descriptive statistics to identify standards variables. To be so designated, at least half of the survey respondents had to meet that specific baseline standard variable (that is, the library didn't have to meet all of the factors' standards). Next, the researchers divided the data set into two groups: one that met all baseline standards, and one that did not meet all baseline standards. A t-test determined that the two groups were significantly different relative to resource and service standards; the most significant difference relative to the baseline standards was the presence of a full-time school librarian. A logistic regression analysis found that several variables related to resources and services further differentiated the two groups: number of subscription databases, library web portal presence, information literacy instruction, Internet instruction, flexible scheduling, planning with teachers, book and non-book budget size, and currency of collection.

Objectives

When the California Model School Library Standards were being developed in 2010, the state economy was in crisis, and as a result school librarian positions were being eliminated. The 2007-2008 data set analyzed by Farmer and Safer (2010) preceded this economic drop, which provided a good baseline. The same researchers used the following year's data (2011-2012) for comparison, and developed four research questions:

1. Has the number of school library programs meeting the standard changed since the standards were approved?
2. How do the significant variables identified in the 2007-2008 data set compare with those in the 2011-2012 data set?
3. Which variables differentiate school library programs meeting state standards from those which do not meet the standards?
4. What statistical model provides the best fit of school library programs meeting state standards?

Methods

To answer the research questions, the current study used the 2011-2012 California school library survey data set, and referred back as needed to the 2010 Farmer and Safer study.

Each year the California Department of Education requests all K-12 schools in California to complete the annual online public school library survey about the prior year's library data. Typically, the site library staff complete and submit the survey, although occasionally a school administrator responds to the survey. The researchers had access to the resulting data, and applied several statistical methods to determine a model that would describe the data in terms of meeting state school library standards.

Data Description

The California Department of Education received 4017 responses (out of a possible 8588 K12 public schools, excluding special education, continuation, and alternative schools) to its survey regarding data about site school library for the academic year 2011-2012. Of the respondent schools, 387 (9.6%) did not have a library so they were removed from the analysis, leaving 3630 useable libraries (observations). Since information about the survey was disseminated to county and district superintendents, and to the state's school librarian listserv, it is reasonable to assume that non-respondents were less likely to have school libraries than respondents. Thus, the resulting data set may be considered representative of school libraries in California.

Most response variables were coded as binary: indicating whether or not the school's library met the specific state standard, with 0 being no and 1 being yes. Three independent variables

were categorical: single or joint use of library, credentialed staff, and grade level. Five other independent variables had continuous values: number of books, average copyright date, book budget, and non-book budget. There were a total of 64 variables, noted in Appendix B. The value of each variable was calculated; frequency and percentage statistics were applied to compare school libraries that met state standards and those that did not.

The researchers used SAS Enterprise Miner to identify which school libraries met all the standards, and that determination was coded (0 being no and 1 being yes) into the data set.

Statistical Predictive Modeling

The researchers wanted to develop a statistical model to predict which school libraries could meet state standards, based on a set of variables. The underlying concept of predictive analysis entails "searching for meaningful relationships among variables and representing those relationships in models" (Miller, 2013, p. 2). Predictive analytics reveals explanatory variables or predictors those factors that can relate to a desired response or outcome. In other words, what is the probability of an outcome given a set of input data? In this study, variable values were compared for library programs that met standards versus programs that did not. Was there a subset of variables that could predict a school library's success at meeting the state standards?

Two main statistical techniques are recognized for developing decision procedures: logistic regression and decision trees. Logistic regression is used to predict a response based on input data. Decision trees are used to predict a categorical response, such as meeting standards or not (Miller, 2013). Within these two types of techniques are several possible versions.

A decision tree diagram looks like a flow chart because it is essentially a sequenced set of if-then decisions based on questions. An example is computer troubleshooting a printer failure, starting with the question: Does the computer have electrical power? Depending on the answer (yes or no), different actions are taken (e.g., If no, is the computer plugged in? If yes, is the computer switched on? This branching continues through a series of decision points). Decision trees are a very useful statistical tool because they make good visual aids that are easy to interpret, and help show the relative importance of variables. They also facilitate predictions if a strong tree is built, and help find profiles of variables that are either much more likely, or much less likely, to occur than the overall average. Statistical programs such as SAS and SPSS can generate tree-based classification models using algorithms.

Two main types of decision tree techniques are CART (Classification and Regression Trees) and C4.5. The CART method allows for just two splits at any node (for example, meeting the standard or not, or budget greater or less than a specific amount), which can work well in this study because of the large number of binary factors. The algorithm is set up to choose a split among all possible splits at each node; it depends on the value of just one predictor variable. The best split point is one in which the resulting variables are unlikely to be mixed in ensuring splits. Using the example above, determining if the power is on or not is a good first node because the ensuing issues are likely to be dependent on that first choice. A competing algorithm to CART is C4.5, where a node may split into more than two branches (Larose, 2005).

The other statistical technique employed was logistic regression, which is used to find a model that relates a dependent binary variable (here, meeting, or not meeting, the state standard) with a set of independent variables. There are several advantages to running regression. First, if one is able to build a strong regression, it can potentially be used as a predictive tool for future data. It is relatively easy to interpret the effect of changing one predictive variable on the

Table 1
Quantitative Factors of School Library Programs Meeting Standards

	2011-2012 Data	**2007-2008 Data**
Average Number of Books	~21,000	~16,000
Average Copyright Date	1994	1995
Average Book Budget	~$8000	~$5000
Average Non-book Budget	~$4000	~$4000

Table 2
Data Set of School Library Programs Meeting Baseline Standards

Level of School	**Total N (2011-2012)**	**N Meeting Standard**	**% Meeting Standard 2011-2012 Data**	**Total N (2007-2008)**	**% Meeting Standard 2007-2008 Data**
Elementary	2303	12	0.5	3312	0.4
Middle School	531	39	7.3	842	8.2
High School	500	161	32.2	688	44.9
TOTAL	3334	212	6.4	4832	7.4

response variable, holding the other predictive variables constant. Regression analysis can also yield valuable information about the data within it. The researchers applied three selection methods — backwards, forward, and stepwise — and compared results to determine which model best fitted the data (Miller, 2013).

To compare the fit of each model, and see error rates across models, ROC (Receiver Operating Characteristic) was used (Larose, 2005). This kind of chart visualizes the effectiveness of a classification model, calculating how well a variable will be assigned the right category, in comparison to being assigned a category randomly. A good predictive model shows a steep incline and remains near the top of the graph, indicating that the model can distinguish between two (or more) groups easily. A model that has a line close to the diagonal would imply that the classification is close to random or guessing. Thus, a good model indicates that the categories are well chosen, and can be used to predict high-quality school library programs. As with decision trees, statistical software programs can calculate the ROC based on the model's sensitivity of the classification schemes.

For each model, the data were partitioned into a training set, for model fitting, and a validation set, for empirical validation. This technique was used in order to generalize to better predict future values. The training set included a random sample of 70% of the observations, and the remaining 30% composed the validation set to see how well the models classified other sets, and to determine possible generalizations (Larose, 2005).

Results

Each library's data were compared with the California state standards (Appendices A and B). Only 212 school library programs met all of the standards. Appendix B details those resource and service variables that were present and independently met the standards. Table 1 contains quantitative values for only the libraries that met all the standards; it lists the mean for those variables having continuous values (rather than simply being available or not, such as access after school).

Table 2 compares the percentage of school libraries that met all of the state standards in

2011-2012 with the percentage of school libraries in 2007-2008 that met the standards before those standards were officially approved and disseminated (Farmer & Safer, 2010).

Decision Trees

Each of the following trees randomly used 70% of the observations as the training set and the remaining 30% for the validation (or test) set.

CART (Classification and Regression Trees)

Figure 1 shows the tree generating the smallest misclassification error (for example, a variable that is classified as meeting a standard when in actuality it does not, or vice versa). Unfortunately, the data set has many variables so the full decision tree is very difficult to read when viewing it as a whole. The right side of the tree is bushier than the left. The root node for this tree is "funding from the state lottery" since about 90% of the libraries did not receive state lottery funding (value of 0). Therefore, the tree is unbalanced, with most decision points appearing on the right branch for that root node.

This decision tree's variable importance is shown in Table 3.

Table 3
Variable Importance in CART Decision Tree

VARIABLE NAME	IMPORTANCE
Automated catalogue	1.0000
State lottery funds	0.8274
Access to online resources	0.8228
Online subscriptions	0.6155
Streaming video subscriptions	0.5047
Budget	0.4978
Librarian helps find resources outside the library	0.4174

C4.5

Running the C4.5 algorithm with target criterion set to entropy (that is, the least probability of a variable result occurring), the tree with the lowest misclassification error (that is, put in the wrong category) was generated. Compared to the tree gained from the CART method, this tree has several more branches and nodes (i.e., it is bushier), which could potentially lead to a higher misclassification error. Upon examination, the researchers found the misclassification rate to be the same as the tree generated with the CART algorithm.

Regression

The next statistical technique used was logistic regression. Three selection methods were applied: backwards, forward, and stepwise. Afterwards, a comparison determined which model best fitted the data.

Main Effects: Backward Selection

The first logistic regression model to be run used only the basic variables, called the main effects model. Using backwards selection initially starts with all the variables and slowly removes the insignificant ones. Fifty-two steps (iterations) occurred during the backwards selection process. The final model selected is shown in Table 4. Many resources emerged, such as an automated online catalogue and automated textbook circulation. Even more so, the kind of funding a library receives also appears on the list frequently.

Fit statistics showed a misclassification error of 15.4%, which is relatively high in comparison to the other analysis performed. The average squared error also gives a percentage of 12.5%, which is high. Both of these rates corresponded to the validation set.

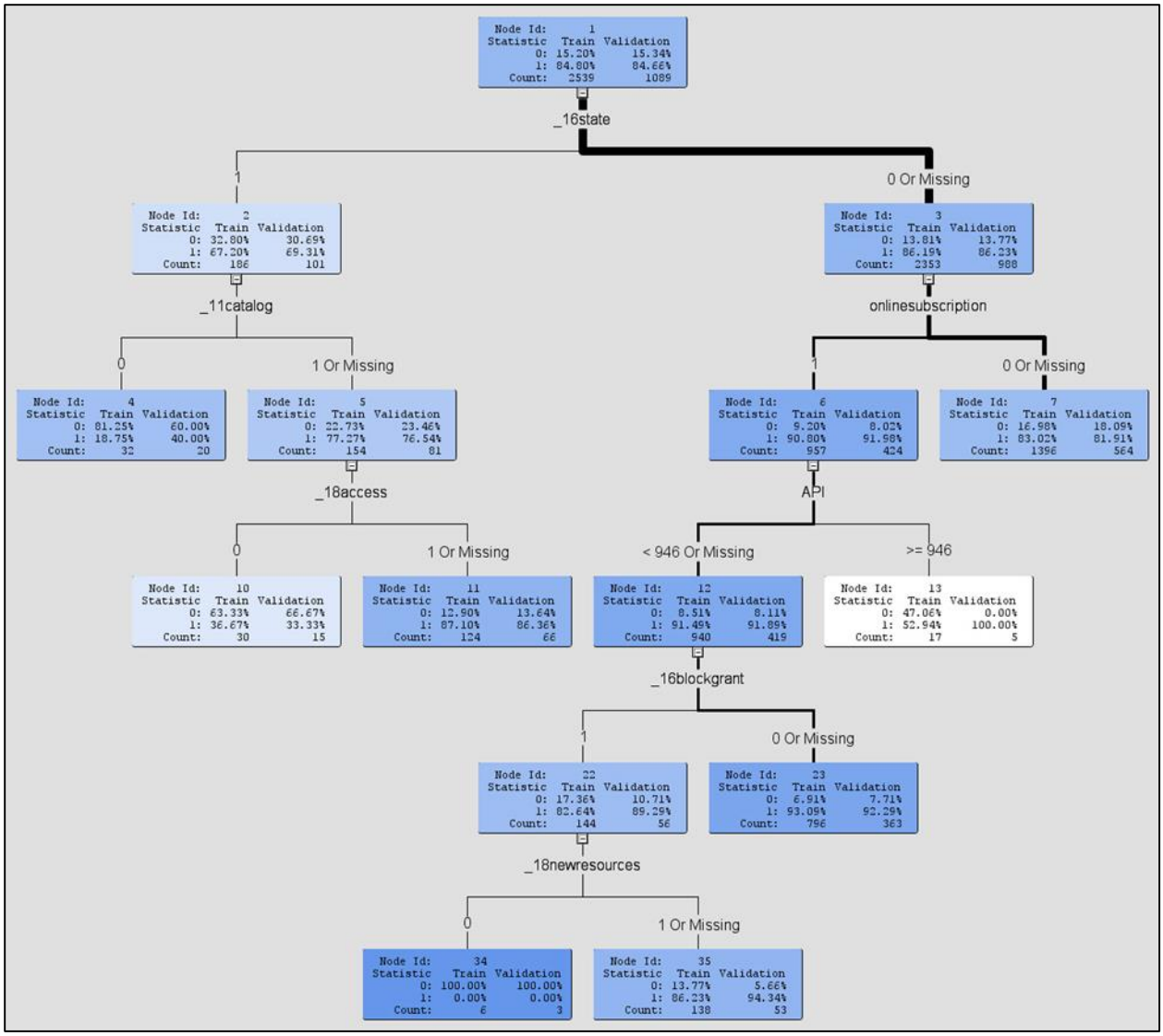

Figure 1
Full decision tree.

Main Effects: Forward Selection

Forward selection was used next on the main effects design. Forward selection starts with zero variables and adds significant variables until the model is complete. The Estimated Selection Plot is a visual way of seeing which variables were selected at which step in the process. State lottery funding was the first variable selected for the model, not surprisingly. Using forward selection, only 15 steps were needed to create the optimal model.

Several variables associated with funding were used, much like the main effects model using backwards selection. This model had a 15.2% misclassification rate on the validation set, which is a slight improvement in comparison to the backwards selection model. However, a classification chart showed that the model incorrectly classified school libraries that did not meet school standards.

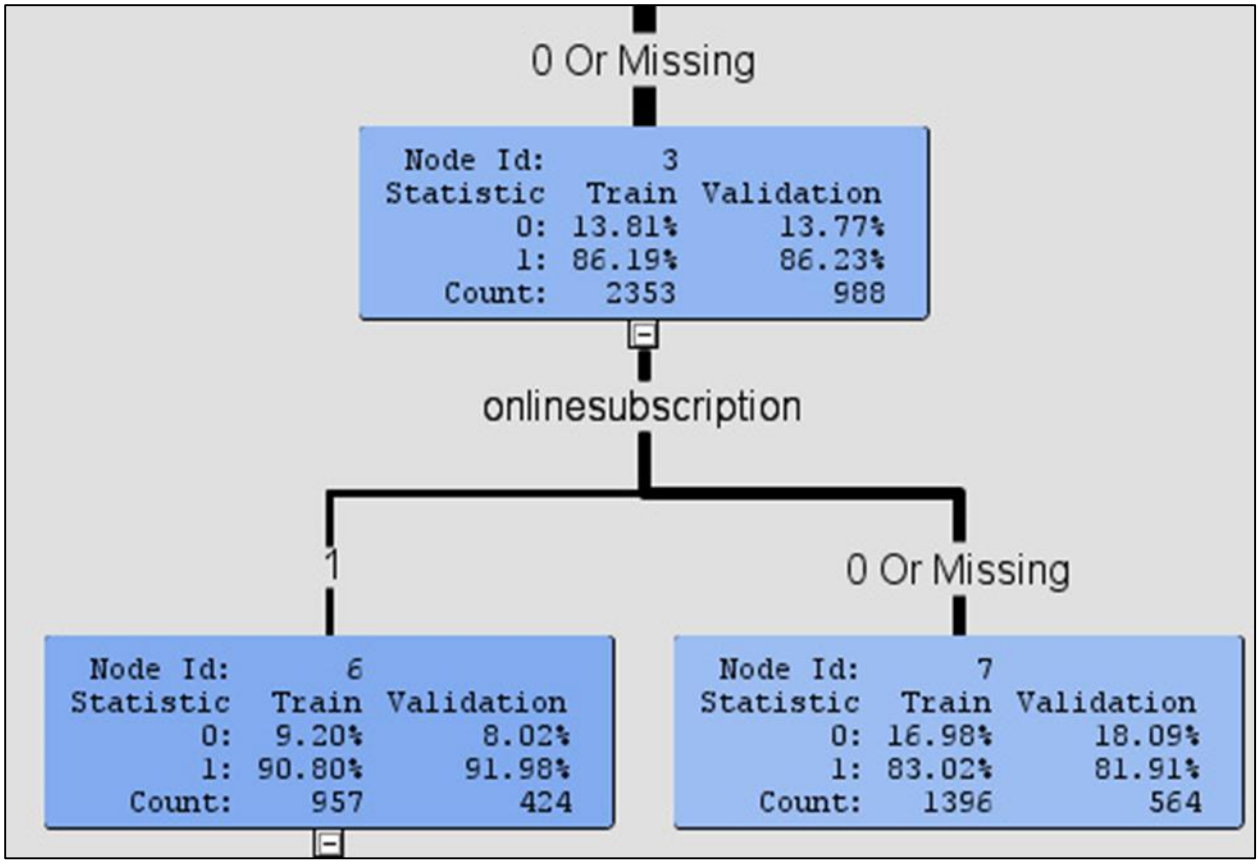

Figure 2
Decision tree detail.

Table 4
Final Model – Regression, Main Effects, Backwards Selection

VARIABLE	SIGNIFICANCE LEVEL
Book budget	0.013
Automated catalogue	< 0.01
Integrated information literacy instruction	<0.01
State Block grants (from federal government)	<0.01
State school library funding	<0.01
Librarian helps find resources outside the library	0.013
Interlibrary loan	0.047
Librarian does online publishing	0.026
Librarian creates wikis	<0.01
Online subscriptions	<0.01

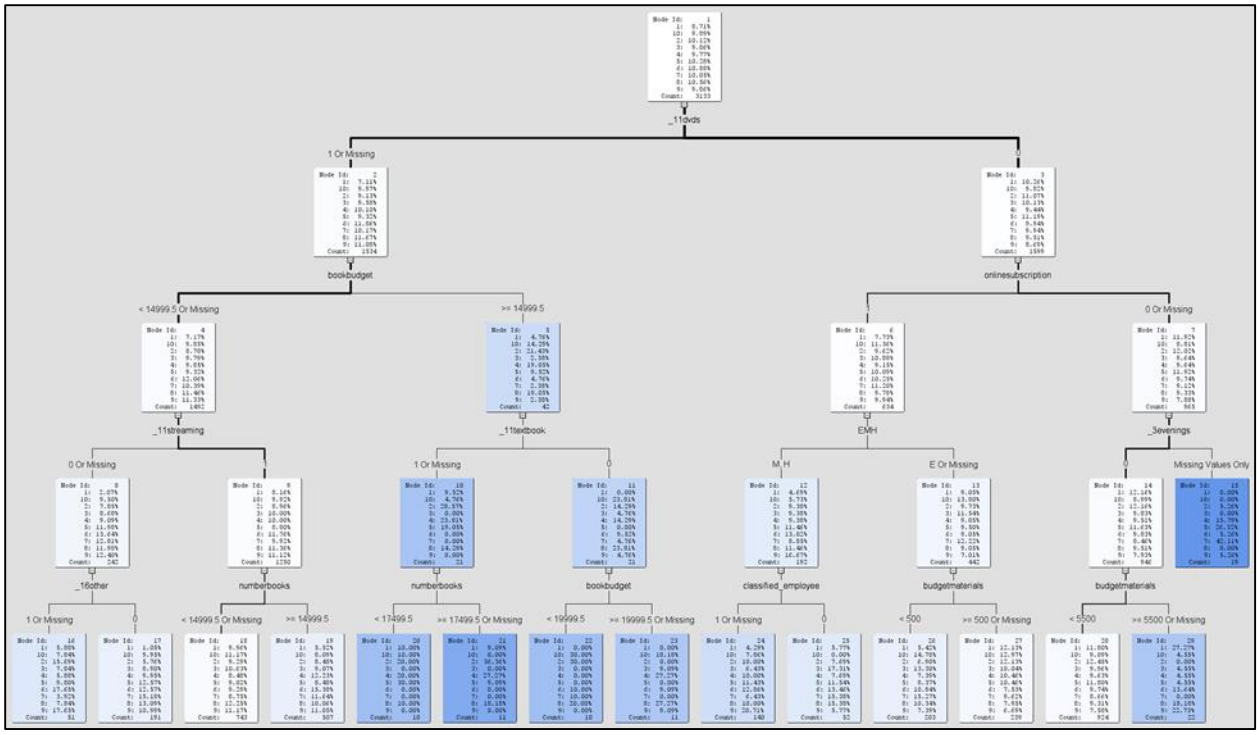

Figure 3
CART decision tree – excluding funding sources

Main Effects: Stepwise Selection

In this selection model a variable can be added or removed at each step, depending on which would make the model better. The process resulted in the same variables as the forward selection one.

Discussion

With the introduction of the California Model School Library Standards, the educational community has metrics by which to assess school library programs and specific targets to aim for in improving those programs. Furthermore, since these metrics were based on the professional literature about significant factors that impact student learning, the standards provide a case for value-added school library programs – and areas that could optimize such value.

School Libraries Meeting Standards

The first research question asked whether the number of school library programs meeting the standard changed since the standards were approved. The short answer is "no" for elementary and high schools, and "yes" for middle schools, at least in terms of percentages. A major confounding external factor was economics, somewhat exacerbated by politics. By fall 2011, the state and federal economy was precarious, and federal funding for school libraries was severely reduced. Not surprisingly, elementary school librarian positions became scarcer. At the high school level, school librarians became more likely to split their time between two (or more) schools, so they no longer met the standard of a full-time librarian at the site. In that respect, it is actually a bit heartening to see that the percentage of full-time middle school librarians increased 2.2%, although it still left almost 90% of middle schools without a full-time school librarian. It

will be interesting to see in future years the extent to which school library programs improve because of the standards – or to which they improve because of the economic outlook. The latter picture would then predict that money more than standards makes the difference, which could lead to an unstable program.

For those school libraries that met the baseline standards, some interesting comparisons emerged. The average number of books increased as did the book budget, but the average copyright date was one year older than for the 2007-2008 data set. The 2011-2012 data set included some libraries build since 2008, which would account for the budget increase (and core collections include classic titles so are not automatically newer). In addition, school libraries may be reluctant to weed their collections in fear of leaving subject gaps, resulting in larger but older collections. Non-book budgets stagnated.

Predictive Variables and Models

The next research question asked what factors (i.e., variables) of a school library program can help determine if any given school library will meet the state standards in California? The accompanying fourth research question asked what statistical model provides the best fit of school library programs meeting state standards. Some variables that stand out are those pertaining to staff, budget, and student accessible resources. These variables make sense since budget often drives resources, and staff manage the school library program.

Each decision tree generated similar sets of variables, even though the trees were formed using different algorithms. These variables included state lottery funding, online access, and average copyright dates. The optimal decision trees formed had an average misclassification rate of 14.4%. These additional variables speak to more advanced school library program efforts, going beyond baseline measures. For instance, not only does the number of materials matter, but their currency impacts their use – and reflects the school's support of the collection.

In examining the CART decision tree, the root node of "funding source" seemed to skew the remaining branches and leaves. Further investigation with the former state school library consultant revealed that lottery and state grant funds were inactive at the time, but it was possible to use carry-over money to help finance school library programs. The survey responder, who was usually library staff, either had to know about this "inside" money stream or naively check off that box; the data seemed to indicate the former scenario. In that respect, the 10% of librarians who indicated this funding source are likely to be "in the know" about budgets or have good communication with the fiscal agents; in either case, this knowledge reflects pro-active management. Such a disposition could be generalizable to other factors of the library programs, such as the availability of resources and related services.

To sidestep this issue of funding sources, a second CART decision tree was generated that excluded the funding sources variables. The result was a more balanced tree, as viewed in Figure 3.

The important variables that emerged included (in order of importance): budget for non-book materials, evening access, book budget, number of books, level of library, availability of DVDs, having classified staff, having online subscriptions (including streaming), and providing textbook service.

Decision Tree Model Comparison.

ROC charts (Figure 4) visualized differences between the CART and C4.5 tree for the training and validation data sets. The highest line signifies the C4.5 tree, the next highest line signifies the interactive tree (i.e., manually built), and the third line signifies the CART tree. For

the training set, the C4.5 algorithm shows better results (higher accuracy), but the validation set shows better results with the decision trees produced by CART. Wanting a tree with good predictive power, the accuracy of the validation set is more important; the higher the line (that is, maximum area above the line), the more accurate the model. With the ROC percentages so close and the misclassification rates the same, the tree with the smallest averaged squared error should be selected as the optimal tree (Larose, 2005). The decision tree which was produced using the CART algorithm showed the best results for the validation set, and was chosen to be the optimal tree.

Regression Model Comparison

Several logistic regression models were also run, including ones such as main effects and polynomial degrees. Final models show significant variables to include the amount budgeted for books in 2011 and state lottery funding. Although the logistic regression models were formed using multiple selection techniques, their misclassification rates did not match up to other models gained using different techniques. To see which logistic regression model was the best at classifying, ROC charts were analyzed (Figure 4). When considering the training set, the model that had the greatest accuracy and was best at classifying existing data was the main effects design using backwards selection. The same regression model also showed the highest accuracy when it came to the validation set based on the ROC chart. Fit statistics had the highest ROC index for the backwards selected model, but that model also had the highest misclassification rate of the regression models. The difference, however, was only .2%, and when drilling down to specific variables, the backwards selection model did not have outstanding single misclassifications as did the forward selection. Therefore, the backwards selection regression model was chosen as the preferred regression model.

Final Model Comparison

Model comparisons were run to determine the best models under each statistical technique. The

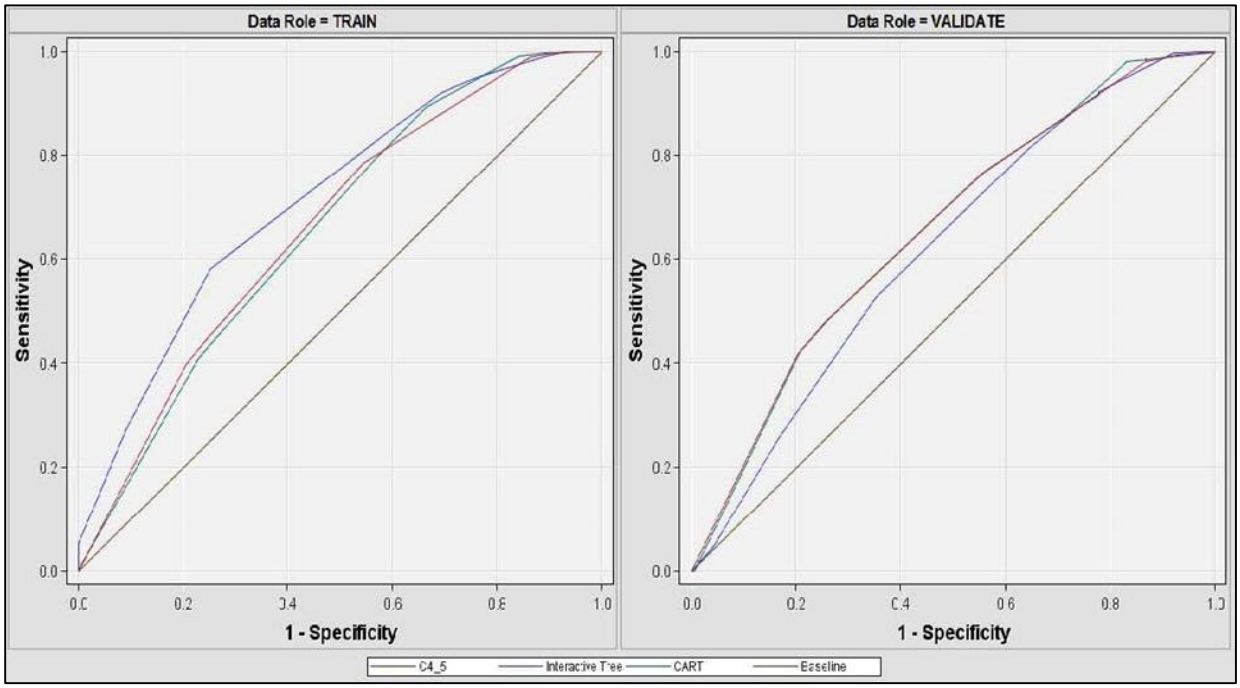

Figure 4
ROC Chart – Comparison of Decision Tree Models

goal at this point was to choose the overall best model, regardless of the method. The CART decision tree received the lowest misclassification rate, but it also has the second lowest training set accuracy (ROC index).The logistic regression model had the highest misclassification rate and average squared error percentage; it also consistently had the lowest accuracy for both the training and validation set. Figure 5 shows the ROC Chart for the validation set. The blue squares indicate the line that represents the CART decision tree model. Even though a few of the models had higher accuracies than this method, CART gave a model with the lowest misclassification rate and average squared error.

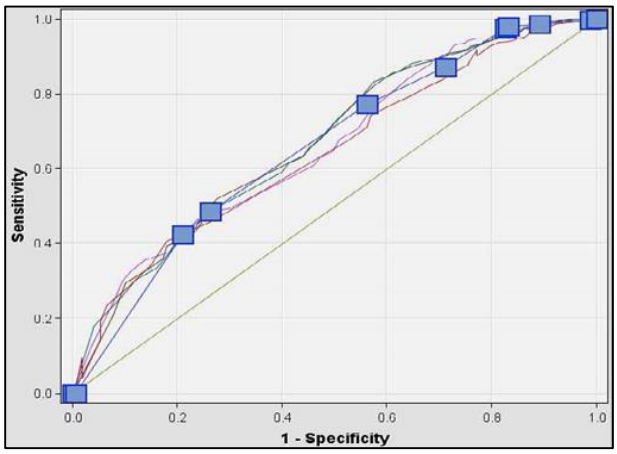

Figure 5
ROC chart – final model comparison

Comparison of 2007-2008 and 2011-2012 Variables

The second research question asked how the significant variables identified in the 2007-2008 data set compared with the 2011-2012 data set. In the 2007-2008 study, the distinguishing variables were: availability of subscription databases, Internet instruction, flexible scheduling, library web portal existence, information literacy instruction, planning with teachers, book and non-book budget size, and currency of collection. Using the CART decision tree, several variables remained the same: availability of subscription databases, book and non-book budget size. Additional variables identified in the 2011-2012 data included evening access and availability of DVDs (probably not included in the earlier data set because of the small N sample size), having classified employees (probably because they were scarcer in 2011-2012), number of books, and textbook service (largely a function of high schools, and may be influenced by changing staffing patterns and online textbook initiatives). Instruction and planning tend not to correlate closely with budget or even resources. So they might occur even in poorer school libraries; what is probably a more significant factor since 2007-2008, however, is the increased importance (and consequences) of high-stakes testing, which has tended to reduce library instruction and co-planning time.

Conclusions

This research study examined California school library programs in light of the state's Model School Library Standards. Using the California school library survey data from 2007-2008, Farmer and Safer's 2010 study helped form these standards, and they discovered a significant difference between school libraries that met state standards and those who did not. The current research used the 2011-2012 school library survey data, which asked the same questions. This study compared the standards data, and explored a number of statistical models to find a best fit for capturing data about school library programs that could be used as predictors of program quality in order to provide the conditions for optimum learning experiences and student academic success.

The current study could not uncover any visible impact of the approved state standards on the 2011-2012 data relative to the 2007-2008 data (research questions 1 and 2), but the time frame was very short to expect any such changes. More current data would be needed, substantiated by interviews with school librarians to explain

possible reasons for changes. Furthermore, the economic and political landscape changed in the interim between the two time frames, which could account for changes.

After conducting several types of analyses, the CART decision tree provided the best fit to explain the data (research question 4). Funding overall, and use of a variety of funding sources, were major factors in school library program status (research question 3). The findings pointed out the need for librarians to be aware of these funding streams, and to take advantage of them, which may require pro-active communication and negotiation with decision-makers. Resources and access to them constituted another important "leg" of school library programs. Books, non-print and online resources are all needed, and some analysis seemed to indicate that both physical and intellectual access through instruction were needed in order to make a difference. In general, there seemed to be a sizable gap between the vast majority of school libraries providing basic resources and services and those stellar libraries with rich collections, innovative services, and expanded access. In that respect, there is a possible Matthew effect (that is, the bad become worse) that shows up more clearly in bad economic times.

As a model and possible predictive tool, the CART decision tree has potential as a way to examine school library programs, and determine the most effective allocation of funding in order to have a high-quality school library. This statistical model can also be used to make funding decisions in other kinds of libraries as well. The same variables could be used when appropriate, but other likely variables could be used as well, such as free parking, self-checkout systems, story hours, thesis workshops, and so on.

Much research remains to be done. California has parallel data sets from 2003 to 2013, which can be analyzed using the CART decision tree model to look for patterns over time, both in terms of meeting standards as well as comparing important variables that make a significant difference in school library programs. Newer survey data can be analyzed to see if the Model School Library Standards impact support of school libraries. The CART decision tree can also be used with data from other states, or compared with national data, to determine possible significant differences between populations – or if a different model should be used.

School library metrics can help both librarians and the educational community analyze school library programs closely and determine effective ways to maximize the school library's impact on student learning. More generally, library resources and services can be measured as data points, and then modeling statistics can be applied in order to optimize library operations.

References

American Association of School Librarians. (2007). *Standards for the 21st century learner*. Chicago, IL: American Library Association.

American Association of School Librarians. (2009). *Empowering learners: Guidelines for school library media programs*. Chicago, IL: American Library Association.

Bartow, C. (2009). How one state established school library/technology standards. *School Library Monthly*, *26*(3), 19-21.

California Department of Education. (2011). *Model school library standards for California public schools kindergarten through grade twelve*. Sacramento, CA: California Department of Education.

Council of State School Library Consultants. (2014). *Standards.* Salem, OR: Council of State School Library Consultants. Retrived from http://cosslc.wikispaces.com/standards

Farmer, L. (2003). *Student success and library media programs.* Westport, CT: Libraries Unlimited.

Farmer, L., & Safer, A. (2010). Developing California school library media program standards. *School Library Media Research* 13. Retrieved from http://www.ala.org/aasl/slr

Hatry, H. (2006). *Performance measurement: Getting results* (2nd ed.). Washington, DC: The Urban Institute.

Kachel, D. (2013). *School library research summarized: A graduate class project*. Mansfield, PA: Mansfield University.

Larose, D. (2005). *Discovering knowledge in data: An introduction to data mining*. Hoboken, NJ: Wiley-Interscience.

Loertscher, D. (2008). Information literacy 20 years later. *Teacher Librarian, 35*(5), 42-43.

Mardis, M. (2011). Evidence or evidence based practice? An analysis of IASL research forum papers, 1998-2009. *Evidence Based Library and Information Practice, 6*(1), 4-23. Retrieved from http://ejournals.library.ualberta.ca/index.php/EBLIP/index

Miller, T. (2013). *Modeling techniques in predictive analytics: Business problems and solutions with R.* Upper Saddle River, NJ: Pearson.

Scholastic Publishing. (2008). *What works!* New York, NY: Scholastic.

Smith, E. G. (2001). *Texas school libraries: Standards, resources, services and students' performance*. Austin, TX: Texas State Library and Archives Commission.

Texas State Library and Archives Commission. (2005). *School library programs: Standards and guidelines for Texas*. Austin, TX: Texas State Library and Archives Commission.

Appendix A
School Library Program Standards (California Department of Education 2011, 34-42)

Full time teacher librarian (.5 for schools with enrollment between 350 and 785 students)
Full time paraprofessional librarian assistant
Library open to students at least 36 hours per week
Integrated library management system with online public access capability
Library web page
Internet access for students
Flexible scheduling at least 20 hours per week
Class set of networked computers (10 for elementary, 15 for middle school, 25 for high school)
Facility to accommodate one class for instruction and small group independent work
Collaborative planning and teaching for at least two grade levels or departments
At least 20 hours of instruction per week
At least 5 hours of management per week
Reading guidance
Current policies, procedures and library plan, including assessment
At least two online subscription databases
Print magazines (25 for elementary, 20 for middle school, 15 for high school)
At least two-thirds of the collection less than 15 years old
At least 28 books per student

One book per student added per year for elementary and middle school; one book per two students for high school

Appendix B
2011-2012 School Library Program Variables Meeting State Standards

Total N=3628 (8 missing) Elementary N= 2591 (6 missing) N= 533 HS N= 498 (2 missing)

VARIABLE	ELEMEN-TARY	MS	HS	TOTAL # MEETING STANDARD	TOTAL % MEETING STANDARD
Credentialed Full-time Teacher Librarian	299 (11.6%)	188 (35.2%)	330 (66.3%)	818	22.6
Paraprofessional	2590 (99.7%)	447 (83.9%)	373 (74.6%)	2927	80.7
Open before school	1069 (41.2%)	421 (79%)	428 (85.6%)	1917	53
Open for classes	2479 (95.5%)	513 (6.2%)	482 (96.4%)	3620	95.9
Open during breaks	1721 (66.3%)	440 (82.6%)	437 (87.4%)	2597	71.7
Open during lunch	1483 (57.1%)	486 (91.2%)	447 (89.4%)	2414	66.5
Open after school	1085 41.8%)	391 (73.4%)	420 (84%)	1895	52.3
Open evenings	55 (2.1%)	16 (3%)	61 (12.2%)	132	3.6
Open weekends	7 (0.3%)	1 (0.2%)	19 (3.8%)	27	0.7
Open summers	51 (2%)	6 (1.1%)	50 (10%)		
Used instructional materials funds	81 (3.1%)	39 (7.3%)	44 (8.8%)	164	4.5
Used state lottery funds	182 (7%)	49 (9.2%)	57 (11.4%)	287	7.9
Used per pupil allotment funds	255 (9.8%)	57 (10.7%)	70 (14%)	383	10.6
Used general funds	569 (21.9%)	190 (5.6%)	252 (50.4%)	1009	27.8
Received block grant	296 (11.4%)	78 (14.6%)	68 (13.6%)	441	12.2
Did fundraising	1398 (53.8%)	298 (55.9%)	128 (25.6%)	1825	50.3
Used Title I funding	23 (7.8%)	72 (13.5%)	56 (11.2%)	364	10.0
Used Title V funding	6 (0.2%)	5 (0.9%)	3 (0.6%)	14	0.4
Use local bond funding	66 (2.5%)	13 (2.4%)	11 (2.2%)	91	2.5
Received other grant funding	1 (0.2%)	1 (0.2%)	2 (0.4%)	7	0.2
Received start-up funds	16 (0.6%)	1 (0.2%)	2 (0.4%)	19	0.5
Received other funding	323 (2.4%)	75 (14.1%)	92 (18.4%)	490	13.5
Did online publishing	165 (6.4%)	74 (13.9%)	135 (27%)	374	10.3

Share photos online	96 (3.7%)	51 (9.6%)	89 (17.8%)	236	6.5
Used a news feed	83 (3.2%)	46 (8.6%)	65 (13%)	195	5.4
Generated digital images	61 (2.3%)	38 (7.1%)	61 (12.2%)	160	4.4
Used social bookmarks	44 (1.7%)	28 (5.3%)	61 (12.2%)	133	3.7
Used wikis	249 (9.6%)	113 (21.2%)	126 (25.2%)	488	13.5
Used online productivity tools	451 (17.4%)	209 (39.2%)	298 (59.6%)	957	26.5
Used online social libraries	95 (3.7%)	42 (7.9%)	68 (13.6%)	205	5.7
Used online videos	402 (15.5%)	154 (28.9%)	218 (43.6%)	774	21.3
Downloaded audio files	182 (7%)	72 (13.5%)	95 (19%)	349	9.6
Used ebooks and audiobooks	421 (16.2%)	141 (26.5%)	188 (37.6%)	751	20.7
Used learning management systems	107 (4.1%)	51 (9.6%)	86 (17.2%)	244	6.7
Provided OPAC	2157 (83.1%)	488 (91.6%)	453 (90.6%)	3097	85.4
Circulate textbooks	779 (30%)	318 (59.7%)	308 (61.6%)	1403	38.7
Provided access to online resources	565 (21.8%)	174 (32.6%)	277 (55.4%)	1017	228
Provided video streaming services	1294 (49.8%)	322 (0.4%)	295 (59%)	1909	52.6
Provided DVDs	1136 (43.7%)	304 (57%)	301 (60.2%)	1739	47.9
Provided audiobooks	655 (25.2%)	232 (43.5%)	248 (49.6%)	1134	31.3
Provided integrated searching portal	618 (23.8%)	168 (31.5%)	186 (37.2%)	972	26.8
Conducted workshops	208 (8%)	104 (19.5%)	168 (33.6%)	481	13.3
Offered integrated information literacy instruction	202 (7.8%)	95 (17.8%)	154 (30.8%)	452	12.5
Informally instructed on resource use	1503 (57.9%)	375 (70.4%)	384 (76.8%)	2261	62.3
Gave reference help	1884 (72.5%)	463 (86.3%)	442 (88.4%)	2789	76.9
Helped find outside resources	947 (36.5%)	297 (55.7%)	358 (71.6%)	1603	44.2
Facilitated interlibrary loan	854 (32.9%)	245 (46%)	186 (37.2%)	1285	35.4
Helped parents realize lifelong learning importance	981 (37.8%)	160 (30%)	141 (28.2%)	1283	35.4

Coordinated in-school production of materials	203 (7.8%)	68 (12.8%)	87 (17.4%)	359	9.9
Collaborated to create AV products	118 (4.5%)	67 (12.6%)	110 (22%)	296	8.2
Did AV programming	90 (3.5%)	51 (9.6%)	64 (12.8%)	206	5.7
Coordinated computer networks	370 (14.2%)	153 (28.7%)	159 (31.8%)	683	18.8
Provided access to OPAC	1717 (66.1%)	438 (82.2%)	411 (82.2%)	2564	70.7
Provided student Internet access	1594 (61.4%)	470 (88.2%)	437 (87.4%)	2499	68.9
Provided access to resource sharing network	258 (9.9%)	121 (22.7%)	172 (34.4%)	551	15.2
Communicated proactively with principal	1817 (70%)	415 (77.9%)	362 (72.4%)	2593	71.5
Attended site council 2x (year or more	499 (19.2%)	161 (30.2%)	182 (36.4%)	842	23.2
Provided online subscription DB	956 (36.8%)	262 (49.2%)	330 (6%)	1547	42.6