Relational Inclusion – Think Piece - Zipf's Law

What is Zipf's Law?

George Zipf was a linguist who studied word frequency. He noticed a pattern: a small number of words dominate usage, while the vast majority occur rarely. When words are ranked by frequency, the pattern is strikingly consistent.

At its simplest, Zipf's Law says this: in many systems, frequency is inversely related to rank. The most common element appears roughly twice as often as the second most common, three times as often as the third, and so on. Not perfectly, but uncannily close.

What's it got to do with school?

For many years now I've said that around 5% of children seem to cause 95% of the problems. Zipf's Law offers a way of understanding why. When schools rank behaviour incidents by frequency, they rarely see an even spread. **Instead, a very small number of pupils account for a very large proportion of incidents**, followed by a long tail of pupils with one-off or rare incidents.

What we do not have are lots of pupils behaving a bit badly. What we have are a few pupils responsible for most significant incidents, and the majority responsible for little to none.

In any school dataset, you will often see something like this:

- The top one to three percent of pupils account for around twenty to thirty percent of incidents.
- The top ten percent of pupils account for fifty to seventy percent.
- The bottom fifty to seventy percent of pupils record one incident or none at all across the entire year.

This pattern remains consistent across schools and across years.

Why does this matter?

Most behaviour systems are built as if incidents are normally distributed. The same rules, sanctions, escalation steps, and expectations are applied to everyone. But Zipf tells us something uncomfortable: **behaviour data is heavy-tailed, not average-based**. Designing systems for the ‘typical’ pupil misses the point entirely. Removing repeat offenders barely changes the shape of the curve. The system keeps blaming children for what is actually structural maths.

If a Child Could Do Better, They Would

Two phrases are heard all too often: ‘*it is chosen behaviour*’ and ‘*they are not school ready*’. Zipf does not deny agency, but it screams constraint.

The pupils at the top of the ranking are often under chronic stress. They have fewer behavioural options, fewer relational buffers, and narrower regulatory bandwidth. **Their probability of repeated incidents is therefore higher**. This is not because they are worse, but because their degrees of freedom are smaller.

Why Sanctions Don’t Work

Sanctions assume that behaviour will spread back toward a middle. Zipf systems do not have a middle. Sanctions concentrate on the same pupils. Relationships erode. Incidents cluster. The curve steepens.

What looks like defiance is often predictable compression under pressure:

- A handful of pupils “become the problem”
- The same specific behaviours dominate logs

- The same names keep reappearing
- Exclusions feel inevitable rather than chosen

The Relational Insight

If Zipf is operating, behaviour is unevenly distributed because regulation is unevenly distributed. Effective responses widen relational safety, increase adult regulatory presence, reduce cognitive and emotional load, and personalise support disproportionately. Zipf systems shift when capacity increases, not when punishment intensifies. Sanctions reduce a child's relational safety, whereas increased adult presence and co-regulation widen the range of possible responses.

The Uncomfortable Conclusion

Behaviour data does not tell you who is naughty. It tells you where the system is leaking regulation. The pupils at the top of the chart are not anomalies. They are signals.

Why Exclusion Doesn’t Work

When a child is permanently excluded, incident numbers may dip briefly. Very quickly, another pupil slides into the vacated position and the curve reforms. This is often misread as evidence that exclusions are necessary. In reality, it is a feature of the system.

Why Some Schools Appear to Have More Behaviour Problems

Zipf also explains the myth that some schools simply have ‘worse behaviour’. Schools under greater community stress will show steeper curves, not because more children misbehave, but because more children are operating with reduced regulatory capacity.

For most pupils, incidents are rare or never happen at all. The difference sits at the top of the distribution. Apply more pressure, and the curve steepens further. What looks like a behaviour problem is often a capacity problem being misnamed.

Looking in the Wrong Place

Most behaviour analysis is built on averages. If we know that in reality a few pupils are carrying most of the load, then instead of asking how many instances of poor (or dare I say bad) behaviour a school has it might be better to ask: where behaviour is concentrated; who is carrying the strain; where is regulation thinning; and what happens to the curve when adult presence increases rather than sanctions.

What if behaviour dashboards were designed to reveal pressure points, not smooth them away.

Might it be, that in terms of behaviour, we have simply been looking in the wrong place. Or perhaps more accurately, we have been looking for the wrong thing.

As John Berger once wrote, "What we see depends mainly on what we look for." If we look for defiance, we will find defiance. If we look for chosen behaviour, we will see children making poor choices. And if we look for who is or is not school ready, the data will quickly point us towards those who appear not to be.

Read through Zipf, behaviour data suggests not a problem child, but a pressure point. Not moral failure, but thinning regulation.

The question is not how to stop pupils appearing in the data, but what it would mean to build systems that no longer require them to carry it for us.

What if, instead of building a system which counted used behaviour points to measure a child's failings, we used data to help us actually see the child and help us understand where the system was failing. That would be a bit different.