The Assessment: Testing the Role of Gene Mutations in Pancreatic Cancer

The assessment asked students to test a hypothesis about the role mutations in key genes play in the severity of pancreatic cancer. Students were presented with data on eight variables gathered from studies on this hypothesis including things such as patient ID, sex, cancer metastasis, and the presence of a mutation in the p53 or KRAS gene. Students were then asked to plot three predictions drawn from the hypothesis by graphing the appropriate variables for each prediction. Each graph represents a different task in the assessment. The data in this assessment is drawn from the graphs they submitted for each task, and questions they answered following each graph.

Graphing Skills scored by this version of GraphSmarts

This version is able to score nine graphing practices (out of 17 that will eventually be targeted), described in the following table.

Graphing Skill	**Description**
Variable Relevance	Selects variables for the graph that are relevant to a scientific claim in the context of a given research question, hypothesis, prediction, or objective
Variable Categorization	Identifies variables as related or causally linked in the context of a stated research question, hypothesis, prediction, or objective
Graph Type	Selects a graph appropriate for the type of data
Data Summarization	Plots individual or summarized data to communicate information efficiently for a given data set and intended purpose
Data Variability	Displays variation in data in a form appropriate for a given graph type and intended purpose
Axis Conventions	Follows disciplinary conventions in assignment of axes of graph
Scaling	Follows disciplinary conventions in scaling
Data Description	Describes the characteristics (i.e. central tendency, variability) and patterns of the graphed data for the plotted values and graph type
Scientific Claim	Interprets the constructed graph to support a scientific claim in the context of a given research question, hypothesis, prediction, or objective

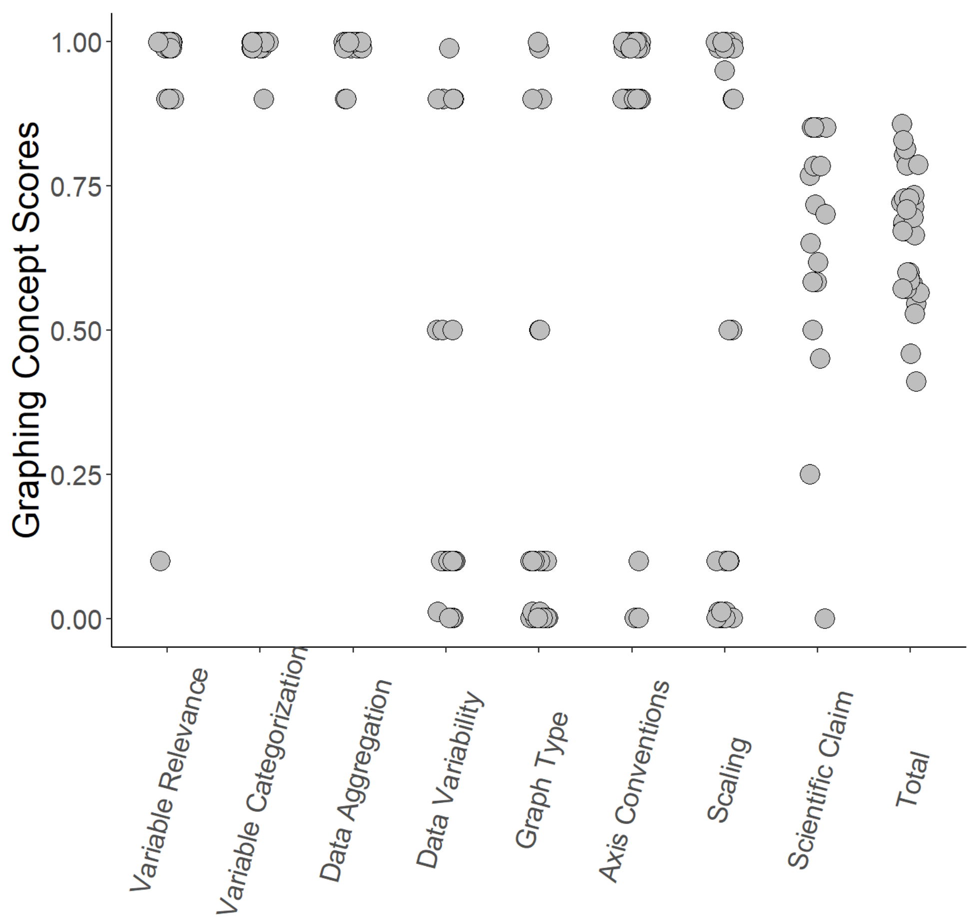
The overall score for each student is the average across all nine graphing practices.
Each practice is scored on a scale from 0 - 1. These scores represent the estimated probability that students are competent with that practice. For most practices, students who scored 0.9 or above did well on at least 2 of the 3 graphs they were asked to make, while students who scored 0.1 or below did poorly on at least 2 of the 3 graphs. On individual practices, those with

scores around 0.5 often were missing one or more of the graphs, and the score reflects one graph showing high performance on that skill and another showing low performance.

Scores for the class across all three tasks in GraphSmarts Assessment
Total Score: 0.66 out of 1.0

Sample Size
A total of 34 students in your class submitted at least one graph, and there was a total of 96 graphs submitted.

While there were a wide range of graphing skills in this class, most students had mid-range to good graphing skills. One of your students made what our algorithm considered a perfect graph on all three tasks. Most of your students showed some good practices in each graph they made. Students did well on identifying variables that are linked to a specific prediction or hypothesis, choosing appropriate variables to graph, data aggregation and following standard axis conventions. They struggled with showing variability, choosing the right type of graph for their data and scaling axes.



GraphSmarts Performance Scores by Practice

Details on Sub-Skills

Variable Relevance

Performance score: 0.95 out of 1.0
Question score: 0.98 out of 1.0

Students received high performance scores if they plotted the most relevant variables for the prediction being tested with each graph, and a low score if one of the variables relevant to the prediction was not plotted. There was also a card-sorting question prior to each graph asking which variables were relevant to the prediction being tested. Students similarly received high scores if all relevant variables were included in that bin on the question, and a low score otherwise.

Overall, your students did well at plotting variables related to the prediction in all three tasks. Only 4 students submitted a graph where there was one correct variable missing, and only one student submitted a graph with two variables that were not relevant.

You students did well on the card sorting task, picking the relevant variables the majority of the time for all three tasks.

Variable Categorization

Score: 0.99 out of 1.0

Students received high scores if they plotted two variables against each other where they could plausibly be related to one another (through causation or correlation) given the presenting problem. For example, plotting any type of mutation (i.e., p53, KRAS or additional mutations) against cancer stage or survival time would receive a high score. Plotting sex as an independent variable against other variables also resulted in a high score. Students only received low scores if they plotted two variables that are independent of each other (i.e., Patient ID vs. sex).

In this data set, this was a relatively easy graphing skill to demonstrate (as most variable combinations had a plausible relationship between them), and most of your students received perfect scores.

Data Aggregation

Score: 0.99 out of 1.0

Students received high scores if they did not use data aggregation features inappropriately. Possible data aggregation features were plotting means, std. Dev, or range, plotting group mean lines, and/or plotting a best-fit line. Students who did not use any of these automatically scored high. Students who used one of these features appropriately also scored high. Students who used a feature inappropriately scored low.

Your students scored high in this category because there were no students who used data aggregation inappropriately. Your class overwhelmingly chose to plot raw data, with only 5 students graphing aggregated data. None of the graphs with raw data had mean lines, which would have been another way of showing aggregation. Your students fit data to a regression line one time, which was done appropriately.

Data Variability
Score: 0.27 out of 1.0

Students received high scores if their plot showed the variance in their data. Scatter plots with individual patient data (i.e. not plotting aggregated data through taking means) on each axis automatically show variance by showing all data points. If students plotted aggregated data on one of the axes, they needed to add std. dev. error bars for full credit. Students received low scores if they plotted bar graphs with individual patient data, as they show variability in a difficult to read way. Line graphs with a mix of quantitative and categorical data similarly receive low scores.

Students performed poorly in this competence since the majority of students submitted a bar graph that plotted individual patient data without any type of aggregation. This results in a bar graph that has a bar for each individual data point, which shows variability in a way that made it difficult to interpret. In addition, 22% of students plotted a line graph with categorical and raw data, which resulted in graphs that are hard to interpret and were not appropriate for this dataset.

Graph Type
Score: 0.17 out of 1.0

Students received high scores if the type of graph was well-suited to the type of data they plotted. For instance, if they plotted individual patient data on scatter plots, or means using a bar graph. Plotting means on one axis versus a categorical variable on a scatter plot was also scored highly (for instance, p53 mutation vs survival time). Line graphs were scored low in this dataset, as there were no variables that could be continuously interpolated between measured points.

Your students struggled with choosing a graph type appropriate for the data they plotted. 57% of students choose to plot a line graph for at least one of the tasks. Although a variable like survival time seems like a variable that can be interpolated, survival is not measured for patients over time (it is represented as number of months) and cannot be interpolated over time. In addition, four students also chose a bar graph to plot two quantitative variables (survival time vs additional mutations), which results in a graph that is difficult to interpret.

Axis Conventions
Score: 0.88 out of 1.0

Students received high scores if the independent variable in a given comparison was on the x-axis. Sex and Patient ID were independent compared to any other variable. Other variables were scored as being placed on the correct axis if the y-axis variable could be reasonably considered dependent on the x-axis variable, or if it was reasonable to correlate the two variables.

Students did fairly well following axis conventions. Few students plotted the independent variable on the y axis instead of the x axis. For example, three students put survival time on the x axis, although mutations impact survival time and thus should be on the y axis.

Scaling
Score: 0.38 out of 1.0

Students received high scores if they chose scales for quantitative axes that showed all data points for the variable being plotted and were not way beyond any of the data points. Qualitative axes did not affect the score.

Your students scored less well on this practice primarily because we did not give credit if the maximum value on an axis was exactly equal to the maximum value in the data. In other words, if Survival Times in the data were between 7-107, we only gave a high score if the scale on the axis went beyond 107. This makes it easier to see that data is not being cut off but is a debatable criterion. The majority of your students included a range that included all the data and did not cut off any data points. However, students often created scales by reading the range of data given by the data table and setting those values as the exact range of the axis scales. There were four students who did not plot the appropriate range and cut off data points.

Data Description
Score: 0.25 out of 1.0

After each graph made students received a question asking them to describe the data trend they saw in the graph. Students scored highly if the variables they used in their data description matched those they plotted, and if they correctly described the trend visible in the graph (i.e. X higher in Y; X had higher values with lower values of Y).

Your students struggled with describing data trends. Around 10% of students described variables that didn't match the variables they plotted on their submitted graphs. In addition, about half of students had a difficult time describing trends that matched the data plotted on their graphs, often describing a trend that was not present in the data.

Scientific Claim
Score: 0.63 out of 1.0

Scores were drawn from analyzing the conclusion statements students made, both for the presence of a conclusion (i.e. “data supports hypothesis”), that the conclusion statement aligned with what they plotted in their graph, describing broad trends in the data, and how well the student used the graphed data to support their conclusion.

Five of your students stated in their conclusions whether their data supported or refuted the prediction they were testing, and most supported that statement by citing specific data or trends in the graph. However, few students made broader statement about the trends in the data. In addition, three students had at least one answer that made claims or cited evidence that wasn’t aligned with what the graph showed.