Enabling teachers to explore grade patterns to identify individual needs and promote fairer student assessment

Sorelle A. Friedler *, Yee Lin Tan, Nir J. Peer, and Ben Shneiderman

Department of Computer Science, University of Maryland, A.V. Williams Building, College Park MD 20742, USA

Abstract

Exploring student test, homework, and other assessment scores is a challenge for most teachers, especially when attempting to identify cross-assessment weaknesses and produce final course grades. During the course, teachers need to identify subject weaknesses in order to help students who are struggling with a particular topic. This identification often needs to happen across multiple assessment data points and should be considered in comparison to the class's progress as a whole. When determining grades, fairness to all is essential, but there are special needs for students who did poorly on one exam or had a steadily increasing grasp of the subject. We present `eduViz`, a visualization tool designed to help teachers explore and assign grades. Teachers can see the trajectory of student scores, the relationship of a particular student to the class, and use categories they have defined in order to filter their assessment information. Query response is immediate and all logical comparisons are possible. Teachers can easily compare their query to the class or per student average as well as view scores by raw point total or percentage. Additionally, `eduViz` provides a grade assignment interface which allows teachers to view sorted student scores in a scatterplot. This scatterplot is coupled with a unique partition slider which allows users to move color coordinated bands on the scatterplot to indicate grade ranges. As these grade ranges are set, a histogram is updated to show the number of students assigned to each grade range. These features give teachers new and powerful ways to explore and assign grades so that they can better understand student strengths and weaknesses and make the most of the time they have available. Interviews with sixteen expert teachers indicate that `eduViz` is a success across fields, provides teachers with a useful tool to understand and help their classes, and encourages reflective practice.

Key words: Human-computer interface, Improving classroom teaching, Evaluation methodologies, Pedagogical issues, Teaching/learning strategies

---

[1] For help videos and code access see http://www.cs.umd.edu/∼sorelle/eduviz/.

* Corresponding author.

Email addresses: `sorelle@cs.umd.edu` (Sorelle A. Friedler), `yeelin@cs.umd.edu` (Yee Lin Tan), `peer@cs.umd.edu` (Nir J. Peer), `ben@cs.umd.edu` (Ben Shneiderman).

1 Introduction

Teachers spend between a third and a half of their time preparing, administering, or grading assessments [Stiggins, 1999]. The remainder of their time is spent in front of the classroom, preparing classes, talking to parents, helping students individually, performing administrative duties, and doing many other tasks necessary during the school day (lunch duty, hall monitoring, etc). School districts, administrators, parents, and the teachers themselves hope and expect that after all these tasks are completed, the teachers will still have time to review assessment results to identify student and class strengths and weaknesses so that strong students can be challenged, weak students can be helped, and extra credit can be added to compensate for unusually difficult tests. It has long been understood that teacher reflection on student assessment can help students learn [Garfield, 1994]. In reality, however, many teachers lack the needed training to effectively assess their students [Doolittle, 2002]. Potential reasons for this general lack of understanding include a fear of "incomprehensible statistics and esoteric concepts [Popham, 1998]." `eduViz` allows teachers to visually explore assessment grades so that they can make the most of the information and time available.

Our own interest in this work comes from our experiences as students and as teachers. The first author spent a year teaching middle school math and all the authors have taught or been teaching assistants for undergraduate courses. Our hope as students is that our work will be examined thoroughly and help will be offered when appropriate. As teachers, we aim to identify and understand our students' struggles and successes. These experiences helped us to appreciate the vast amounts of data teachers attempt to process and the need for tools to help teachers easily analyze and understand the meaning behind the numbers.

Given the necessity for ease of use and for flexibility of queries based on a teacher's current interest, there were some unique spreadsheet and database design issues. Our solutions are explained in Sections 2 and 3.

`eduViz` supports two main functions. The first is grade exploration. In this section, users are able to compare students by any combination of three categories and see the results immediately displayed in bar graph form with markers indicating student or class average. For example, a math teacher could choose to view all of Alice's grades organized by assignment and data for word problems and compare her progress to the class average. Two exploration components are shown next to each other on the screen so that teachers can easily compare two queries. For example, a teacher may wonder if the final exam grades for their students were representative of their progress overall — these two graphs could be examined side-by-side. Grade exploration functionality and

implementation is described in more detail in Section 4.1.

The second `eduViz` function is grade assignment. Teachers can create statistics which will weight assignments as they prefer — for example, the final exam could be weighted as 30% of the grade. The resulting total will be displayed as a scatterplot with student grades sorted in ascending order. Letter grades can then be assigned based on these total scores. A novel multiple ranges slider, *Partition Slider*, is provided where each sliding marker represents a division between grades, and intervals between markers correspond to grade ranges. Dragging any marker resizes the corresponding bands of color on the scatterplot, indicating which students fall within the grade ranges with values below and above the marker. At the same time, the numerical grades are updated in a visible table and a histogram is updated indicating how many students or what percent of the class is currently assigned each letter grade. To our knowledge, `eduViz` is the only tool to support visual grade manipulation. More information about grade assignment options can be found in Section 4.2.

Though we all have teaching experience, we sought the advice of sixteen professional teachers. These were junior, intermediate, and senior teachers in computer science, biology, and math, and intermediate and senior teachers in non-technical subjects (e.g. dance, math, education, etc.). Their experiences using `eduViz` and their thoughtful suggestions for improvement are detailed in Section 5. These comments suggest that while `eduViz` has room for improvement, it is ultimately a success.

For help videos and code access see our website listed on the bottom of the title page.

1.1 Related Work

For the most part, use of computers in computer science education has fallen into two categories; the use of computers to aid student understanding [Byrne et al., 1999, Chang et al., 2001, Dewhurst et al., 2000, Weller, 1996, Kreijns et al., 2003], and the use of computers to help teachers by automatically grading student assignments (for a survey of the second see Douce et al. [2005]) [Page, 1994, Spacco et al., 2006, Jackson and Usher, 1997, Cheang et al., 2003, Edwards, 2003, Naur, 1964]. `eduViz` is a tool to aid in teacher understanding of student grades, and so it falls into a category of its own with no precisely comparable existing research literature. However, as this is an area of great importance, there are many existing commercial tools.

Existing comparable tools fall mainly into two categories; general use tools with similar capabilities which are not specifically for an educational context, and electronic grade book tools. Two such general use tools are `Spotfire` and

Excel [Ahlberg, 1996, Microsoft, 2007]. Spotfire is an information visualization tool with great flexibility. It is capable of generating all the desired graphs. Its drawbacks are that it is not accessible to teachers, since it is designed for industry use. It would also require teachers to create separate spreadsheets not in standard grade book format for each query, which would not easily allow for spur of the moment exploration. Excel is a highly useful and accessible spreadsheet application. It is probably used by many teachers already to create and maintain class information. One great advantage of this tool is that more advanced users can add logic which would allow them to, for example, drop the lowest quiz grade. Excel has visualization tools requiring some but not extensive training built in which could help to visualize the already present data. However these visualization tools also do not allow for equivalent spur of the moment exploration and, when using a standard grade book setup, require extra spreadsheet manipulation for visual or numerical comparison across categories.

Electronic grade books are now very common. Some examples which mirror most basic Excel capabilities, but add a few teacher specific features like per student progress reports are MyGradeBook, Gradelink, OnCourse Systems, Class Action Gradebook, and MathZone [Pearson Education, 2007, Gradelink, 2007, OnCourse Systems for Education, 2007, CalEd Software, 2007, McGraw-Hill Higher Education, 2007]. These tools have the advantage of being designed to perform the tasks that teachers desire, so teachers do not need to perform additional steps or reformat the data in order to get the information they need. Some of these tools (Class Action Gradebook and MathZone) also perform basic visualization actions — they allow student progress to be viewed as bar graphs or pie charts. However, they do not allow easy comparison of a student against the whole class. Turnitin [iParadigms LLC, 2007] is another grade book tool, however its visualization options are more extensive. Along with basic bar graphs and pie charts, it allows teachers to view individual student grades over time as compared to the class average and maximum possible points. This is the most advanced existing grade visualization tool that we know of. However, the view which includes the class average and maximum is graphed over discrete time or assignment intervals, but uses a line graph so that there is a misleading impression that there is a class average at intermediate time values. Turnitin also has the limitation that it does not allow teachers to dynamically explore their grades and design their own queries.

Extensive research has also been done on the decision making process. For a grade book and visualization application, decision making literature is relevant when letter grades are assigned. One of the best known and often used choice methods is the analytic hierarchy process [McCaffrey, 2005]. This process requires users to detail their hierarchy of desires and weight these wishes. For example, when deciding on a vacation, you might weight minimizing the

Figure 1: Example input spreadsheets

Assignment Type	midterm	midterm	midterm	final	final
Question Number	Q1	Q1	Q2	Q1	Q2
Subject	area	volume	volume	area	volume
Total Points	32	45	70	25	73
Alice	30	42	68	20	68
Bob	24	44	65	14	60

Assignment Type	midterm	midterm	final	final
Question Number	Q1	Q2	Q1	Q2
Total Points	77	70	25	73
Alice	72	68	20	68
Bob	66	65	14	60

Fig. 1. An example of a user's input spreadsheet. Note that some category labels appear more than once (this is common when using `Excel`'s cell filling).

cost as three times more important than the location. These weights are then put into a matrix and analyzed using linear algebra techniques to help decision makers arrive at a logical choice. In an educational and grading context, the hierarchy is often collapsed so that test grades are based only on one level of the hierarchy (the questions on the test) and final grades are also based on only one level of the hierarchy (the tests). The complete hierarchy process is therefore not required, but the idea of weighting options (tests) differently is frequently used by teachers. `eduViz` allows teachers to easily manipulate these percentages and results are immediately displayed in a context which allows teachers to examine the consequences of these choices and change their minds accordingly. In this way `eduViz` also supports the sense-making loop: Users begin by giving their students assessments (gathering information), then examine their information using the exploration tools (they re-represent it), then begin to understand their students' strengths and weaknesses (develop insight) and finally assign grades (produce results) before starting the loop again [Thomas and Cook, 2005].

Range sliders have long been used in visualization applications. Traditional range sliders [Ahlberg and Shneiderman, 1994, Li and North, 2003] allow a subrange on the slider to be specified by confining it between high and low draggable markers. Data Visualization Sliders [Eick, 1994] utilize the space

inside the slider to convey additional information about the selected data linked to the slider. They also support defining multiple intervals dynamically. Treemap [Chintalapani et al., 2004] uses a color binning slider which is closer in spirit to our approach. It allows dynamically adding and removing draggable bin separators as well as precisely setting their position by specifying their value in an associated edit box.

1.2 Main Contributions

`eduViz` addresses the limitations of current general visualization tools by allowing teachers to do grade book specific tasks and to view information by categories that they have assigned without needing to do additional data manipulation. Thus, `eduViz` is more teacher-friendly than a general tool. `eduViz` also improves on existing teacher focused grade book tools by allowing advanced visualizations and visual exploration of data. Teachers are not limited to a particular set of options, but can view any subset or combination of the data in which they are interested. In addition, they can compare data using our two exploration panels placed side-by-side or by viewing the data against the class or per student average. Our emphasis on the tools and ease of use the teacher needs, combined with the flexibility and power of a general visualization tool make `eduViz` unique.

In addition, there are several distinguishing attributes of our Partition Slider. First, our slider is applicable to general partition selection situations and it makes efficient use of space allowing the placement of multiple such sliders within one view. This compact form naturally extends to potentially turning any axis of a Cartesian coordinate system into a partition slider. For example, making the y-axis of a scatterplot a partition slider could be useful, though we don't extend the slider in this way in `eduViz`. Second, in our design the ranges between markers are first class objects, each having a distinct color. Thus, the movement of each marker can be visually correlated with selection in other widgets (e.g., the bands in the grades scatterplot with the corresponding color). Finally, our placement of the marker grip closer to the axis will improve the accuracy of partitioning using the mouse, as it avoids eye movement. To our knowledge, this is the first use of a partition slider of this form.

2 Spreadsheet Design

Our user input spreadsheet format is designed to accommodate both manual and automatically generated input. It mimics common grade book formats for easy import. User input is in the form of a .csv spreadsheet in which the

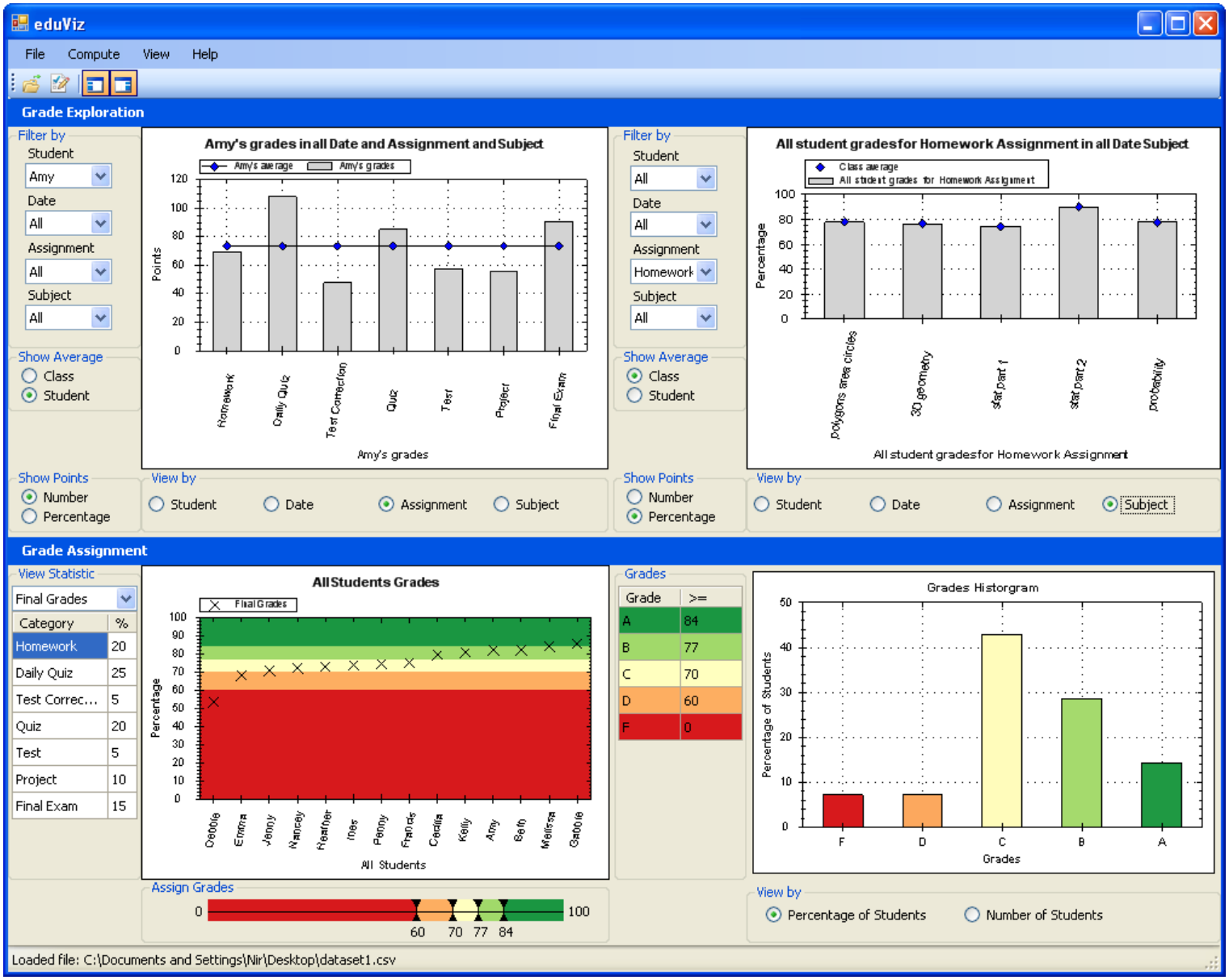

Fig. 2. The main screen of `eduViz`. The top panel allows side-by-side exploration of grades based on user defined categories (Date, Assignment, and Subject in this example). Information can be filtered and distinct views chosen so that the user can compare grades as desired. The bottom panel allows grade assignment using the partition slider shown on the bottom left. The resulting grades are shown in the scatterplot (bottom left) and histogram (bottom right).

first column contains student names and the first three rows contain category information. The fourth row contains the total number of points possible for that entry. All other cells contain scores which pertain to that student in those categories. If users do not want to use all three category possibilities they can leave the extra row(s) blank. See the spreadsheets in Figure 1 for examples.

The categories referred to here are user defined. For example, users may choose to record information about the type of assignment (test, quiz, project, etc.), the question number, and the subject tested. Or, they may wish to record the date of an assignment instead. We leave these choices up to the users to allow for the greatest flexibility of use — they may choose and enter information about any three categories of data that they have recorded. However, we do restrict this number to three, which is a reasonable restriction to the depth of categorization of a score given the domain. In future versions, we hope to allow this number to be dynamically set by creating an interface which would allow users to enter their data directly into `eduViz`.

3 Database Design

The database allows three user defined categories as well as students, scores, and maximum possible points to be tracked. These categories were made purposefully independent — no hierarchical assumptions are made. It is important for this application to maintain this independence so that teachers can use `eduViz` for the largest number of queries. For example, a teacher may want to know how all students did on the latest test and want to see that broken down by category within the test or they may want to see how students did on a subject and want to see that broken down by test within that subject. Our database design is correspondingly independent. Each score is stored in its most basic, most specific form. The unit which is stored in the scores table is the score for a particular student in the intersection of three categories. If a query asks only for two of these three categories, the score returned is the sum over the scores table of any entry identified by these two categories. For the detailed database design, see the database tables in Figures 3, 4, 5, and 6.

4 eduViz Implementation

`eduViz` is implemented in C# and the graphing capabilities are provided by `ZedGraph` [Champion, 2007]. Figure 2 shows the main screen for `eduViz`. The main screen is divided into a Grade Exploration panel on the top and a Grade Assignment panel at the bottom. The purpose of the Grade Exploration panel is to allow users to explore their dataset, discover any patterns or anomalies,

Database Tables

Table 1: students

Field Name	Type	Description	Example
pk_student_id	integer	Primary key for the student	236
student_name	string	Student's Name	John Smith

Fig. 3. Contains the list of all students taken from column 1 of the input spreadsheet.

Table 2: category 1

Field Name	Type	Description	Example
pk_cat1_id	integer	Primary key for the category	1748
cat1_string	string	String version of the category	Quiz #1
cat1_desc	string	Title/description of the category	Assessments
cat1_max	double	Max points for this category	100

Fig. 4. Contains the list of all categories taken from row 1 of the input spreadsheet. For example, all assessments the class was graded on. The description is taken from the first cell of row 1. Tables for categories 2 and 3 are constructed similarly.

Table 5: max_scores

Field Name	Type	Description	Example
pk_maxscore_id	integer	Primary key for the subcategory	7
fk_cat1_id	integer	Foreign key to category 1	1748
fk_cat2_id	integer	Foreign key to category 2	983
fk_cat3_id	integer	Foreign key to category 3	245
max	double	Max points for this subcategory	10

Fig. 5. Contains a list of the maximum scores which could be earned in each subcategory. For example, the maximum points that could be earned on the midterm on question 4 for computation.

Table 6: scores

Field Name	Type	Description	Example
pk_score_id	integer	Primary key for the score	643
fk_maxscore_id	integer	Foreign key to maxscore (table 5)	7
fk_student_id	integer	Foreign key to student (table 1)	236
points	double	Student's score	3

Fig. 6. Contains a list of the scores given to students for each category.

as well as confirm any speculation users have about how a particular student is performing. The grade assignment panel allows users to assign grades. These panels can be collapsed or expanded so that users can choose to focus on one at a time. However, the main screen is purposely designed so that all components

can be viewed at the same time without having to switch between windows to reduce short-term memory load as suggested by the "eight golden rules of interface design" [Shneiderman and Plaisant, 2004].

4.1 Grade Exploration Panel

The Grade Exploration panel (top of Figure 2) consists of two identical plots areas which serve to provide users with an overview of the dataset. Each plot area allows users to select a different subset of the dataset to view. Each plot area is associated with a graph and two control panels, one to the left of the graph and one at the bottom. The graph supports zooming and panning via mouse scrolling. In addition, the x-y values are displayed when a cursor hovers over individual data points in the plot. `eduViz` also provides for copying the graph to the clipboard, saving the graph as an image, and printing the graph directly.

Two exploration plots are provided so that users can easily compare the responses to two queries against each other. For example, a teacher may wonder if the grades their students earned on the final exam were representative of their overall progress in the class. Each of these graphs could be viewed in an exploration plot and then they could be compared side-by-side.

The control panel to the left is divided into 3 groups: a filter-by group, a show-average group, and a show-points group. The bottom control panel consists of the view by group. The x-axis of the graph is determined by the user's selection in the view-by group while the y-axis is determined by the selection in the show-points group.

Control Panel

(1) The *filter-by* group allows users to filter the dataset by any combination of student and the three categories in the dataset.
(2) The *show-average* group allows users to determine, for a given view, whether the average for the class or the student is displayed.
(3) The *show-points* group of radio buttons allows users to plot either the raw data values or percentages.
(4) The *view-by* group provides users with the option of projecting the filtered data onto different views and to selectively see each of the views.

Supported Functionality We will illustrate how the functionalities provided by the graph exploration panel can be used to achieve a few example

tasks using a sample dataset collected for a period for three months where each student has grades for assignments which are identified by date, assignment type, and subject. Multiple assignments can be assigned for each date, and assignments can be of different types (e.g. homework, daily quiz, test correction, etc). Subject refers to the material covered by the assignment. The date, assignment type, and subject map naturally to unique categories in `eduViz`.

Users who are exploring this dataset may be interested to see how the class as whole did in each assignment type. `eduViz` supports this by allowing users to view the (unfiltered) dataset by assignment type. Alternatively, users may also want to see how the class did by subject material. This can be accomplished by selecting the subject view on the same (unfiltered) dataset.

Users may also want to see how a particular student, say Amy, is doing in class. Users can filter the dataset by Amy and then select either the assignment type or the subject view depending on whether they are interested in seeing Amy's performance in the different types of assignments or material covered in class. Users can compare Amy's performance on each of the components of the assignment type (or subject) category to the class by selecting to show the class average. The class average for each of the category components will be plotted on the same graph. Alternatively, users can compare Amy's performance on various assignments or in various subjects to her own average by selecting to view the student average. In this case, the student average is simply a straight line across the graph.

4.2 Grade Assignment Panel

The Grade Assignment panel (bottom of Figure 2) consists of two plot areas, a scatterplot and a histogram. The scatterplot is based on a derived statistic while the histogram shows the number or the percentage of students that fall in each grade category. Both the scatterplot and grades histogram are displayed after users specify a derived statistic on which the grades would be assigned.

One of the main design objectives was to use a unified color scheme for the different parts of the Grade Assignment panel. By having the scatterplot, histogram, partition slider, and grades grid all "speak" the same color language, the coordination between them is significantly enhanced and emphasized. Using consistent colors for different grades makes it easier to correlate various views of the same data. The multiple views work together by having changes reflect in all of them simultaneously. For example, manipulating the partition slider makes the grade bands in the scatterplot resize accordingly, the values in the grades grid update, and the histogram bins show the current grade distribution. Our choice of colors was based on the convention of using red to

signify caution, or in our case a *failing* grade. Green was chosen to symbolize a grade of `A` since it has good and desirable connotations. Matching colors for the other grades were chosen using Color Brewer [Harrower and Brewer, 2003].

Scatterplot with Bands After loading a raw dataset, users can create a derived statistic via the Compute New Statistic dialog (Figure 7) such as the overall score based on the weighted average of a selected set of category components. These components (for example, assignment types) are loaded automatically and the user enters the percentage totals.

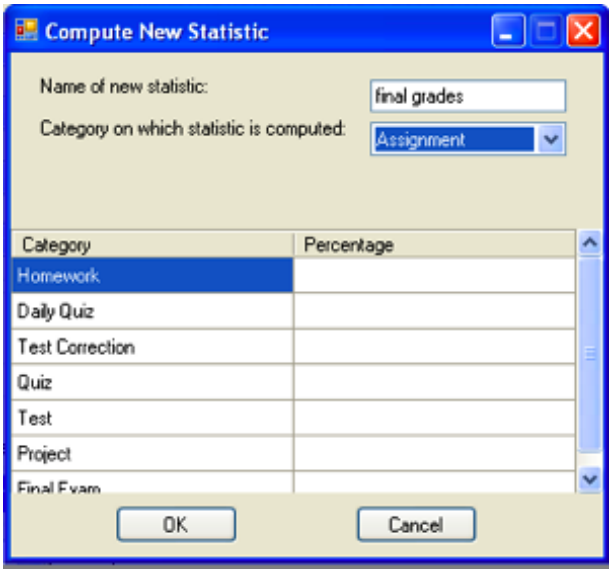

Fig. 7. The compute new statistic dialog.

The student independent variable is plotted on the x-axis of the scatterplot while the derived statistic is plotted on the y-axis. The derived statistic is sorted in increasing order. Sorting by the derived statistic helps to see the natural clusters of students with close grades. The scatterplot is overlaid with colored bands corresponding to different letter grades to help in the grade assignment process (see bottom of Figure 2). The position of these bands is controlled using the partition slider, which allows teachers to immediately see the effect of their grading choices.

Partition Slider `eduViz` currently provides a default set of 5 grade categories (`A`, `B`, `C`, `D`, and `F`). Correspondingly, there are 4 markers on the partition slider interspersed between 5 color bands. Coordinated color bands are also overlaid on the scatterplot, showing the boundaries of each grade category (see Figure 8). These bands help users visualize the cutoff points for the grades. The range for each grade category can be manipulated using the partition

slider since each band separation in the scatterplot is synchronized with one marker on the slider. That is, dragging a marker on the slider, updates the position of the two corresponding bands on the scatterplot.

The numerical values in the grades data grid as well as underneath each marker in the slider show the cutoff scores for each grade category. The grades grid also shows the letter grades corresponding to the different score ranges. Manipulating the partition slider causes the histogram and bands in the scatterplot to be updated as shown in Figure 8 part A (before, with markers set at 60, 70, 77, and 84) and part B (after, with markers set at 60, 68, 82, and 88). All parts of the display are updated dynamically as the markers are being dragged.

Fig. 8. Manipulating the partition slider updates the histogram and scatterplot dynamically. Top (A) and bottom (B) show the view before and after the change, respectively.

The scatterplot superimposed with bands helps users who do not want to know just the percentage of students that fall into each category, but also if there are any clusters that might be divided into two when a grade cutoff is set. For instance, in the scatterplot section of Figure 8, if there is an obvious cluster of students with overall scores between 77–83, and a cluster of students with overall scores between 70–76, the user may, on looking at the cutoff scores, decide that it is not desirable to break the 77–83 cluster at 80 and would rather give every student in that cluster a `B` grade. Similarly, selecting the `A` grade at the range of 84–100 is shown to be a logical choice because it preserves a cluster at about 84–86 that is significantly different from the next lower cluster

of scores.

Histogram The purpose of the grades histogram (Figure 8) is to help users see the number or percentage of students that fall into each grade category based on the ranges specified using the partition slider. This visualization is particularly helpful to users who are grading on a curve, as it allows them to adjust the slider to achieve the desired grades distribution.

5 Domain Expert Consultations

While constructing and testing `eduViz`, we had the opportunity to work with sixteen teaching domain experts. The goal of these interviews was not experimental, but to glean informal advice, criticism, and suggestions which would allow `eduViz` to evolve and improve and would allow us to assess the achievement of our goals. These expert reviews were conducted in the vein of the "cognitive walkthrough" as described in [Shneiderman and Plaisant, 2004] where experts carry out the common tasks of the software and meet with the developers to discuss their experience.

The 16 experts we interviewed fell into three categories based on their years of teaching and grading experience; those at Associate or Full Professor rank or with more than 7 years of experience (who we will refer to as "senior experts"), those with between 6 semesters and 7 years of experience (who we will refer to as "intermediate experts"), and education students who have completed student teaching (who we will refer to as "junior experts"). These experts were diverse in their subject field of expertise. Among the eight senior experts, half were from technical fields (including computer science and mathematics) and half were from social sciences or humanities (including dance, german, education, etc). The six intermediate experts were similarly split, with half from technical fields and half from social sciences and humanities. The two junior experts were training to be teachers of mathematics. We met with each of these experts once individually towards the end of our development process. Half of these sixteen experts were interviewed in a first round of interviewing; half were interviewed after software updates were made.

5.1 Discussion Format

Our discussions with these 16 experts were somewhat structured, though they allowed for frequent interjections and guidance by the experts. We interviewed

these experts one-on-one for a total of 30 minutes to an hour each. One developer and one expert were present at each interview, and the developer took observational notes by hand. While exploring, the experts were encouraged to talk about what they were experiencing and doing and were prompted to do so when they fell silent. The general format was as follows:

(1) **Explanation and introduction to** `eduViz`: First we explained our basic goals which were to allow teachers to explore grades and get an understanding of their students' strengths and weaknesses and allow teachers to assign grades. Next we described the example dataset that we used to demonstrate the tool. This was a detailed grade book spreadsheet from three months of an 8th grade math class which the first author taught in Spring 2005. The dataset includes three categories associated with each grade — the date the assignment was due, assignment type, and subject covered on that assignment. The total and individual scores were also provided in the spreadsheet format given in Section 2. Student names were changed to provide anonymity.
(2) **Guided introduction to** `eduViz`: To provide users a basic introduction to the tool, we explained how to do the following tasks and asked them to perform the actions. Each action was accompanied by an explanation of the general interface and the current display.
    (a) Load file
    (b) Choose a specific student to view
    (c) Choose to view by assignment
    (d) Choose to show the average per class
    (e) Create a new statistic for the class total scores
    (f) Use the partition slider to manipulate grades and make the histogram change
(3) **Guided exploration of** `eduViz`: Our experts were then given some suggestions of questions they could try to answer about the dataset. This was done so that they could see for themselves the kinds of exploration and discoveries that are possible with `eduViz`. When experts had trouble viewing the information in the way they wished, we reminded them of the options and basic functions and helped them find the appropriate solution. Here are some examples of the questions we offered.
    (a) Find a student who was below the class average for every assignment type.
    (b) Find a student who did much worse on geometry than their usual progress.
    (c) Did the final exam grades reflect students' progress throughout the class?
    (d) Why is the class average for March 22nd lower than usual?
(4) **Unguided exploration and feedback**: Finally, we allowed experts to explore the tool on their own while thinking about any feedback they had for us. In order to elicit a more extended, thoughtful, and consistent

response from all experts, we made sure to ask each expert the following questions.

(a) What do you think of the exploration section? ...of the grade assignment section?
(b) What else would you like the tool to have?
(c) What was confusing?
(d) What was beneficial?
(e) How does `eduViz` compare to other tools that you've used or know about? Benefits/disadvantages?
(f) Is there anything else we should think about or that you want to share?

5.2 Expert feedback

Throughout this section we will use a naming system to provide information about which experts made the following comments. The first set of one or two letters refer to the experience level of the expert ("SR" for senior experts, "I" for intermediate experts, and "JR" for junior experts). The second set of letters refer to the discipline in which that expert teaches ("T" for technical fields and "NT" fields in the social sciences or humanities). The suffix contains one number which distinguishes between multiple experts with the same background and one letter which distinguishes between the two rounds of interviews ("A" for experts interviewed in the first round of interviews, "B" for experts from the second). A full naming might be "SR_NT_2B" for an expert interviewed in the second round of interviews who was the second senior expert in a non-technical field. We begin by categorizing the expert responses thematically and then move to more anecdotal descriptions and analyses of the interviews based on these categorizations.

Using thematic interview analysis [Rubin and Rubin, 2004, Kvale, 1996] we categorize the feedback given by the experts with the codes as described in Figure 9. This coding scheme allows categorization of general comments as well as comments specific to the grade exploration, assignment, and help sections. Within these specific areas comments were additionally described as positive, negative, or suggestive improvement comments on specific features a s well as a few descriptive usage comments.

For additional analysis, the codes were grouped according to their broader categorizations — positive, negative, or suggestive improvement comments. These broader categories included the codes ending in "GOOD," "BAD," and "IMPROV" respectively, except for the codes beginning in "COMPETITOR," since the goal of this analysis is to evaluate `eduViz`. These broader groupings were then analyzed for trends based on the differing expert characteristics. A

chart giving the number of comments of each type as normalized based on the total number of experts with the given characteristic is shown in Figure 10.

Figure 9: Code Dictionary

Code	Meaning
NO-CODE	Any independent thought in the log that we do not code
EDUVIZ-GOOD	Positive comment about `eduViz`
EDUVIZ-BAD	Negative comment about `eduViz`
EDUVIZ-USAGE	Description of how `eduViz` can be used or of expert's use
EDUVIZ-VIDEO-GOOD	Positive comment about the video tutorials
EDUVIZ-VIDEO-BAD	Negative comment about the video tutorials
EX-GRAPH-GOOD	Positive comment about the grades exploration graphs
EX-GRAPH-BAD	Negative comment about the grades exploration graphs
EX-SECTION-GOOD	Positive comment about the grades exploration section, which is not specific to the graphs or the view-by/filter-by options
EX-SECTION-BAD	Negative comment about the grades exploration section, which is not specific to the graphs or the view-by/filter-by options
VIEWBY-GOOD	Positive comment about the view-by radio buttons
VIEWBY-BAD	Negative comment about the view-by radio buttons
FILTERBY-GOOD	Positive comment about the filter-by menus
FILTERBY-BAD	Negative comment about the filter-by menus
AS-GRAPH-GOOD	Positive comment about the grades assignment graphs
AS-GRAPH-BAD	Negative comment about the grades assignment graphs
AS-SECTION-GOOD	Positive comment about the grades assignment section, which is not specific to the graphs or the range slider
AS-SECTION-BAD	Negative comment about the grades assignment section, which is not specific to the graphs or the range slider
SLIDER-GOOD	Positive comment about the range slider
SLIDER-BAD	Negative comment about the range slider
EDUVIZ-IMPROV	Idea on how to improve `eduViz`
EDUVIZ-HELP-IMPROV	Idea on how to improve the help provided by `eduViz`
STUDY-IMPROV	Idea on how to improve the user study
EX-SECTION-IMPROV	Idea on how to improve the grades exploration section
EX-GRAPH-IMPROV	Idea on how to improve the grades exploration graphs
VIEWBY-IMPROV	Idea on how to improve the view-by radio buttons
FILTERBY-IMPROV	Idea on how to improve the view-by menus
AS-SECTION-IMPROV	Idea on how to improve the grades assignment section
AS-GRAPH-IMPROV	Idea on how to improve the grades assignment graphs
SLIDER-IMPROV	Idea on how to improve the range slider
COMPETITOR-GOOD	Positive comment about a competitor tool
COMPETITOR-BAD	Negative comment about a competitor tool
COMPETITOR-USAGE	Description of how a competitor tool is used, or description of how participant likes to use a competitor tool

Fig. 9. The code names and associated descriptions used to categorize interview comments.

General Impressions General feedback and impressions of `eduViz` by our 16 expert consultants were positive. They included the comments "I like this program" and "It's something I would consider purchasing" (SR_NT_2B), "I would use this" (I_T_3B), "I like the concept" (SR_NT_3B), and other simi-

Figure 10: Normalized Comment Numbers by Expert Characteristics

	Non-Technical (7)	Technical (9)	Junior (2)	Intermediate (6)	Senior (8)
GOOD (87)	7.1	4.1	6.0	7.0	4.1
BAD (25)	1.1	1.9	1.0	1.7	1.6
IMPROV (93)	5.4	6.1	3.0	5.2	7.0

Fig. 10. Numbers of comments for type of comment and characteristic of expert. Normalized by dividing the total number of comments by the number of experts with that characteristic. Parenthesized values indicate total number of experts or comments per category.

larly positive comments. One expert additionally expressed the desire to get the final version to share with colleagues (SR_NT_1A). These general positive comments were coded as "EDUVIZ-GOOD," and there were 61 such comments overall, an average of 3.8 comments per expert. Despite these comments, some experts were worried that parts of the tool might not be useful for their needs and encouraged us to make improvements. This interest in our tool led the experts to make a total of 93 suggestive comments about `eduViz`. These were skewed so that that senior experts had more recommendations to offer than those with less teaching experience — this trend held for intermediate experts as compared to junior experts as well. Overall, the teachers seemed excited and interested in the prospect of a new tool which could potentially make assessment easier, but wary that the tool would actually cause more work than their current method. A few also commented that because their classes are small they don't need information of this detail (SR_T_4B, SR_NT_4B), though SR_T_4B later commented that `eduViz` might be a tool which initially seems unnecessary but once available becomes something you use all the time. Perhaps this observation is more simply summed up by JR_T_1A's comment that "this is fun!"

The experts we spoke to were enthusiastic about the possibility of being able to visualize their student data. A benefit of the simplified analysis process is that it supports in-depth and exploratory student evaluation. It allows teachers not just to assign grades, but also focus attention and intervene to assist students. One professor of education wanted to make sure that `eduViz` made these links to the real world clear and suggested incorporating these options into the help videos (SR_NT_3B). The experts envisioned many other ways to use `eduViz` as well. Two experts mentioned the advantages of being able to compare a current class against previous classes' performance on comparable assignments (SR_NT_1A, I_T_1A) — this is something that `eduViz` can do if the class semester is chosen as one of the three optional categories. Another expert commented similarly that it would be useful to compare multiple sections of the same class against each other, especially if worried about the comparative grading of different sections' teaching assistants (I_T_3B). SR_T_4B suggested that `eduViz` could be used to investigate disparity in performance between tests and homework, when students complain that they "freeze" under the pressure of exams. Another use he suggested was middle or high school

teachers using `eduViz` to prepare for parent teacher conferences. Finally, he identified three ways `eduViz` could be used to increase teacher effectiveness: identify individual student problems, see problems that the class is having as a whole, and see issues that the teacher is having. One expert, who studies education professionally and is interested in teacher training, said that `eduViz` will encourage teachers to be "reflective practitioners" and that it motivated her to change her practices (I_NT_2B). Expert SR_NT_3B also saw `eduViz` as a tool that could be used in teacher training programs and commented that it could also be used for more general research by rendering descriptive demographic displays which are not necessarily grades related.

While some experts were skeptical about this tool, worried that it would create more or unnecessary work, others found it appealing that `eduViz` saved them a lot of calculations. Data analysis that is normally labor intensive was made easy. "Since you don't have to average or calculate anything, it's really useful, and you never need to do anything other than input (I_NT_1B)." I_NT_3B commented that `eduViz` "does the computing for you" and presents an "easy visual so you don't have to consider numbers." Other experts have commented that for the benefit it provides it wouldn't take too much extra time and so would be possible to use (I_T_3B, I_NT_3B), though they did not comment directly on the benefit provided by the automatic calculations. Many of these experts already have established methods and tools for avoiding manual laborious calculations — they mostly use Excel or another competitor tool to help calculate averages and other class statistics.

In addition to the critiques we received about `eduViz`, our experts had a plethora of suggestions for enhancing `eduViz`. Many of these suggestions are given below in the context of the specific section to which they refer. However, some suggestions are more broadly applicable and we include those here. Several interviewees wanted to see greater intelligence injected into eduViz, bordering sometimes with an expert system approach. This included suggesting a selection of interesting views to look at (I_NT_2B, I_T_3B) and a list of suggested real world remedies (SR_NT_3B). I_NT_1B recommended automatic grade anomaly detection to find grade input mistakes, and SR_T_2A and SR_NT_2B wanted the means to readily see which students improved most. Several experts wanted to be able to view and edit grades directly from within `eduViz` (I_NT_2B, SR_NT_4B) and also synchronize with other software (e.g., Excel or specialized gradebook systems). Synchronization with online servers would be useful for students to track their own performance and progress (I_T_3B, SR_NT_2B). In addition there were a number of smaller requests inherent to the discipline: SR_T_2A, SR_NT_2B, and SR_T_4B desired native support for extra credit that does not go into the overall percentage, and the ability to select multiple students for comparison was requested by SR_NT_3B. The excitement over the tool and the ability to suggest electronic teaching needs to a receptive audience also led to many suggestions outside

the scope of this project which would push `eduViz` to become an all-in-one tool to help teachers manage everything from attendance to communication with students to complex statistical analyses (SR_T_2A, SR_T_3A, SR_NT_1A, I_NT_1B, SR_NT_2B).

Comparison With Other Tools Currently, the 16 experts we interviewed use `Excel`, Blackboard, WebCT, department run grading and submit servers, and paper grading books as their tools of choice. One professor who was especially impressed by `eduViz` mentioned that she had tried other tools, but found them frustrating to learn and use (SR_NT_1A). In general, users of Excel did not use the graphing feature since they found it to be too time consuming and confusing (SR_NT_1A, I_T_2A, JR_T_2A, SR_NT_2B). One expert commented that his department's grades server can assign grades but doesn't show the histogram which would be useful (SR_T_1A), though another expert who uses the same grades server said that the functionality is sufficient for the level of effort in evaluation he can afford to spend (SR_T_4B). Two users of the Blackboard system indicated that the statistics are given numerically, making any visual comparison of student performance hard to do (I_NT_1B, I_NT_2B) since the graphing capabilities of Blackboard are restricted and do not exist for grades (I_NT_1B). While WebCT does include graphing capability, I_T_3B commented that its visualization choices are not as dynamic or "intuitive" as those provided by `eduViz`. However, `eduViz` does not allow for complex cell formulas as Excel does (I_T_2A) or for students to access their grades as they are entered into the spreadsheet as Blackboard, WebCT, and the department run servers do (I_NT_1B) and some experts noted that these functions are also desirable (I_T_3B).

Though some experts commented that figuring out how to best use `eduViz` would take some time, they felt practice would make the usage clear (I_T_3B, JR_T_2A). Other experts found the point and click interface "inviting" (I_NT_2B) and disagreed with the idea that using `eduViz` would take practice saying, "I like that it's so intuitive and you don't have to go back to the help pages (I_NT_3B)." One professor verified the research which shows that teachers spend much of their time grading assessments by her comments that "I really struggle because I know that whether I say `B+` or `A-` makes a big difference" and "my guess is that this is where a lot of professors spend their time." She was especially excited that `eduViz` would save her time and allow her to reason about and justify her grades more effectively (SR_NT_1A). Overall, `eduViz` could empower these teachers to perform comparisons and analysis that they rarely do (I_NT_2B).

Spreadsheet, Database, and Visual Design Our spreadsheet design was purposefully made to be similar to a grade book so that teachers would find

it easy to input grades in their usual format. Our experts verified that we had chosen an appropriate format which mimicked the grade books that they keep. We also designed our database with a great amount of hierarchical flexibility. Many of the experts commented on this flexibility, appreciating that `eduViz` is "capable of doing logical comparisons by anything you want (SR_T_3A)," that you can "tailor it to what you want to see (I_T_3B)," that `eduViz` is flexible and allows comparisons "across all parameters (I_NT_3B)," and that you can compare "anything (I_NT_2B)." One commented that it is necessary to have this level hierarchy in order to examine the strengths and weaknesses of individual students (SR_NT_1A). Our experts strongly supported our database design choices.

The 16 experts also appreciated many of our design choices. We especially concentrated on reducing short-term memory load by creating a single screen display as suggested by the eight golden rules of [Shneiderman and Plaisant, 2004]. This single screen display and the logical grouping within it was specifically complimented by I_T_3B and SR_T_3A. In addition, several experts complimented the "clean, simple interface (SR_T_4B)," saying that it "looks pretty and clear (SR_NT_4B)," and appreciating the easy access to different graph plots (JR_T_2A, I_T_2A).

Grade Exploration Panel The experts we spoke to made many specific comments about the grade exploration section regarding general layout, functionality, and specific features. We won't list all of these comments here, but will give a general sense of the trend of the issues and benefits of the current version of `eduViz`. Many experts commented, as mentioned earlier, that they appreciated the functionality of the grade exploration panel — allowing them to evaluate individual student performance on particular topics (SR_T_4B), compare students against themselves and the class (SR_NT_2B), and examine different slices of the data (SR_T_2B, SR_NT_4B).

Some users were confused about the difference between the data selection options and the view-by options and could not figure out how to find the view they wanted (SR_T_1A, I_NT_1B, I_NT_2B, SR_NT_4B). The interaction of these data viewing and filtering options can lead to degenerate graphs that are of little interest — expert I_NT_1B suggested that `eduViz` could warn users if they were going to reach one of those views. However, as noted earlier, experts commented that after practice they believed this would become easier and some contradicted the negative comments by stating that the interface is intuitive. To address this confusion, we reorganized the layout to make the view-by buttons more visible and added help videos. Several experts also commented that the side-by-side comparison requires back-and-forth eye movement and this could be avoided by using double bars (SR_NT_2B, SR_NT_3B, I_NT_2B, and SR_NT_4B). One expert commented on the current default value

of zero which is given to any missed assignments (a student with no homework score is by default given a zero) (SR_T_2). We had originally thought that this could be a useful way for teachers to identify students with missing work; a student with an unusually low score for an assignment as compared to the class average may have never turned it in. However, after discussion with this expert, it seems that this use is not readily apparent. If `eduViz` is also used as a grading tool, this assumption of zero may even be unfair to the student. In future versions of `eduViz` we will allow the teacher to choose if the default value should be zero or if a flag should be set so that missing work is clearly and visually marked. In addition to these comments, there were also suggestions about general panel organization (I_NT_1B, I_T_3B, SR_NT_4B) and additional data filtering options (SR_NT_3B, SR_NT_2B).

Grade Assignment Panel Several experts appreciated the visual clarity of the grade assignment panel — I_NT_2B found the bottom panel easy to process immediately, aided by the choice of colors, and I_T_3B also singled out the color scheme as a useful feature. The control of the grade divisions by the partition slider was also complimented (I_NT_1B, JR_T_1A). The ability to easily visualize grade distribution is also a valuable contribution of this section (I_T_3B).

Several suggestions for improvement were proposed. SR_NT_2B would like to be able to automatically see where to break the grades in order to achieve a normal curve, and visualize how far the chosen grade binning is from a normal curve. SR_T_4B, more generally, suggested superimposing different statistics. Experts also wanted more individualized manipulation of grading statistics. For example, SR_T_3A wanted to be able to drop a student's lowest quiz grade and SR_T_1A would like support for curving individual assignments before averaging the grades, due to variations in standard deviation between assignments. In future versions we will address this by allowing teachers to overwrite the general class assignment weights with different weights for individual students as well as by adding a basic programming language. We also had a few requests from our first round of interviews to sort the grades in the scatterplot by ascending order so that clusters would be more apparent (JR_T_1A, SR_T_1A, SR_T_2A) and to change our original cutoff lines to bands of color showing grade ranges (I_T_2A, SR_T_2A). We changed this section and accompanying partition slider and histogram accordingly.

Our discussions with the experts crystallized the distinct user types of `eduViz` and their varying needs. Experts felt that for users who grade by curve the grade assignment part of the interface would be particularly helpful, especially if it saved them manual `Excel` work (I_NT_3B). These users were treating the grading process as an approximate clustering algorithm, trying to assign the same letter grade to a group of students with similar overall grade. For these

users, usually the ability to change weighting of different assignments is of little value, since instead of using this repetitive trial-and-error process they can more conveniently move the grades cutoffs to achieve the desired effect (SR_T_1A).

Other users, who use fixed cutoffs, required the weighting manipulation interface. However, even they commented that they are bound by published weights for each assignment category (e.g., for quizzes, projects, homework, etc.) but can manipulate freely the weighting of individual assignments within each category (I_T_1A). Some of these users pointed out that `eduViz` could help them better choose test questions in the long run, allowing them to spot aberrations that may point to problematic questions or topics that need to be taught differently [Walvoord and Anderson, 1998] (I_T_1A).

Usability A few senior experts mentioned usability concerns, mostly associated with older users and eyesight. One commented that our color choices of blue and grey are especially easy on older eyes (SR_NT_1A). Another mentioned that the radio buttons being used to change views were too small to see easily (SR_T_3). There are also usability concerns with respect to platform, since `eduViz` is written in C# and so works only on Windows machines. Experts who work on a Macintosh were especially disappointed by this (SR_NT_1A, .SR_NT_2B).

Scalability Another problem which has the potential to limit usability is the issue of scalability. Two experts in the first round of interviews worried that `eduViz` would not be able to handle the datasets they work with containing, potentially, 100–200 students and numerous assignments (I_T_1A, SR_T_2). We have since tested `eduViz` on a dataset of over 100 students and found the response times to be observably similar to the times for smaller datasets. While the bars on the graphs for this larger dataset were very thin, they were still able to fit on the screen and `eduViz` has zoom capability for users who want to see individual bars more clearly (Figure 11 shows the effect of using zoom on a large dataset). The default layout splits the screen into four sections with one graph per section, but these sections can be hidden (i.e. collapsed) or resized to cover the entire screen. Larger data sets can also be understood by using the filtering and sorting options provided. Given the application domain and therefore, the constraints on the number of students or assignments, this amount of scalability seems reasonable. The constraint that is more likely to limit the usage of `eduViz` under scalability is the constraint on the teacher's time. In future versions, we hope to implement a rank-by-feature framework to help identify struggling students [Seo and Shneiderman, 2005]. This would allow teachers to avoid searching through large class data themselves for unusual or troubling grades. Still, as noted by one of our experts, even without

the rank-by-feature framework, `eduViz` could prove useful to professors who have teaching assistants do their grading and therefore can sometimes lack insight into understanding their students' progress. Since `eduViz` can handle large classes, this exploration could be very useful to the teacher (I_T_1A). In our second round of interviews, multiple experts pointed out that `eduViz` would actually be especially useful when dealing with many data points (students or assignments) since it would help to filter and understand the data (SR_NT_4B, I_NT_1B). The only skeptic in the second round of interviews (I_NT_1B) was convinced after being shown an example with many students that `eduViz` would still be useful for large classes.

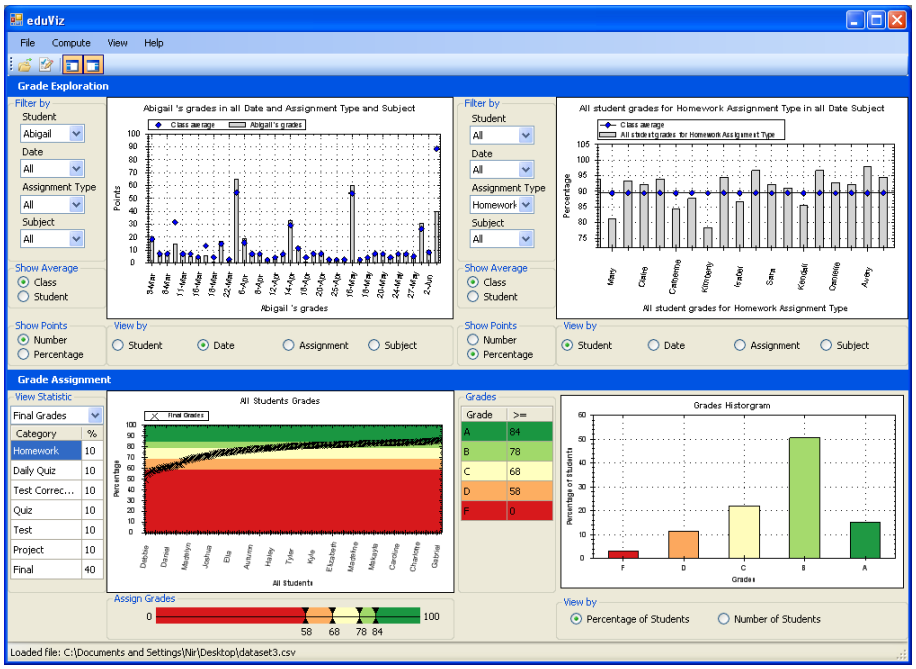

Fig. 11. Using zoom (top-right view) on a large dataset.

Expert Characteristics Analysis One interesting pattern we noticed while speaking to experts was that they fell clearly into the marketing categories outlined by Moore [1999]. The younger (junior and intermediate) experts and the experts who teach non-technical subjects were generally more excited about the tool and interested in adopting it in their own assessment practices as reflected by their higher normalized number of GOOD comments (see Figure 10). Although these users had many comments about ways in which `eduViz` could be improved, the overall impression they conveyed to us was positive. These are characteristics of what Moore calls "early adopters," users who are comfortable enough with the tool so that they can and are willing to make up for any deficiencies they find. In contrast, the senior experts and experts from technical fields were much more interested in knowing if `eduViz` had the functionality that their current grading methods did. They were less interested

in the grade exploration aspects. One commented that he had too many students in lower-level classes to actually have time to explore students' grades (SR_T_1A). Another wanted ways to drop the lowest quiz grade, give the student a higher grade if they did well on the final exam, or perform other, less generic manipulations of student grades (SR_T_3A). These experts said they were unlikely to adopt the use of `eduViz` unless these features were available. The impression they conveyed was that they were interested in the tool, but wouldn't use it unless it was more polished, had all the features they wanted, and would obviously save them time and effort. Many of the senior experts also felt that their current methods of grading were sufficient for their purposes, and so saw no reason to adopt new strategies (SR_T_4B, SR_NT_3B, SR_NT_4B). These are characteristics of what Moore calls "pragmatists," users who won't use the tool until it is fully functional and in mainstream use. Since `eduViz` is still in production, it makes sense that while "early adopters" would be excited by what they see, "pragmatists" would still be skeptical.

6 Future Work

During our consultations with the experts, it became apparent that they often look for certain phenomena in the data. For example, it would be interesting to find students who had an exceptional grade in a specific assignment, students that are below/above class average, etc. Considering the significant workload of teachers, it is improbable that they would have time to look for these patterns manually, especially in bigger classes. Therefore, a system that could flag situations corresponding to preset queries could improve teaching quality without imposing an additional burden on teachers. We think that adding a rank-by-feature capability could direct and improve the quality of the data exploration [Seo and Shneiderman, 2005]. Since we are dealing with exploration of a specific type of dataset, it will be easier to categorize certain queries as representing desired situations and other as needing special attention. Ultimately, we envision a system that calls teachers' attention to any potential problem and then allows them to explore the data to determine whether intervention is needed and get further insight for devising a plan of action. To automate this warning system, we intend to provide a basic internal macro language to automate certain processing. The challenge would be to provide this advanced programming interface without alienating novice users.

Currently, `eduViz` defines weighting of grades in a uniform manner for all students. This does not leave room for special consideration due to personal student circumstances which may affect the student's performance. Also, there is no way to benefit a student that is showing improvement. Finally, it does not consider less quantifiable grade components such as class participation. We would thus like to support individual student statistics that will override

the class statistics based on the teacher's judgment. A related topic is that of missing work. These may sometimes be counted as a zero grade in a specific assignment and in other occasions, other strategies would fit better. `eduViz` needs to provide this flexibility in order to gain wide acceptance.

7 Conclusion

We have presented `eduViz`, a teacher aid for grade exploration and assignment. `eduViz` aims to transform raw evaluation in the form of grades into actionable insight that allows teachers to spot areas of difficulties in the subject matter, assist students, and provide more appropriate feedback. We see great potential in the application of Information Visualization techniques to evaluation processes and believe it can improve assessment accuracy as well as the teaching process itself. The sixteen expert teachers who we interviewed verified this potential and provided many suggestions for the future improvement of `eduViz`.

References

Christopher Ahlberg. Spotfire: an information exploration environment. *ACM SIGMOD*, 25(4):25–29, December 1996.

Christopher Ahlberg and Ben Shneiderman. Visual information seeking: Tight coupling of dynamic query filters with starfield displays. In *Human Factors in Computing Systems. Conference Proceedings CHI'94*, pages 313–317, 1994.

Michael D. Byrne, Richard Catrambone, and John T. Stasko. Evaluating animations as student aids in learning computer algorithms. In *Computers and Education*, volume 33, pages 253–278, 1999.

CalEd Software. Class action gradebook. www.classactiongradebook.com, July 2007.

John Champion. Zedgraph. www.zedgraph.org, July 2007.

Kuo-En Chang, Yao-Ting Sung, and S. F. Chen. Learning through computer-based concept mapping with scaffolding aid. In *Journal of Computer Assisted Learning*, volume 17, pages 21–33, 2001.

Brenda Cheang, Andy Kurnia, Andrew Lim, and Wee-Chong Oon. On automated grading of programming assignments in an academic institution. In *Computers and Education*, volume 41, pages 121–131, 2003.

Gouthami Chintalapani, Catherine Plaisant, and Ben Shneiderman. Extending the utility of treemaps with flexible hierarchy. In *Proc. International Conference on Information Visialization*, pages 335–344, 2004.

David G. Dewhurst, Hamish A. MacLeod, and Tracey A. M. Norris. Inde-

pendent student learning aided by compters: An acceptable alternative to lectures? In *Computers and Education*, volume 35, pages 223–241, 2000.

Allen E. Doolittle. Classroom assessment: What teachers need to know (second edition). *Journal of Educational Measurement*, 39(1):85–90, 2002.

Christopher Douce, David Livingstone, and James Orwell. Automatic test-based assessmed of programming: A review. In *Journal on Educatinoal Resources in Computing (JERIC)*, volume 5, page Article No. 4, 2005.

Stephen H. Edwards. Teaching software testing: automatic grading meets test-first coding. In *Companion of the 18th annual ACM SIGPLAN conference OOPSLA*, volume 18, pages 318–319, 2003.

Stephen G. Eick. Data visualization sliders. In *UIST '94: Proceedings of the 7th annual ACM symposium on User interface software and technology*, pages 119–120, New York, NY, USA, 1994. ACM Press.

Joan B. Garfield. Beyond testing and grading: Using assessment to improve student learning. *Journal of Statistics Education*, 2(1), 1994.

Gradelink. Gradelink. www.gradelink.com, July 2007.

Mark Harrower and Cynthia A. Brewer. Colorbrewer.org: An online tool for selecting colour schemes for maps. In *The Cartographic Journal*, volume 40, pages 27–37, June 2003. www.colorbrewer.org.

iParadigms LLC. Turnitin: Digital assessment suite. www.turnitin.com, July 2007.

David Jackson and Michelle Usher. Grading student programs using assyst. In *Proc. SIGCSE Technical Symposium on Computer Science Education*, volume 28, pages 335–339, 1997.

Karel Kreijns, Paul A. Kirschner, and Wim Jochems. Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: a review of the research. In *Computers in Human Behavior*, volume 19, pages 335–353, 2003.

Steiner Kvale. *Interviews: An Introduction to Qualitative Research Interviewing*, pages 187–209. Sage Publications, Inc, 1996.

Qing Li and Chris North. Empirical comparison of dynamic query sliders and brushing histograms. In *Proc. IEEE Symposium on Information Visualization 2003*, pages 147–154, October 2003.

James McCaffrey. Test run: The analytic hierarchy process. 20(6):139–144, June 2005.

McGraw-Hill Higher Education. Mathzone. www.mhhe.com/math/mathzone, July 2007.

Microsoft. Excel. office.microsoft.com/excel, July 2007.

Geoffrey A. Moore. *Crossing the Chasm: Marketing and Selling High-Tech Products to Mainstream Customers.* Harper Business, 1999.

Peter Naur. Automatic grading of students' algol programming. In *BIT*, volume 4, pages 177–188, 1964.

OnCourse Systems for Education. Oncourse: Systems for education. www.oncoursesystems.com, July 2007.

Ellis Batten Page. Computer grading of student prose, using modern concepts

and software. In *Journal of Experimental Education*, volume 62, pages 127–142, 1994.

Pearson Education. Mygradebook. www.mygradebook.com, July 2007.

William James Popham. *Classroom Assessment: What Teachers Need to Know (Second Edition)*. Prentice Hall, 1998.

Herbert J. Rubin and Irene S. Rubin. *Qualitative Interviewing: The Art of Hearing Data*. Sage Publications, Inc., second edition. edition, 2004.

Jinwook Seo and Ben Shneiderman. A rank-by-feature framework for interactive exploration of multidimensional data. *Information Visualization*, 4(2): 99–113, June 2005.

Ben Shneiderman and Catherine Plaisant. *Designing the User Interface: Strategies for Effective Human-Computer Interaction (4th Edition)*. Addison Wesley, 2004.

Jaime Spacco, David Hovemeyer, William Pugh, Jeff Hollingsworth, Nelson Padua-Perez, and Fawzi Emad. Experiences with marmoset: Designing and using an advanced submission and testing system for programming courses. In *ITiCSE '06: Proceedings of the 11th annual conference on Innovation and technology in computer science education*. ACM Press, 2006. ISBN 1-59593-055-8.

Richard J. Stiggins. Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Isues and Practice*, 18(1): 23–27, 1999.

James J. Thomas and Kristin A. Cook, editors. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. US DOH National Visualization and Analytics Center, 2005.

Barbara E. Walvoord and Virginia Johnson Anderson. *Effective Grading: a tool for learning and assessment*. Jossey-Bass Publishers, 1998.

Herman G. Weller. Assessing the impact of computer-based learning in science. In *Journal of Research on Computing in Education*, volume 28, pages 461–485, 1996.