WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Seeing Where Learning Breaks Down: Design and Simulation-Based Demonstration of a Python and Power BI Dashboard for Topic-Level Mastery in High-School Calculus and Trigonometry

Jannathul Firdous[1], and Shajahan Abdul Salam[2*]

[1]Department of Mathematics, Advance International Academy, Al Hamra District, Dammam, Saudi Arabia
[2]Independent Researcher, Dammam, Saudi Arabia. ORCID: 0009-0008-1568-0041

A B S T R A C T :

Secondary mathematics teachers record marks question by question, but gradebooks usually keep one total per test. A total hides the topic a student has not yet mastered. This paper presents the design of a low-cost dashboard that turns item-level marks into topic-level mastery views for upper-secondary calculus and trigonometry, and demonstrates it on a simulated dataset. Marks entered in a spreadsheet are validated, pseudonymised and aggregated by a short Python (pandas) script. They are shown in Microsoft Power BI as a class heat map, a re-teach list, a student profile and a topic trend view, each linked to one classroom decision. To test the tool without using student data, we simulated two Grade 11 sections (60 students, 32 questions, four assessments, six topics) with a logistic model of performance. In it, 55 topic-specific learning gaps were deliberately planted, including a prerequisite link between limits and differentiation. The script processed 1,917 item scores in 0.13 seconds and caught all three deliberately injected data-entry errors. The topic view flagged 74.5% of the planted gaps. A total-only view, which flags every topic of a student below 60% overall, flagged 41.8%. Precision rose from 27.4% to 43.6%. Across 100 repeated simulations, mean sensitivity was 79.3% for the topic view and 42.7% for the total-only view. The advantage held at every cut-off from 45% to 75% and at every gap size tested. Fifteen of the 33 students with an overall total of 70% or more had at least one topic in the re-teach band, most often trigonometric identities. The planted link between limits and differentiation was visible in the dashboard. The design borrows the idea of leading indicators from safety-critical industries. The paper discusses privacy safeguards under Saudi Arabia's Personal Data Protection Law, the limits of simulated evidence, and a planned classroom evaluation.

Keywords: learning analytics, teacher dashboard, formative assessment, simulation, calculus, trigonometry, Power BI, Python, secondary mathematics, Saudi Arabia

1. Introduction

A test total of 62% tells a mathematics teacher that a student is struggling. It does not tell her whether the student cannot apply the chain rule, cannot read the unit circle, or lost marks only on one long word problem. In upper-secondary calculus and trigonometry, topics build on each other closely. A gap in limits weakens the derivative from first principles, and a gap in identities weakens equation solving. So the useful unit of information is the topic, not the test.

Most schools already hold the data needed to see topics. Teachers mark item by item, and each item maps to a syllabus objective. The data is usually reduced to one total in a gradebook, because rebuilding the topic view by hand takes too long. Commercial learning-analytics platforms can produce topic views. But they are often tied to a particular assessment product, cost money, or store student data outside the school's control.

This paper reports a low-cost alternative that a teacher with basic data skills can build and maintain. It uses a spreadsheet the teacher already keeps, a short Python script, and Microsoft Power BI, which many schools already license. It was designed by a practising teacher of Grades 10–12 pure mathematics at a private international school in Dammam, Saudi Arabia. Before using it with real students, we tested it on a simulated class in which the true learning gaps are known.

The design also borrows an idea from safety-critical industries. In construction and process plants, routine field observations are turned into *leading indicators*. These indicators show where controls are weakening early enough to act before an incident (Abdul Salam, 2026a; Abdul Salam, 2024). A topic-level mastery view plays the same role in a classroom. It flags the weakening topic while there is still time to re-teach it, rather than after the end-of-term exam.

A simulation has one advantage over a first classroom trial: we know where every gap really is, so we can check whether the dashboard finds it. The study addresses three research questions:

1. **RQ1 (feasibility):** Does the pipeline process a realistic term of item-level marks quickly and catch data-entry errors?
2. **RQ2 (detection):** How well does the topic view find planted learning gaps, compared with a total-only view?
3. **RQ3 (patterns):** Does the dashboard reveal patterns a teacher would act on, such as hidden gaps in high-scoring students and prerequisite links between topics?

The contribution is practical. The paper gives a reproducible data model, a pipeline that pseudonymises data before it reaches the dashboard, and a simulation test with known ground truth. A classroom evaluation of teacher use and learning gains is planned as a follow-up study.

2. Background and related work

2.1 Formative assessment and mastery

Formative assessment raises achievement when evidence about learning is used to adjust teaching (Black & Wiliam, 1998; Wiliam, 2011). Feedback works best when it tells the learner and the teacher where the gap is and what to do next (Hattie & Timperley, 2007). Bloom's mastery-learning model makes the same point at the level of the unit. Teachers give a formative check, then corrective teaching for those who have not reached mastery, then a second check (Bloom, 1968; Guskey, 2007). The model is well supported, but its correctives step needs exactly the topic-level information that a test total hides.

2.2 Why calculus and trigonometry need topic-level views

Research on students' understanding of limits shows that their working concept image can differ from the formal definition. This mismatch causes errors that look random at the level of a total score (Tall & Vinner, 1981). In trigonometry, a study of early university students found that many treat sine and cosine as procedures on triangles rather than as functions on the unit circle. This limits their work with identities, equations and graphs (Weber, 2005). Both findings suggest that errors cluster by concept, so a view organised by concept should point the teacher to the right correction.

2.3 Learning-analytics dashboards

Learning analytics is the measurement, collection, analysis and reporting of data about learners to understand and improve learning (Long & Siemens, 2011). Dashboards are its most common teacher-facing form (Verbert et al., 2013). Systematic reviews find that most dashboards are built for universities and online courses, and few are tested in regular school classrooms (Schwendimann et al., 2017; Bodily & Verbert, 2017). Teachers turn dashboard information into action only when the display links directly to a decision they can take in class (Molenaar & Knoop-van Campen, 2019). A wider review of data-driven decision making in schools finds that teachers' data skills and the time cost of data handling are the main barriers (Datnow & Hubbard, 2016; Mandinach & Gummer, 2016). The design in this paper responds to both barriers. It keeps the data in tools the teacher already uses, and each dashboard page answers one classroom decision.

2.4 Lessons from safety-critical industries

High-hazard industries have long experience of turning routine observations into early-warning displays. A risk-based resilience framework for brownfield construction organises safety around six dimensions, four of which form a loop: anticipate, monitor, respond and learn (Abdul Salam, 2026a). Monitoring there works only when it is tied to a defined response. That is the same condition Molenaar and Knoop-van Campen (2019) found for teacher dashboards. In permit-to-work coordination, heat maps of where hazards interact help supervisors decide where to look first (Abdul Salam, 2024). The class heat map in Section 3 uses the same layout. Work on worker exposure profiles argues that a person's record must stay continuous across employers and projects, so that patterns are not lost at each handover (Abdul Salam, 2025). The same applies to students moving from Grade 10 to Grade 12. A topic record that follows the student lets the next teacher see which foundations are still weak. Finally, digital and predictive monitoring is a stated priority of Saudi Vision 2030 programmes, in workplaces and in education (Abdul Salam, 2026b; Kingdom of Saudi Arabia, 2016).

2.5 Gap

Few studies describe a teacher-built, low-cost dashboard for upper-secondary calculus and trigonometry, and fewer still come from Gulf schools. Fewer still test whether such a dashboard finds the gaps it claims to find. This paper addresses that gap with a simulation in which the true gaps are known, before any classroom use.

3. System design

The system has four stages: capture, process, model and act. The teacher enters item marks in a spreadsheet, and a Python script cleans and pseudonymises them. Power BI turns them into four views, and each view ends in a classroom action. Figure 1 shows the flow.

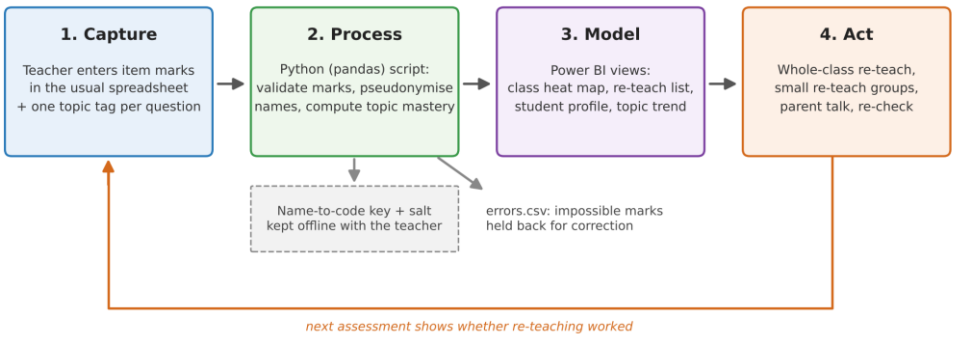


Fig. 1 - Data pipeline, from marks to classroom action

3.1 Design principles

1. **No new data entry.** The teacher records item marks, which she already does when marking. Only the topic tag per question is added, once per assessment.
2. **One page, one decision.** Each dashboard page answers one question the teacher acts on: whom to re-teach, what to re-teach, and what to tell parents.
3. **Privacy by design.** Names never reach the dashboard. Students appear only as pseudonymous codes, and the key that links codes to names stays offline with the teacher.
4. **Transparent thresholds.** Mastery bands are fixed, simple and stated on every page, so students and parents can understand them.

3.2 Data model

Each assessment question is tagged with one primary topic code from the syllabus map (Appendix B). Table 1 lists the four tables in the model, which are arranged as a star schema.

Table 1 - Data model

Table	One row per	Key fields
ItemScore (fact)	student × question	student code, question ID, marks awarded, marks available
Student	student	student code (pseudonymous), section, grade
Question	question	question ID, assessment ID, topic code, marks available
Topic	syllabus topic	topic code, topic name, strand (calculus/trigonometry), grade

Topic mastery for student *s* on topic *t* is the share of available marks earned across all questions tagged to that topic in the period:

$$M_{s,t} = \frac{\sum_{q \in t} \text{awarded}_{s,q}}{\sum_{q \in t} \text{available}_{q}} \times 100$$

Mastery is reported in three bands: **Secure** (80% or more), **Developing** (60–79%) and **Re-teach** (below 60%). The 80% cut-off follows common practice in mastery learning (Guskey, 2007). A topic needs at least 3 questions or 6 marks before a band is shown, so that one question does not decide a student's band.

3.3 Python processing script

The script (outlined in Appendix A) uses the pandas library. On the simulated term (1,920 item scores) it ran in 0.13 seconds on a laptop. It does five things:

1. Reads the mark sheet for each section and checks that the marks awarded never exceed the marks available.
2. Replaces each student name with a code made from a salted SHA-256 hash. The salt and the name-to-code key are kept in a separate file that never leaves the teacher's computer.
3. Joins the question-to-topic map and calculates M for each student and topic, then the class mean for each topic.
4. Flags each student-topic pair that is below 60%, and has enough evidence. Rows with impossible marks are held back in an errors file for the teacher to correct.
5. Writes clean CSV files that Power BI reads with one refresh.

3.4 Power BI dashboard

The dashboard has four pages.

Table 2 - Dashboard pages and the decision each one supports

Page	What it shows	Decision it supports
Class heat map	Students as rows, topics as columns, cells coloured by mastery band	Which topics need whole-class re-teaching
Re-teach list	Students grouped by the topic they need, with their mastery %	Who joins which small re-teach group this week
Student profile	One student's topic bands over time	What to say in a parent meeting or a one-to-one talk
Topic trend	Class mean per topic across assessments, by section	Whether re-teaching worked, and whether one section lags

The class heat map uses the same layout as hazard-interaction heat maps in permit-to-work systems: rows and columns that the reader already knows, with colour used only for the band (Abdul Salam, 2024). A colour-blind-safe palette is used, and each band also has a text label.

4. Simulation study method

4.1 Design

We used a simulation study with planted ground truth (Morris et al., 2019). We generated a class in which we decided in advance which students had a gap on which topic. Then we ran the real pipeline on the simulated marks and checked how many of those gaps the dashboard flagged. This tests whether the tool works. It does not test whether teachers use it or whether students learn more, which needs a classroom study (Section 7).

4.2 Simulated class and assessments

The simulated class has two Grade 11 sections (11A and 11B) of 30 students each, all with fictional names. It covers six topics (Appendix B): four in trigonometry (T1–T4) and two in calculus (C1 limits, C2 introduction to differentiation). There are four assessments with 32 questions and 134 marks in total, giving 1,920 item scores.

Table 3 - Simulated assessments

Assessment	Topics	Questions	Marks
Quiz 1	T1, T4	6	20
Unit test 1	T1–T4	9	40
Quiz 2	C1	5	20
Mid-term	all six	12	54

4.3 Performance model

Each mark on a question was earned with a probability given by a logistic model, a simple form of the Rasch model (Rasch, 1960):

$$P(\text{mark}) = \frac{1}{1 + e^{-(\theta_s + \delta_{s,t} - b_t + \varepsilon_q)}}$$

Here θ is the student's general ability, drawn from a normal distribution with mean 1.2 and SD 0.8. δ is the student's skill on topic t, with SD 0.35 around zero. b is the topic's difficulty, and ε is a small random effect for each question (SD 0.3). The marks awarded on a question follow a binomial distribution over its available marks. Topic difficulties ranged from −0.6 (T1, the easiest) to 0.5 (T2, identities, the hardest).

Gaps were planted in three ways, each with a known location:

1. **Isolated gaps:** each student had a 12% chance of a gap on each topic, which lowered δ by 1.6.
2. **Prerequisite links:** a student with a T1 (unit circle) gap had a 60% chance of a further T3 (equations) gap, and a C1 (limits) gap gave a 60% chance of a C2 (differentiation) gap. These links reflect the research in Section 2.2.
3. **Section effect:** section 11B's skill on T2 was lowered by 0.5, as if identities had been taught in a shortened week. This was not counted as a planted gap.

The model planted 55 gaps across the 360 student-topic pairs. The random seed was fixed (2026), so the dataset can be reproduced exactly.

4.4 Data-entry errors

To test the validation step, we deliberately typed three impossible marks into the spreadsheet: 15 on a 5-mark question, 44 on a 4-mark question and 9 on a 6-mark question.

4.5 Measures and analysis

We compared two ways of deciding who needs re-teaching on a topic:

- **Topic view (the dashboard):** a student-topic pair is flagged when it is in the Re-teach band (below 60%).
- **Total-only view (a normal gradebook):** a teacher who sees only totals re-teaches students whose overall total is below 60%, so all six of their topics count as flagged.

For each view, **sensitivity** is the share of planted gaps that were flagged, and **precision** is the share of flags that were planted gaps. To check that one random class was not unusual, we repeated the whole simulation and pipeline with 100 new random seeds. For the patterns in RQ3, we counted hidden gaps, meaning students with an overall total of 70% or more but at least one Re-teach topic. We tested the prerequisite links with Fisher's exact test, and the T2 section difference with a Mann–Whitney U test.

4.6 Ethics and data protection

No human participants or real student data were involved, so ethics approval was not required. The pipeline is built for later use with real data under Saudi Arabia's Personal Data Protection Law (Royal Decree M/19, 2021). Names are pseudonymised with a salted hash before analysis. The name-to-code key stays offline, and the dashboard holds only codes.

5. Results

All results below come from the simulated dataset described in Section 4.

5.1 Feasibility (RQ1)

The script read all 1,920 item scores, pseudonymised the 60 names and wrote the dashboard files in 0.13 seconds. The validation step flagged exactly the three injected errors, with no false alarms. Those three rows were held back, and the other 1,917 were processed. All 360 student-topic pairs had enough evidence for a band (at least 3 questions).

Overall test totals had a mean of 71.4% (SD 14.2, range 41.0–94.8%), and 14 students were below 60%. Of the 360 student-topic pairs, 155 (43.1%) were Secure, 111 (30.8%) Developing and 94 (26.1%) Re-teach.

5.2 Detecting planted gaps (RQ2)

Table 4 - How well each view found the 55 planted gaps (seed 2026)

View	Pairs flagged	Planted gaps found	Sensitivity	Precision
Topic view (dashboard)	94	41	74.5%	43.6%
Total-only view (gradebook)	84	23	41.8%	27.4%

The topic view found 18 planted gaps that the total-only view missed. All 18 belonged to students whose overall total was 60% or more. Across the 100 repeated simulations, the topic view had higher sensitivity and higher precision than the total-only view in every run.

Table 5 - Results across 100 repeated simulations

View	Mean sensitivity (SD)	90% of runs between	Mean precision	90% of runs between
Topic view	79.3% (7.0)	66.7–89.7%	34.7%	26.0–45.7%
Total-only view	42.7% (11.9)	23.9–60.0%	20.6%	12.8–27.1%

Precision against planted gaps is a strict measure. In the main run, 53 topic flags were not planted gaps. Of these, 28 belonged to students whose overall total was below 60%, who would need support anyway. And 27 of the 53 scored between 50% and 59.9%, just under the line.

5.3 Patterns a teacher would act on (RQ3)

Hidden gaps. Fifteen of the 33 students with an overall total of 70% or more had at least one topic in the Re-teach band. Trigonometric identities (T2) was the most common, appearing for 9 of the 15. Figure 2 shows the heat map for section 11A, with rows sorted by overall total, and Figure 4 shows one example student. That student's overall total was 86%, but their T2 score was 52%. In a total-only gradebook, they would not be flagged at all.

Re-teach groups. The re-teach list turns the heat map into group sizes (Table 6). In 11B, 17 of 30 students needed T2, which points to whole-class re-teaching. In 11A, the needs were spread over T2, T3, C1 and C2, which points to small groups.

Table 6 - Students in the Re-teach band, by topic and section

Section	T1	T2	T3	T4	C1	C2
11A (n = 30)	2	11	12	8	9	12
11B (n = 30)	2	17	10	2	4	5

Prerequisite links. The planted link from limits to differentiation was visible. Of the 13 students in the Re-teach band on C1, 7 (53.8%) were also in it on C2, compared with 21.3% of the other students (Fisher's exact p = 0.035). The planted link from the unit circle to equations was not visible. T1 was the easiest topic, so only 4 students were in the Re-teach band on it, too few to show a link (p = 1.0).

Section differences. Figure 3 shows class means by topic. The planted T2 weakness in 11B appeared as a lower mean (60.6% against 67.2%) and more students to re-teach (17 against 11). However, it was not statistically significant with 30 students per section (Mann–Whitney p = 0.16). Section 11B was also 10–15 points higher on T4, C1 and C2, although no difference was planted there. Those gaps arose by chance in the random draw.

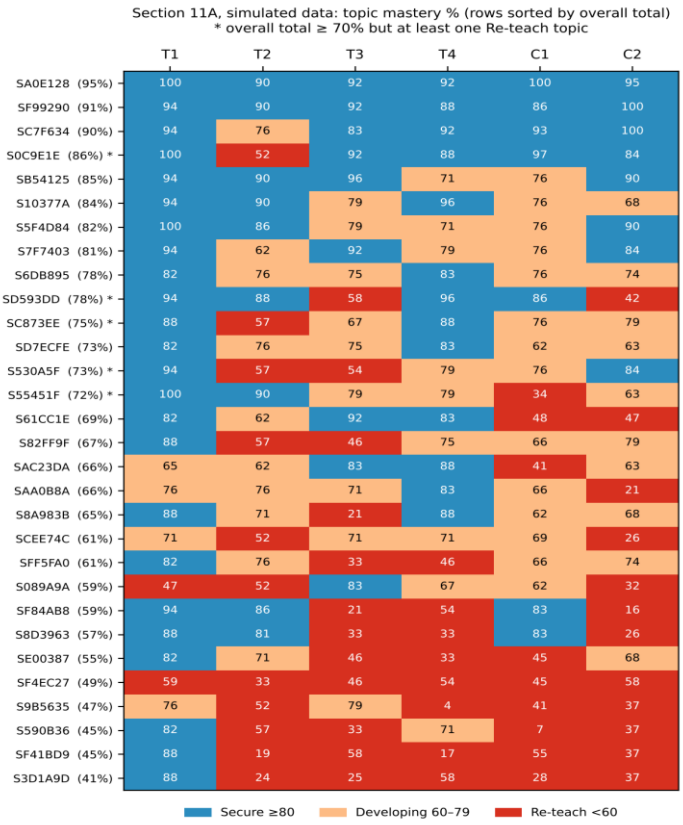


Fig. 2 - Class heat map, section 11A (simulated data). Rows are pseudonymous students sorted by overall total, and columns are topics. Stars mark students with a total of 70% or more and at least one Re-teach topic.

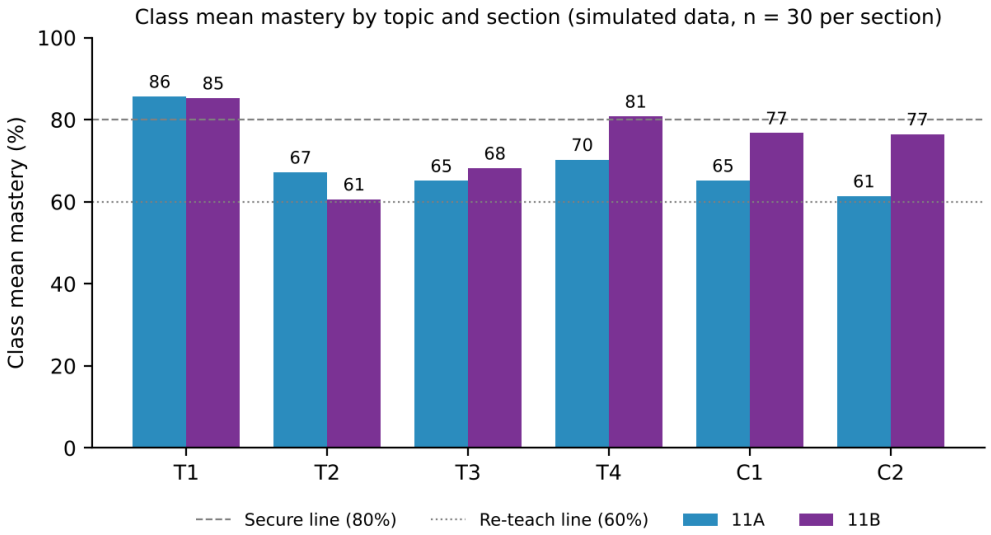


Fig. 3 - Topic trend view: class mean mastery by topic and section (simulated data).

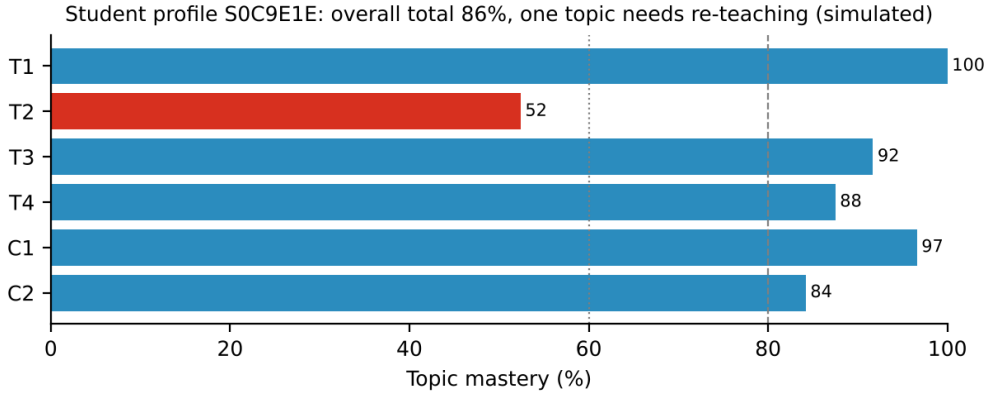


Fig. 4 - Student profile view for a student whose 86% overall total hides a Re-teach band on trigonometric identities (simulated data).

5.4 Comparison and sensitivity analyses

Six further analyses test how robust these findings are. All use the same simulation and pipeline.

Detection across 100 classes (Figure 5). In 90 of the 100 simulated classes, the topic view found more planted gaps than the total-only view managed in its single best class (69.2%). Its precision was also higher in every class.

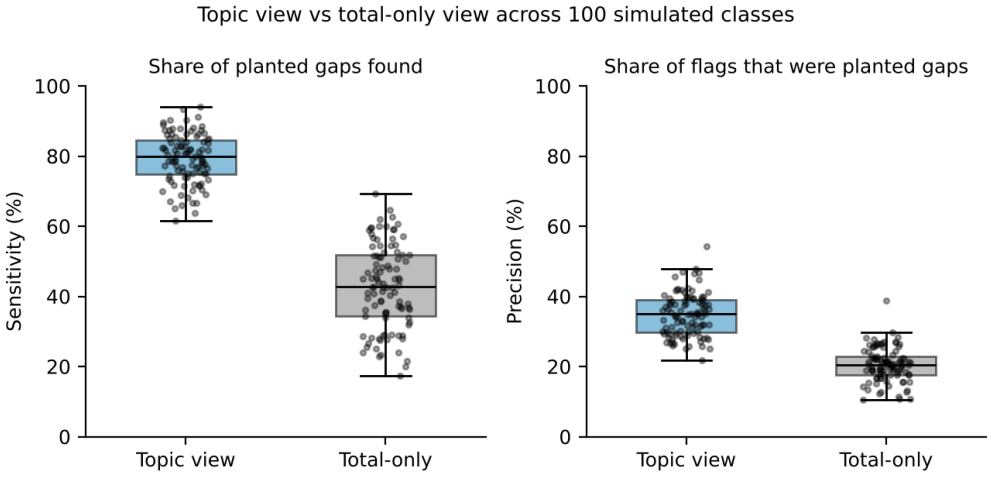


Fig. 5 - Sensitivity and precision of the topic view and the total-only view across 100 simulated classes. Each dot is one class.

A good total can hide a weak topic (Figure 6). Overall totals and weakest-topic scores were strongly correlated (r = 0.91). Even so, at any given total the weakest topic varied widely. The 15 hidden-gap students had totals of 70% or more, but their weakest topic scored only 33.3–58.8%.

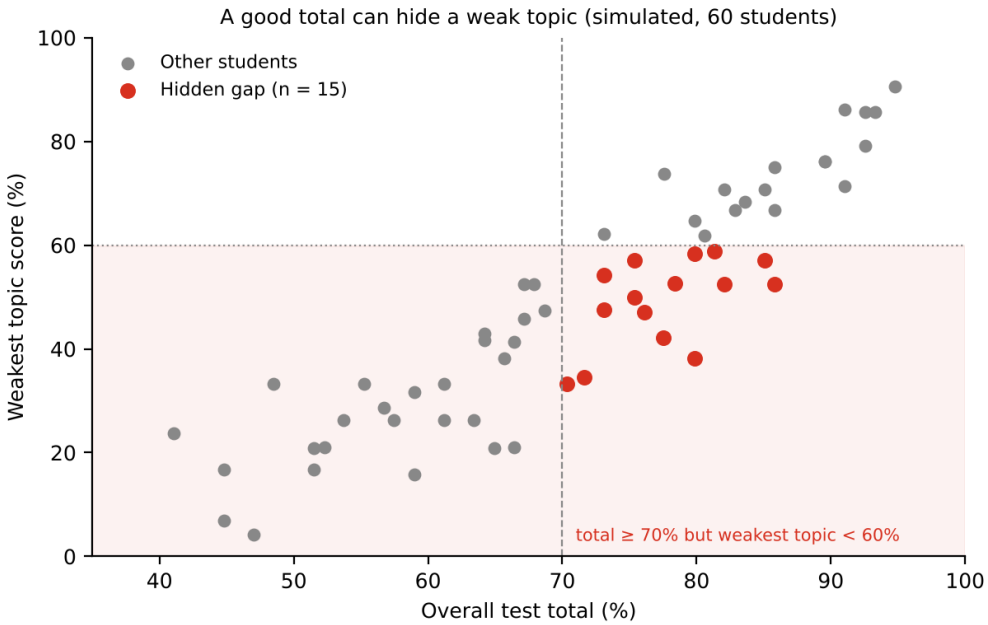


Fig. 6 - Overall test total against weakest topic score. Red points are students with a total of 70% or more and at least one topic below 60%.

Band profile by section (Figure 7). In 11B, 57% of students were in the Re-teach band on T2, against 37% in 11A. In 11A, the Re-teach share was 27–40% on five of the six topics. This makes the different responses clear: one whole-class lesson in 11B, and several small groups in 11A.

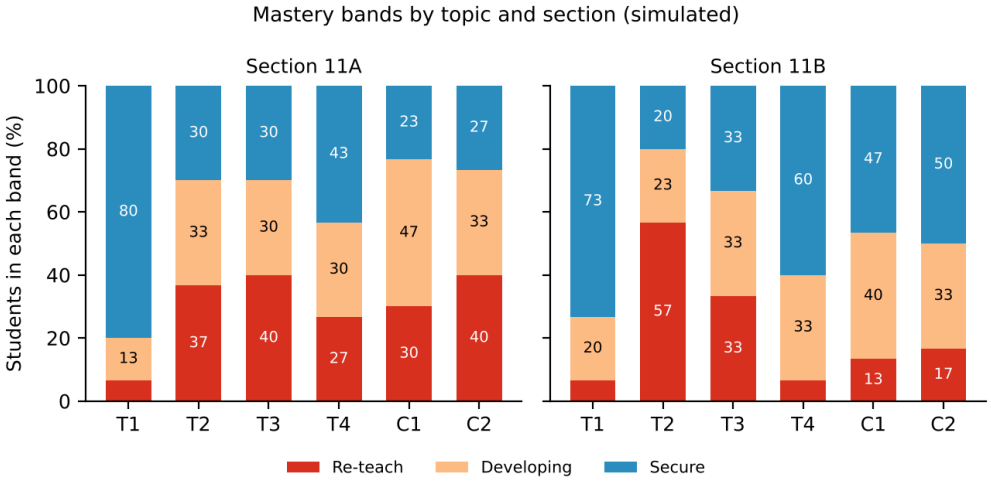


Fig. 7 - Share of students in each mastery band, by topic and section (simulated data).

Which weak topics occur together (Figure 8). Weak topics clustered, mainly because generally weaker students were weak on several topics. For example, 69% of the students in the Re-teach band on C1 were also in it on T2. The matrix helps a teacher see where re-teach groups overlap, so sessions can be combined. On its own, though, it cannot separate a prerequisite link from general weakness. That is why the formal comparison in Section 5.3 was needed. The T1 row rests on only 4 students.

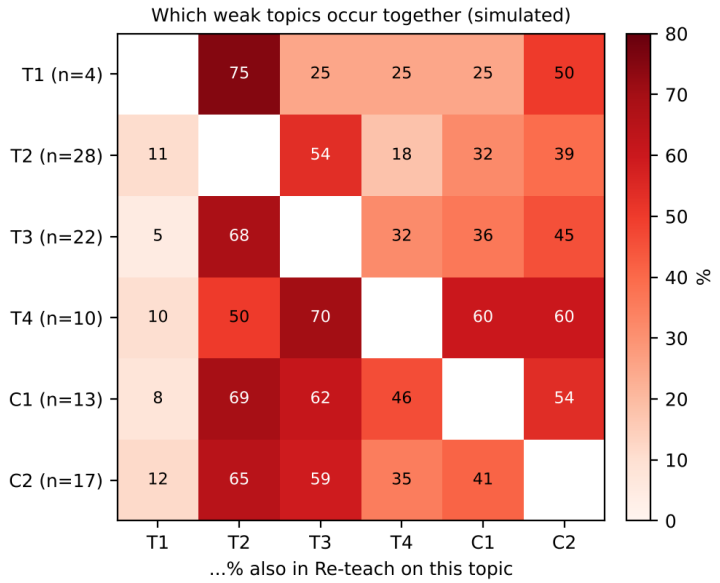


Fig. 8 - Of the students in the Re-teach band on the row topic, the percentage also in it on the column topic (simulated data).

Effect of the cut-off (Figure 9). At every cut-off from 45% to 75%, the topic view had both higher sensitivity and higher precision. At a cut-off of 65% or more, the total-only view flagged more pairs than the topic view but still found fewer planted gaps. At 65%, for example, it flagged 120 pairs and found 52.7% of the gaps, while the topic view flagged 114 and found 85.5%. Raising the topic-view cut-off from 60% to 65% added 20 flags and raised sensitivity from 74.5% to 85.5%. Precision fell only slightly, from 43.6% to 41.2%. A teacher with time for slightly larger groups could therefore use 65%.

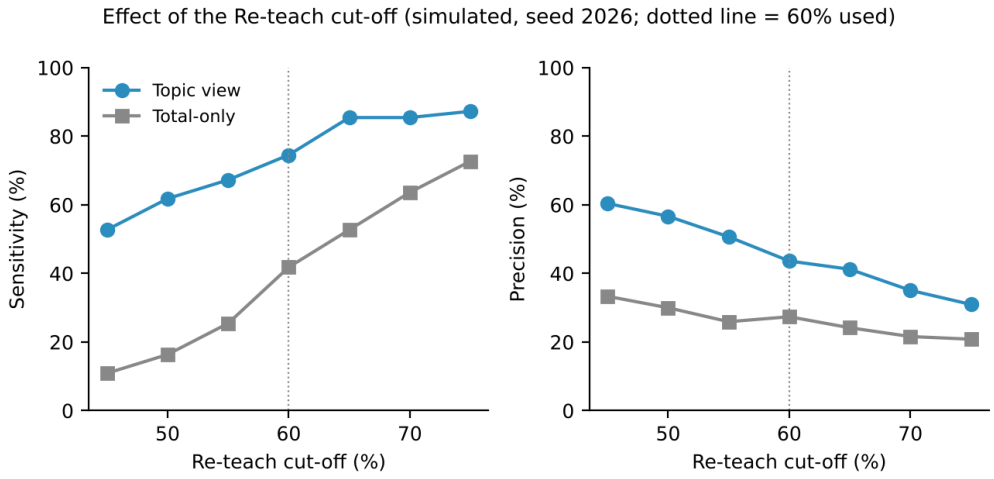


Fig. 9 - Sensitivity and precision of both views at Re-teach cut-offs from 45% to 75% (seed 2026). The dotted line marks the 60% cut-off used in this paper.

Effect of gap size (Figure 10). We re-ran the simulation with planted gaps of four sizes, using 30 classes for each size. As gaps grew from small (0.8) to large (2.0), the topic view's mean sensitivity rose from 56.4% to 87.1%, and the total-only view's from 35.1% to 50.6%. The topic view led at every size. Its lead grew from 21 to 37 percentage points as gaps became larger.

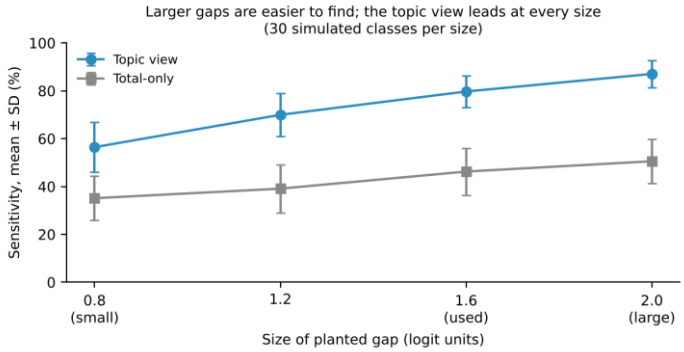


Fig. 10 - Mean sensitivity (± SD) of both views by size of planted gap, with 30 simulated classes per size.

6. Discussion

The topic view nearly doubles what a teacher can see. With the same marks, the topic view found 74.5% of planted gaps against 41.8% for a total-only view. It beat the total-only view in all 100 repeated simulations. The extra gaps it found all belonged to students whose totals looked acceptable. This is the information that the correctives step of mastery learning needs (Bloom, 1968; Guskey, 2007), and that a gradebook total throws away.

It is cheap. Processing took a fraction of a second, and the validation step caught every injected error. The main cost is the one-off work of tagging questions to topics. That work can be reused every year and shared within a department, which answers the time-cost barrier reported in reviews of data use in schools (Datnow & Hubbard, 2016).

Each page leads to an action. The re-teach list turned the heat map into concrete group sizes, such as whole-class re-teaching of identities in 11B and small groups in 11A. Teachers act on dashboard information when it links to a decision (Molenaar & Knoop-van Campen, 2019). This matches the resilience principle that monitoring adds value only when it is tied to a defined response (Abdul Salam, 2026a). The heat-map layout follows hazard-interaction displays in permit-to-work systems, where supervisors scan rows and columns they already know (Abdul Salam, 2024).

Flags are prompts, not verdicts. Precision was modest, at 43.6%. Many flags that were not planted gaps belonged to generally weaker students or sat just under the 60% line. A teacher should read a flag as "check this", and a short re-check question settles it. The cut-off is a dial the teacher controls. Moving it from 60% to 65% found more gaps (85.5% against 74.5%) for 20 extra flags, with almost no loss of precision (Figure 9). In safety terms, a leading indicator is useful because it raises early, cheap questions, even when some turn out to be false alarms.

Not every pattern shows, and some patterns are chance. The limits-to-differentiation link was detected, but the unit-circle link was not, because too few students failed the easiest topic. The planted T2 weakness in 11B was visible but not statistically significant. Large section differences also appeared on topics where nothing was planted. With 30 students per section, a dashboard will show chance differences as clearly as real ones. So teachers should look for patterns that persist across assessments before changing a scheme of work.

A record that follows the student. The model stores bands by topic code, not by test. So a Grade 12 teacher could see that a student's Grade 11 limits topic is still at Developing. Keeping records continuous across handovers is central to longitudinal exposure management in construction (Abdul Salam, 2025), and it transfers directly to school transitions. Saudi Vision 2030 and its education programmes call for data-informed, digitally enabled teaching (Abdul Salam, 2026b; Kingdom of Saudi Arabia, 2016). The approach shows that a classroom teacher can contribute to this without a large platform, provided privacy is built in from the start.

7. Limitations

- **Simulated data.** The class follows a model we chose. Real students' errors may cluster differently, topics may overlap more, and marking adds noise the model ignores. The results show that the tool works as designed, not that it works in every classroom.
- **Known answers.** We planted the gaps, so the detection rates depend on our choice of gap size and cut-off. Figures 9 and 10 show that the topic view's advantage held across the cut-offs and gap sizes tested, but small gaps were found only about half the time.
- **No teachers or learning outcomes.** The study does not show how teachers use the dashboard or whether re-teaching improves scores.
- **One topic per question.** Each question was given one topic code. Real questions often test two topics.
- **Figures from Python.** Figures 2–10 were drawn in Python from the same files that Power BI reads, so the Power BI layout itself was not tested here.

The next step is a classroom study. One or more teachers would use the dashboard for a term, with school permission and student opt-out. That study would log their decisions, compare re-taught topics before and after, and measure usability with the System Usability Scale (Brooke, 1996).

8. Conclusion

A test total hides where learning breaks down. This paper presented a low-cost dashboard that recovers that information from marks a teacher already records, using a spreadsheet, a short Python script and Power BI. On a simulated class of 60 Grade 11 students with 55 planted gaps, the topic view found 74.5% of the gaps, against 41.8% for a total-only view. It beat the total-only view in all 100 repeated simulations. It also revealed that 15 of the 33 high-scoring students had a topic needing re-teaching, and it showed a planted prerequisite link between limits and differentiation. The approach is fast, protects privacy and can be reproduced. It borrows a lesson from safety-critical industries: watch the leading indicators, and act before the final exam. The dataset, scripts and data model are provided so that other teachers can test and adapt the tool before a classroom evaluation.

Declarations

Conflict of interest. The authors declare no conflict of interest. The second author is the author of four works cited in Sections 2.4 and 6.

Author contributions. Jannathul Firdous: conceptualisation, topic map and assessment design, simulation design, interpretation, writing. Shajahan Abdul Salam: data-pipeline and dashboard design, framing from safety-critical industries, review and editing.

Ethics. The study used only simulated data with fictional names. No human participants were involved.

Data availability. The simulated dataset, the simulation, pipeline and analysis scripts, and the replication results are available from the corresponding author on request.

Use of generative AI. A generative AI assistant (Claude, Anthropic) was used to help write the simulation and analysis code and to structure and draft the manuscript text. The authors specified the study design, checked every result against the script outputs and every reference against its source, and take full responsibility for the content.

Appendix A. Python processing script (outline)

The input is one sheet per assessment. Each row is a student, each column is a question ID, and each cell holds the marks awarded. A second file, questions.csv, lists question_id, assessment_id, topic_code and marks_available.

```
"""Processing script. Reads marks.xlsx + questions.csv,
pseudonymises names, validates marks and writes mastery.csv for Power BI."""
```

```
import hashlib, time, pandas as pd

t0 = time.perf_counter()
SALT = open("salt.txt").read().strip()          # kept offline, never shared

def code(name):
    return "S" + hashlib.sha256((SALT + name.strip().lower()).encode()).hexdigest()[:6].upper()

q = pd.read_csv("questions.csv")
marks = pd.read_excel("marks.xlsx", sheet_name=None)

rows = []
for assessment, df in marks.items():
    long = df.melt(id_vars=["name", "section"], var_name="question_id", value_name="awarded")
    long["student"] = long["name"].map(code)
    rows.append(long.drop(columns="name"))
scores = pd.concat(rows).merge(q, on="question_id")

bad = scores[scores.awarded > scores.marks_available]
bad.to_csv("errors.csv", index=False)
scores = scores[scores.awarded <= scores.marks_available]      # hold flagged rows back for correction

key = pd.concat(marks.values())[["name"]].drop_duplicates()
key.assign(student=key["name"].map(code)).to_csv("KEY_offline.csv", index=False)

mastery = (scores.groupby(["student", "section", "topic_code"])
               .agg(awarded=("awarded", "sum"), available=("marks_available", "sum"),
                    n_items=("question_id", "nunique")).reset_index())
mastery["pct"] = (100 * mastery.awarded / mastery.available).round(1)
mastery["band"] = pd.cut(mastery.pct, [-1, 59.99, 79.99, 100],
                         labels=["Re-teach", "Developing", "Secure"]).astype(str)
mastery.loc[(mastery.n_items < 3) & (mastery.available < 6), "band"] = "Too little evidence"

mastery.to_csv("mastery.csv", index=False)
scores.drop(columns=["section"]).to_csv("item_scores.csv", index=False)
print(f"rows={len(scores)} flagged={len(bad)} runtime={time.perf_counter()-t0:.2f}s")
```

In Power BI, load mastery.csv, item_scores.csv, questions.csv and a topics.csv table, then relate them on student, question_id and topic_code. Use a matrix visual with conditional formatting on band for the heat map. Never load KEY_offline.csv.

Appendix B. Topic map used in the simulation (Grade 11)

Code	Strand	Topic	Difficulty b in simulation
T1	Trigonometry	Radian measure and the unit circle	−0.6
T2	Trigonometry	Trigonometric identities	0.5
T3	Trigonometry	Solving trigonometric equations	0.3
T4	Trigonometry	Graphs of trigonometric functions	−0.3
C1	Calculus	Limits and continuity	0.2
C2	Calculus	Introduction to differentiation	0.0

REFERENCES

- Abdul Salam, S. (2024). A field-ready permit-based coordination protocol for dynamic SIMOPS hazard interaction review, critical control verification, and operational reassessment. *International Journal of Engineering Applied Sciences and Technology*, 8(10), 189–198. https://doi.org/10.33564/IJEAST.2024.v08i10.031
- Abdul Salam, S. (2025). Worker exposure profiles for construction: A conceptual framework for exposure record continuity. *International Journal for Multidisciplinary Research*, 7(1). https://doi.org/10.36948/ijfmr.2025.v07i01.88492
- Abdul Salam, S. (2026a). A risk-based safety resilience framework for brownfield projects. *Journal of Mechanical, Civil and Industrial Engineering*, 7(5), 67–89. https://doi.org/10.32996/jmcie.2026.7.5.5
- Abdul Salam, S. (2026b). Transforming occupational health and safety under Saudi Vision 2030: A strategic framework for predictive, digital and resilient workplaces. *Iconic Research and Engineering Journals*, 9(8), 2637–2675. https://doi.org/10.64388/IREV9I8-1723472
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. https://doi.org/10.1080/0969595980050102
- Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment*, 1(2), 1–12.
- Bodily, R., & Verbert, K. (2017). Review of research on student-facing learning analytics dashboards and educational recommender systems. *IEEE Transactions on Learning Technologies*, 10(4), 405–418. https://doi.org/10.1109/TLT.2017.2740172
- Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester & I. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189–194). Taylor & Francis.
- Datnow, A., & Hubbard, L. (2016). Teacher capacity for and beliefs about data-driven decision making: A literature review of international research. *Journal of Educational Change*, 17(1), 7–28. https://doi.org/10.1007/s10833-015-9264-2
- Guskey, T. R. (2007). Closing achievement gaps: Revisiting Benjamin S. Bloom's "Learning for Mastery". *Journal of Advanced Academics*,

19(1), 8–31. https://doi.org/10.4219/jaa-2007-704

- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. https://doi.org/10.3102/003465430298487
- Kingdom of Saudi Arabia. (2016). *Saudi Vision 2030*. https://www.vision2030.gov.sa
- Kingdom of Saudi Arabia. (2021). *Personal Data Protection Law* (Royal Decree No. M/19 of 9/2/1443H, amended by Royal Decree No. M/148 of 5/9/1444H). Saudi Data & AI Authority.
- Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, 46(5), 30–40.
- Mandinach, E. B., & Gummer, E. S. (2016). *Data literacy for educators: Making it count in teacher preparation and practice*. Teachers College Press.
- Molenaar, I., & Knoop-van Campen, C. A. N. (2019). How teachers make dashboard information actionable. *IEEE Transactions on Learning Technologies*, 12(3), 347–355. https://doi.org/10.1109/TLT.2018.2851585
- Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, 38(11), 2074–2102. https://doi.org/10.1002/sim.8086
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Danish Institute for Educational Research.
- Schwendimann, B. A., Rodríguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., Gillet, D., & Dillenbourg, P. (2017). Perceiving learning at a glance: A systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies*, 10(1), 30–41. https://doi.org/10.1109/TLT.2016.2599522
- Tall, D., & Vinner, S. (1981). Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics*, 12(2), 151–169. https://doi.org/10.1007/BF00305619
- Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist*, 57(10), 1500–1509. https://doi.org/10.1177/0002764213479363
- Weber, K. (2005). Students' understanding of trigonometric functions. *Mathematics Education Research Journal*, 17(3), 91–112. https://doi.org/10.1007/BF03217423
- Wiliam, D. (2011). *Embedded formative assessment*. Solution Tree Press.