Teacher guidance

Key focus

This objective builds on the TARSOG framework (Trend, Association, Relationship, Scatter, Outliers, Grouping) by requiring students to read and quantify values directly off a scatter plot, not just label its features verbally. Every task in this LO shows a real plot: students must name the explanatory (independent) and response (dependent) variable, use the line of best fit to predict a value, classify that prediction as interpolation or extrapolation, and quantify how far an outlier sits from the trend. The beta textbook reference is Ex 31.01, pg 579, alongside the CorbettMaths Scatter Graphs worksheet.

Suggested teaching sequence

1. Recap TARSOG briefly, then move straight to a plot: model naming which axis is the explanatory (independent) variable and which is the response (dependent) variable, and why (the explanatory variable is the one that plausibly influences the other).
2. Model describing trend and association together with the actual data (e.g. "practice hours and quiz score show a strong, positive, linear association").
3. Introduce reading a prediction off the line of best fit at a value between two plotted points, and label this *interpolation*.
4. Introduce predicting beyond the plotted range and label this *extrapolation*; discuss why it is less trustworthy (the pattern is not guaranteed to continue, and results can even become nonsensical, e.g. a negative price).
5. Show a plot with one outlier. Have students read its coordinates, calculate the trend's prediction at that x-value, and quantify the gap between the two.
6. Show a keyed two-group scatter plot (e.g. two classes). Have students compare each group's relationship strength and explain why one combined line of best fit across both groups would be misleading.
7. Practise judging relationship strength (weak/moderate/strong) purely from how tightly points cluster around the line, separately from the outlier question.
8. Have students critique a flawed real-world claim about a line of best fit (e.g. "every extra dollar spent always adds exactly this much in sales"), identifying both the outlier and the interpolation/extrapolation issue.

Core examples to model

- Explanatory vs. response: practice hours (explanatory) vs. quiz score (response) – practice plausibly affects the score, not the reverse.
- Interpolation vs. extrapolation: predicting a value at $x = 6$ within a plotted range of 1–9 is interpolation; predicting at $x = 15$ is extrapolation, and can even give an impossible negative car value.
- Outlier, quantified: a viral social-media post gives \$3,000 of ad spend an \$85,000 sales result, against a trend-predicted \$30,000 – a \$55,000 gap.

- Grouping: two classes' sleep-vs-score data sit in visibly different clusters on a keyed plot, so one combined line of best fit would blend two different real patterns into a misleading average.

Watch for

- Naming a TARSOG letter without reading actual numbers: this LO always expects a predicted value, an outlier's quantified gap, or a percentage, not just the label.
- Confusing positive/negative association (scatter plots) with increasing/decreasing trend (time series) – the vocabulary is deliberately different between the two topics.
- Trusting an extrapolated prediction as much as an interpolated one: always state which type a prediction is, and flag extrapolation as less reliable.
- Swapping which variable is explanatory and which is response: the explanatory variable is the one that plausibly drives the other, not simply whichever is plotted on the x-axis by habit.
- Drawing (or accepting) one combined line of best fit across two visibly different groups shown by a key.

Checks for understanding

- Can students correctly name the explanatory and response variable for an unfamiliar context?
- Do they describe trend and association together, backed by the actual data?
- Can they judge relationship strength from how closely points cluster around the line of best fit?
- Can they locate an outlier and quantify how far it departs from the line's prediction?
- Can they use the line of best fit to predict a value, and correctly classify it as interpolation or extrapolation?
- Can they explain why an extrapolated prediction is less reliable than an interpolated one?
- Can they explain why a single combined line of best fit across two distinct groups would be misleading?