Automated feedback on student attempts to produce a set of dimensionless power products from a set of physical quantities that describe a physical problem

By Karl Lundengård[1], Peter Johnson[2], Phil Ramsden[1]

[1]Department of Mathematics, Imperial College London
[2]Department of Mechanical Engineering, Imperial College London

k.lundengard@imperial.ac.uk; peter.johnson@imperial.ac.uk; p.ramsden@imperial.ac.uk

PREPRINT

Formative feedback is important in learning. Automating the provision of specific, objective, constructive feedback to large cohorts requires complex algorithms that most teachers do not have time to develop, suggesting that a community effort is needed to create a library of specialised algorithms. We present an exemplar algorithm for a class of tasks in 'dimensional analysis' relating to the Buckingham Pi theorem. The challenge arises because there are infinitely many valid and invalid answers, but any valid answer is sufficient to complete a task. We present an algorithm that, given one valid reference answer, can evaluate any response to a task. The algorithm uses a vector-space formulation of the Pi theorem. Deployment across seven tasks for 380 students provided feedback 3,090 times, including stating why a response is invalid. The most common reason was that a student-proposed set was not dimensionless, but other educationally relevant reasons were also identified.

1 INTRODUCTION

Formative feedback that is timely, specific, objective, and constructive can help students learn (Hattie & Timperley, 2007; Shute, 2008). In this paper, we focus specifically on how to automate feedback with the purpose of serving large cohorts.

In a systematic review of 109 feedback automation systems, Deeva et al. (2021) classified automated feedback as non-exclusively one or more of: informative, corrective, suggestive, and/or motivational. In this paper, we focus on informative and corrective feedback. We omit suggestive and motivational feedback from automation and assume that either they are the role of the tutor, noting especially the arguments of Polya (2004, p.20), or, if automated, require a different type of automation outside the scope of this paper. The focus here is on the algorithm for generating technical feedback.

Models and typologies have been suggested to determine what 'good' feedback is, but there is no consensus in the literature (Lipnevich & Panadero, 2021; Panadero & Lipnevich, 2022). Therefore, 'feedback design' remain local and contextual, according to the intent of the teacher.

For mathematical tasks in general, we conceive of feedback generation as comparing a mathematical expression given by a student (a *candidate expression*) to some reference criteria provided by a teacher or teaching resource as

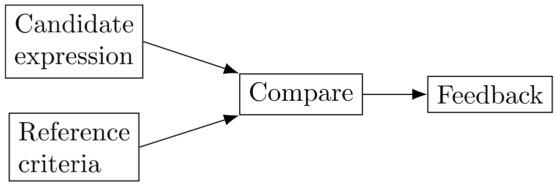

Figure 1. A conceptual process of feedback generation.

illustrated in Figure 1. If the candidate expression is invalid, then information on why it is invalid is required; ideally, information upon which the learner can act.

A reference criterion can be simple, for example, checking for equality with a reference expression, or more complex criteria may be required. For example, if there are many, perhaps even infinitely many, unique and valid answers to an educational task, then only comparing to one reference answer could result in inappropriate feedback for students who responded with an alternative valid answer.

When teachers manually generate feedback for students, they solve this 'feedback problem' using a complex combination of tacit rules, contextual awareness, and broad mathematical knowledge. To provide feedback automatically by writing computational algorithms, solving the feedback problem in a more explicit way is required.

The requirement for an explicit algorithm can be surprisingly challenging to meet. This paper presents one such algorithm as an exemplar. It is not realistic for individual teachers to develop their own algorithms for each task because too much time and specialist knowledge are required to develop each algorithm. We therefore argue that as a community, we need to collectively develop a library of such algorithms, each specialised to an educational task.

We give an exemplar algorithm in this paper for an educational task in 'dimensional analysis', which commonly appears in undergraduate studies of engineering and applied science (e.g. White, 2016, §5.3). A distinctive property of the task is that there are infinitely many unique valid responses.

The paper has three goals: first, to describe an algorithm and the feedback that it can provide to students. Second, to show that when deployed, it produces useful data for teachers. Third, to show by example that good practice for a feedback algorithm is to be robust, general, and simple to configure.

The paper is structured as follows. The theory of dimensional analysis is summarised in §2, together with an example task and a definition of the feedback problem. In §3, we present an algorithm to solve the feedback problem and highlight the type of feedback it can generate. In §4, we present results from deploying the algorithm to students on a web platform as part of their studies. In §5, we conclude that our initial goals of showing the functionality of the algorithm have been achieved and discuss the wider implications and future needs to evaluate the educational impact of this work.

2 THEORY AND PROBLEM DEFINITION

This section provides a summary of 'dimensional analysis' to give context to the feedback problem we solved in this paper. We define a dimensional matrix, the Buckingham Pi theorem, and the feedback problem.

2.1 BACKGROUND

Engineering students are often taught that if a physical problem is defined as finding a relation between a given list of n physical quantities, q_j,

$$g(q_1, q_2, \dots, q_n) = 0, \tag{1}$$

then the problem can be simplified to that of finding the relation between a shorter list of k power products, π_{i}, that are necessarily dimensionless and composed of the quantities q_i:

$$G(\pi_1, \pi_2, \dots, \pi_k) = 0, \tag{2}$$

where $k < n$. This statement is the Buckingham Pi theorem; a more precise version is given in §2.2.

For a given set of physical quantities, q_j, a student may be asked for a set of dimensionless quantities, π_{i}. The feedback problem for this task includes evaluating the validity of the candidate set of quantities, π_{i}^C, offered by the student. The technical challenge is that the candidate set may not resemble the reference set but may still be valid. If, alternatively, the student is incorrect, then constructive feedback is desirable.

To demonstrate the main aspects of the problem, consider, for example, the force, f, on an object of length scale ℓ, flowing at a speed v, through a fluid of density, ρ, and viscosity, μ. The relationship $\phi(f, \ell, v, \rho, \mu) = 0$ has $n = 5$ quantities but can be simplified, according to the Pi Theorem, to a relationship between $k = 2$ dimensionless power products. A typical reference expression is

$$\pi_1^R = \frac{f}{\frac{1}{2}\rho v^2 \ell^2}, \qquad \pi_2^R = \frac{\rho v \ell}{\mu}. \tag{3}$$

A student may propose a candidate set that is different from the reference set but still valid. For a simple relation, e.g., with candidates given by $2\pi_1^R$ and $1/\pi_2^R$, the validity of the candidate set is evident. However, consider the candidate set

$$\pi_1^C = \frac{f}{\mu v \ell}, \qquad \pi_2^C = \frac{f\rho}{\mu^2}. \tag{4}$$

Since $\pi_1^C = \frac{1}{2}\pi_1^R \pi_2^R$ and $\pi_2^C = \pi_1^R (\pi_2^R)^2$, the candidate set (4) is valid, but this is not immediately obvious, even to someone familiar with the subject matter.

Examples of candidate sets with corresponding feedback are given in Table 1, including the two foregoing examples. Given that infinitely many valid and invalid candidate sets are possible, and validity is not obvious, a general algorithm is required to compare any candidate set π_i^C with a single reference set π_i^R.

	Candidate set	Valid?	Feedback
A	$\frac{f}{\rho v^2 \ell^2}$, $\frac{\rho v \ell}{\mu}$	✓	Correct
B	$\frac{f}{\mu v \ell}$, $\frac{f\rho}{\mu^2}$	✓	Correct
C	$\frac{fv}{\mu \ell}$, $\frac{f\rho}{\mu^2}$	✗	The power product $\frac{fv}{\mu\ell}$ is not dimensionless.
D	$\frac{f}{\mu v \ell}$, $\frac{\mu v \ell}{f}$	✗	The power products are not independent.
E	$\frac{f}{\mu v \ell}$	✗	Too few power products. Expected 2, 1 was given.
F	$\frac{f}{\rho v^2 \ell^2}$, $\frac{\rho v \ell}{\mu}$, $\frac{f}{\mu v \ell}$	✗	Too many power products. Expected 2, 3 were given. Power products not independent.
G	$\frac{f}{P v^2 \ell^2}$, $\frac{P v \ell}{X}$	✗	Unknown symbols: P, X
H	$f/$	✗	Expression could not be interpreted

Table 1: Example feedback on a sample of candidate sets.

2.2 DIMENSIONAL MATRIX

Throughout this paper, the concept of a *dimensional matrix* is needed and is defined here. A physical quantity, q_j, has physical dimensions, $[q_j]$, which are the product of some set of fundamental physical dimensions. If the fundamental physical dimensions are denoted by $[u_1], [u_2], \dots, [u_m]$ then $[q_j]$ can be written in the form

$$[q_j] = \prod_{i=1}^{m} [u_i]^{a_{ij}}.$$

For example, if q_1 is a force, it has dimensions

$$[q_1] = [M]^1 [L]^1 [T]^{-2},$$

where the $[M]$ (mass), $[L]$ (length), and $[T]$ (time) are fundamental physical dimensions $[u_j]$. Hence,

$$a_{11} = 1,\ a_{21} = 1,\ a_{31} = -2.$$

For a set of n quantities with a total of m fundamental physical dimensions, the elements a_{ij} comprise an $m \times n$-

dimensional matrix, $\boldsymbol{A}$. The element a_{ij} of $\boldsymbol{A}$ corresponds to the exponent of the ith fundamental physical dimension of the jth quantity. For the example in §2.1, the dimensional matrix would be

$$\boldsymbol{A} = \begin{array}{c|c} & \begin{matrix} f & \ell & v & \rho & \mu \end{matrix} \\ \begin{matrix} M \\ L \\ T \end{matrix} & \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & -3 & -1 \\ -2 & 0 & -1 & 0 & -1 \end{bmatrix} \end{array}. \qquad (6)$$

The row and column labels in (6) show the chosen order of the fundamental physical units and quantities but are not part of the matrix. If a column of $\boldsymbol{A}$ is all zero, then the corresponding quantity q_j is said to be 'dimensionless', i.e. $[q_j] = [0]$. The matrix $\boldsymbol{A}$, or its elements a_{ij}, define a task.

The appropriate set of fundamental quantities (such as mass, length, and time) is arbitrary, and the choice is contextual. In physics, the smallest viable subset of the seven quantities used to define the fundamental SI units (time, length, mass, electric current, thermodynamic temperature, amount of substance and luminous intensity) is usually chosen.

2.3 THE BUCKINGHAM PI THEOREM

The Buckingham Pi theorem was named by Bridgman (1922) after Buckingham (1914). The theorem has a long history and several variants and generalisations (Longo 2021, §1.4). The theorem is stated for use in this paper as follows:

Let q_1, q_2, ... , q_n, be n physical quantities. Then any 'complete' (i.e. true for all systems of units, see Bridgman 1922, Ch. IV) equation,

$$g(q_1, q_2, \dots, q_n) = 0$$

can be restated as

$$G(\pi_1, \pi_2, \dots, \pi_k) = 0$$

where $k = n - r$ and $r = \operatorname{rank}(\boldsymbol{A})$.

For the power products π_i to be valid they must be dimensionless power products of the form

$$\pi_i = \prod_{j=1}^{n} q_j^{b_{ij}}, \ \ i \in \{1, \dots, k\}, \ \ b_{ij} \in \mathbb{R}, \qquad (7)$$

and the π_i must be independent of each other, i.e. no π_i can be rewritten as $\pi_i = \pi_1^{\sigma_1} \dots \pi_{i-1}^{\sigma_{(i-1)}} \pi_{i+1}^{\sigma_{(i+1)}} \dots \pi_k^{\sigma_k}$, where all exponents are real numbers.

The exponents b_{ij} in (7) may be restricted to a smaller set, such as integers or fractions, for physical reasons. The algorithm proposed in §3.1 is not affected by such a restriction; hence, the formulation here is more permissive.

2.4 THE FEEDBACK PROBLEM

A teacher, in creating a task of the type considered here, may have a particular valid set of power products in mind; however, there are infinitely many valid student responses. For this reason, the algorithm presented here is to evaluate any candidate set based on a single *reference set* of power products, Π^R:

$$\Pi^R = \left\{\pi_i^R \mid i \in \{1, \dots, n-r\}\right\},$$

where

$$\pi_i^R = \prod_{j=1}^{n} q_j^{\rho_{ij}}, \ \ i \in \{1, \dots, n-r\}, \ \rho_{ij} \in \mathbb{R}$$

are known to be valid (i.e. satisfy the conditions in §2.3).

A student will provide a *candidate set*, i.e. a set of (not necessarily dimensionless) quantities,

$$\Pi^C = \left\{\pi_i^C \mid i \in \{1, \dots, s\}\right\},$$

where s is a characteristic of the student response, and

$$\pi_i^C = \prod_{j=1}^{n} q_j^{c_{ij}}, \ i \in \{1, \dots, s\}, \ \ c_{ij} \in \mathbb{R}.$$

To construct an algorithm to determine the validity of a candidate set, we will rely (in §3.1) on matrices of exponents for the reference and candidate sets, denoted as follows:

$$\boldsymbol{R} = \begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k1} & \rho_{k2} & \cdots & \rho_{kn} \end{bmatrix}, \quad \text{(reference)} \qquad (8)$$

$$\boldsymbol{C} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{k1} & c_{k2} & \cdots & c_{kn} \end{bmatrix}. \quad \text{(candidate)} \qquad (9)$$

As a concrete example of the reference and candidate matrices of exponents in (8) and (9), in the example in §2.1, the comparison is between the reference set

$$\Pi^R = \{\pi_1^R, \pi_2^R\} = \left\{\tfrac{1}{2} f \rho^{-1} v^{-2} l^{-2}, \rho v l \mu^{-1}\right\}$$

and one of the candidate sets, e.g.

$$\Pi^C = \{\pi_1^C, \pi_2^C\} = \{f \mu^{-1} v^{-1} l^{-1}, f \rho \mu^{-2}\}.$$

In matrix form:

$$\boldsymbol{R} = \begin{bmatrix} 1 & -2 & -2 & -1 & 0 \\ 0 & 1 & 1 & 1 & -1 \end{bmatrix}, \quad \text{(reference)}$$

$$\boldsymbol{C} = \begin{bmatrix} 1 & -1 & -1 & 0 & -1 \\ 1 & 0 & 0 & 1 & -2 \end{bmatrix}. \quad \text{(candidate)}$$

The goal of this paper is to construct an algorithm that solves the following problem:

Assuming a reference set Π^R is known — but without needing the dimensional matrix $\boldsymbol{A}$ — determine the validity of a candidate set Π^C; i.e. determine whether Π^C satisfies the conditions in §2.3. The algorithm will compare the $\boldsymbol{R}$ and $\boldsymbol{C}$ matrices and provide outputs in the form of Table 1.

3 ALGORITHM TO EVALUATE RESPONSES

In this section, we present an algorithm to solve the problem stated in §2.4. We cover validity in §3.1, feedback to students in §3.2, and deployment in §3.3.

3.1 VALIDITY OF A CANDIDATE SET

We will use vector spaces as a framework to determine the validity of a candidate set. Vector spaces were used to formalise dimensional analysis and prove the Buckingham Pi theorem from the 1950s onwards (e.g., Birkhoff, 1960), but in more recent literature it is common to either use less abstract mathematics for practical application, e.g. descriptions similar to Bridgman (1922), or use more complex mathematical structures to capture more of the desired properties of physical dimensions; for more details see e.g. Longo (2021, §1.4). For the current use case, the vector space description is sufficient.

The conditions for a valid set were given in §2.3 as

1. Every power product in the set is dimensionless
2. The power products are independent
3. The number of power products is $k = n - r$

A set described by $\boldsymbol{B}$ (with elements b_{ij}, defined in §2.3) is dimensionless if it satisfies $\boldsymbol{A}\boldsymbol{B}^{\top} = \boldsymbol{0}$, where $\boldsymbol{A}$ is the dimensional matrix as in §2.2. In words, all vectors that form a column of $\boldsymbol{B}^{\top}$ must be in the *null space* (also known as the *kernel*) of $\boldsymbol{A}$ (see e.g. Langhaar 1951, Ch. 3). A set of k power products is independent if $\operatorname{rank}(\boldsymbol{B}) = k$. Thus, validity requires the span of the rows of $\boldsymbol{B}$ to be the null space of $\boldsymbol{A}$.

Checking validity does not fully solve the problem defined in §3.2 for two reasons. First, a reason for invalidity is needed to give useful feedback; second, to reduce the workload on the teacher configuring the task, we want to only require a reference set, Π^R, rather than specifying the entire dimensional matrix $\boldsymbol{A}$.

A different way to determine the validity of a candidate set Π^C, which does not rely on knowing $\boldsymbol{A}$, is to determine if the span of the rows of $\boldsymbol{C}$, i.e. $\operatorname{span}(\Pi^C)$ if each power product is represented by a vector row of $\boldsymbol{C}$, is the same as the span of the rows of $\boldsymbol{R}$, i.e. $\operatorname{span}(\Pi^R)$. An explanation follows.

From the rank-nullity theorem. the null space of $\boldsymbol{A}$ is a unique vector space with dimension k (Kreyzig 2006, §7.4). A vector space of dimension k can be described using any k linearly independent vectors as a basis. Thus, if Π^R satisfies the Pi Theorem (§2.3), then the span of Π^R is the null space of $\boldsymbol{A}$. If Π^R and Π^C have the same span, then the span of Π^C is also the null space of $\boldsymbol{A}$. Thus, comparing the spans of Π^R and Π^C can establish the validity of Π^C.

We can determine that $\operatorname{span}(\Pi^C)$ is the same as $\operatorname{span}(\Pi^R)$ by testing if $\operatorname{span}(\Pi^C)$ is a subset of $\operatorname{span}(\Pi^R)$ and vice versa. If $\operatorname{span}(\Pi^C)$ is a subset of $\operatorname{span}(\Pi^R)$, then for any $\pi_i^C \in \Pi^C$ the corresponding row in $\boldsymbol{C}$ can be written as a linear combination of rows in $\boldsymbol{R}$. One way to determine if this is the case is to create a new matrix composed of $\boldsymbol{R}$ and $\boldsymbol{C}$, i.e.

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{C} \end{bmatrix} = \begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k1} & \rho_{k2} & \cdots & \rho_{kn} \\ c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{k1} & c_{k2} & \cdots & c_{kn} \end{bmatrix}$$

and check if $\operatorname{rank}(\boldsymbol{R}) = \operatorname{rank}(\boldsymbol{D})$. Similarly, all $\pi_i^R \in \Pi^R$ are in $\operatorname{span}(\Pi^C)$ if $\operatorname{rank}(\boldsymbol{C}) = \operatorname{rank}(\boldsymbol{D})$. Thus, Π^R and Π^C have the same span if $\operatorname{rank}(\boldsymbol{R}) = \operatorname{rank}(\boldsymbol{C}) = \operatorname{rank}(\boldsymbol{D})$.

This section provided a mathematical basis for an algorithm that determines candidate set validity. A candidate Π^C is valid if and only if $\operatorname{rank}(\boldsymbol{R}) = \operatorname{rank}(\boldsymbol{C}) = \operatorname{rank}(\boldsymbol{D})$ and $\operatorname{rank}(\boldsymbol{C}) = |\Pi^C|$ (i.e. $\boldsymbol{C}$ has full rank).

3.2 GENERATING FEEDBACK

In this section, we extend the test for validity given in §3.1 to provide a reason for why a candidate set is invalid.

Cases that generate different feedback are listed in Table 2. Cases 5 and 6 are assumed to be handled before executing the algorithm. A candidate set is valid if Case 0 applies and none of Cases 1–6 apply. A flowchart describing the algorithm is given in Figure 2.

In Case 1 (see Table 2), where $\operatorname{rank}(\boldsymbol{R}) < \operatorname{rank}(\boldsymbol{D})$, at least one power product in the candidate set is not dimensionless. A further enhancement (not present in Figure 2) is to identify which power product(s) is/are not dimensionless. The offending power product(s) is/are found by checking each row of $\boldsymbol{C}$ separately: append row i of $\boldsymbol{C}$ to $\boldsymbol{R}$, to create $\boldsymbol{D}_i$,

$$\boldsymbol{D}_i = \begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k1} & \rho_{k2} & \cdots & \rho_{kn} \\ c_{i1} & c_{i2} & \cdots & c_{in} \end{bmatrix}.$$

If $\operatorname{rank}(\boldsymbol{D}_i) \neq \operatorname{rank}(\boldsymbol{R})$ then the π_i^C is not dimensionless.

For Case 2, the number of rows in $\boldsymbol{C}$ is determined by the number of elements in Π^C so $\operatorname{rank}(\boldsymbol{C}) \leq |\Pi^C|$. The case $\operatorname{rank}(\boldsymbol{C}) < |\Pi^C|$ means that the power products in the candidate set are not independent.

Case 3, $\operatorname{rank}(\boldsymbol{R}) = \operatorname{rank}(\boldsymbol{D})$ and $\operatorname{rank}(\boldsymbol{C}) < \operatorname{rank}(\boldsymbol{D})$, implies that the power products in the candidate set are dimensionless but that there are too few independent ones. The validity of the total number of power products ($|\Pi^C|$) cannot be determined using the ranks, so it is done by comparing the number of elements in Π^R, e.g., as in Case 4.

Case 4 corresponds to too many power products being included in the candidate set (since the reference set contains a minimal number of power products). If Case 4 is found, then either Cases 1, 2 or 3 must also apply.

Case	Condition	Description	Examples
0	$\text{rank}(\boldsymbol{R}) = \text{rank}(\boldsymbol{C}) = \text{rank}(\boldsymbol{D})$	The candidate set has sufficiently many independent power products	A, B, F
1	$\text{rank}(\boldsymbol{R}) < \text{rank}(\boldsymbol{D})$	The candidate set has at least one dimensionless power product	C
2	$\text{rank}(\boldsymbol{C}) < \lvert\Pi^C\rvert$	The power products in the candidate set are not independent	D, F
3	$\text{rank}(\boldsymbol{C}) < \text{rank}(\boldsymbol{R}) = \text{rank}(\boldsymbol{D})$	The candidate set has too few independent power products	D, E
4	$\lvert\Pi^R\rvert < \lvert\Pi^C\rvert$	The candidate set has too many power products	F
5	Parser	Candidate set contains at least one unknown symbol	G
6	Parser	Expression could not be interpreted	H

Table 2. Cases detected by the algorithm. Examples refers to which of the examples in Table 1 the case is found to apply.

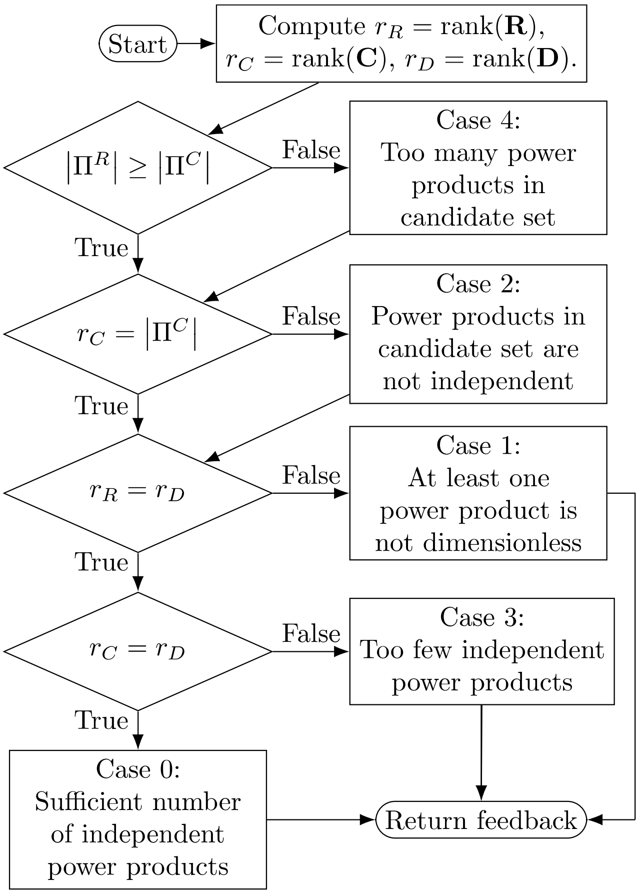

Figure 2. Flowchart illustrating the proposed algorithm. Cases are defined in Table 2.

Cases are not mutually exclusive. The possible case combinations are: (0), (1), (3), (1, 2), (2, 3), (1, 4), (0, 2, 4), (1, 2, 4), and (2, 3, 4). Only combination (0) corresponds to a valid candidate set. For two of the examples in Table 1, more than one case applies, D (2, 3) and F (0, 2, 4).

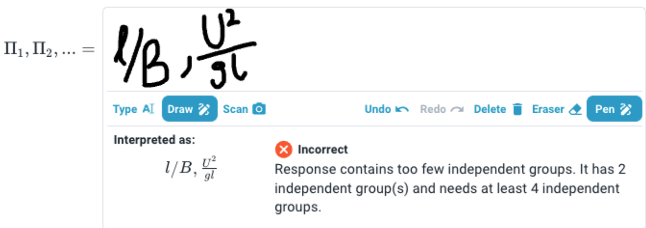

Figure 3. Screenshot of feedback in the web interface

3.3 DEPLOYMENT

The authors implemented the proposed algorithm using Python and SymPy (Meurer et al., 2017) and integrated it with a web interface that lets teachers author tasks. The web interface was available as a standard part of the curriculum for two consecutive cohorts of ~190 students in an undergraduate Mechanical Engineering degree. In each cohort, seven tasks, which were part of a module on Fluid Mechanics, were deployed to the students with automated feedback determined as described in §3.2. Table 3 describes the tasks where the reference sets contained one to four power products and involved two to seven quantities. The collected responses and an implementation of the algorithm, as described in §3.1-3.2, are publicly available on FigShare (Lundengård et al., 2023).

Students completed the tasks as self-study, which was an expected part of their curriculum but not enforced by summative assessment or other form of credit, which was separate. Students had access to answers and worked on solutions on the web interface, including branched solutions for the most common different valid approaches. Candidate sets were submitted via a web interface that supported both typing and handwriting (see Figure 3) and received automated formative feedback within a few seconds. Submission was optional, with the only incentive being to receive formative feedback without seeing the answer. There was no limit to the number of times students could use the automated feedback.

Cases 5 and 6 were handled before calling the algorithm. For Case 5, *'Unknown symbol'*, only the unknown symbols are identified in the feedback message. To reduce the occurrence of this kind of error, the teacher could specify other symbols (e.g. upper- and lowercase versions of a symbol and similar characters like 't' for 'τ') to mean the same thing.

Case 6, *'Invalid expression'*, occurred in early deployment but was subsequently prevented by displaying an interpretation of the response to the student before submission, i.e. a 'preview' of the response. Submission was not allowed if the expression was not interpreted (i.e. parsed) successfully. After implementing this control, Case 6 was completely prevented, with no noticeable reduction in the number of student responses, suggesting that, as well as assuring quality feedback, it did not hinder students from using the system.

4 RESULTS

A total of 3,090 student responses were collected; the results are summarised in Table 4 and plotted in Figure 4. Complete results are in the supplemental data for this article (Lundengård et al., 2023). Approximately half of the student responses were valid. Invalid responses were always given qualitative feedback from one or more of cases 1 to 5. All case combinations listed in §3.2 occurred in at least one task.

For all tasks except 2 and 7, the most common responses included both valid and invalid candidate sets. In Tasks 1, 3, 6 and 7, one mathematically unique response occurred more frequently (by at least a factor of four), but for Tasks 2, 4 and 5, there were two or three common valid (mathematically unique) responses. Task 6 elicited the most diverse set of responses, with 39 unique, valid responses and 185 unique, invalid responses.

The most common mistake, accounting for one-third of all responses, was that one or more proposed power products were not dimensionless, i.e., Case 1.

Case 2 (products not independent) and Case 3 (too few products) only applied to Tasks 5–7, where multiple products are required. Students rarely proposed linearly dependent groups; Case 2 was very rare in Tasks 5–7 (2%, 4%, and 0%, respectively). Case 3 occurred in Tasks 5–7 at 6%, 15%, and 7%, respectively, approximately proportional to the expected number of products (2, 4, and 3, respectively), i.e. students tend to express too few products when more are required.

Case 4 (too many products) occurred more for Task 1 than Tasks 2–4, despite all tasks only requiring one product. All occurrences of Case 4 in Task 1 consisted of listing some of the quantities used in the expression, e.g. 'm, l, T, f', indicating that the students misunderstood the task. Since Tasks 2–4 are structurally very similar to Task 1, students are unlikely to repeat the same mistake – suggesting that students learned from the first task.

The highest frequency of Case 4 is seen in Task 5, possibly because Task 5 is the first task that requires more than one power product. That there are only two occurrences (0.4%) of Case 4 in Task 6 and none in Task 7 can possibly be explained by a tendency of students to assume a smaller number of power products when they are uncertain; this explanation is consistent with the higher occurrence of Case 3.

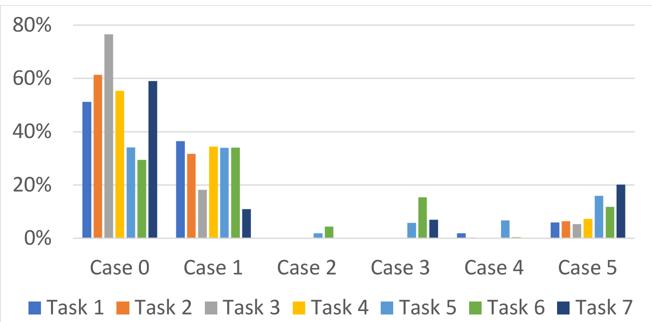

Figure 4: Summary of feedback from 3,090 responses. For a description of tasks, see Table 3, and for cases, see Table 2.

5 DISCUSSION AND CONCLUSION

The goal of the work presented here is to show that feedback can be automatically provided to students on educational tasks, including when there are infinitely many valid answers. In practice, there was a diversity of (mathematically) unique responses by students, including 39 unique valid responses for one task and hundreds of unique invalid responses for most tasks. The diversity of answers validates the need for an algorithmic approach rather than a lookup table of common answers.

The results show that timely, specific, objective feedback was provided every time in over 3,000 cases. The feedback goes beyond the binary question of validity and provides a meaningful message to students in all cases where their response was invalid. The most common such feedback was that proposed power products were not dimensionless. Submitting a response was optional for students who also had access to the answer. The high number of submissions across all tasks is strong evidence that students valued the feedback.

Automated feedback has value in the following ways. First, it is timely. Second, the student can avoid seeing the 'correct answer', which avoids spoiling the task if they are incorrect, but also avoids confusion if the student is correct but their answer looks very different from the reference answer. Third, automated feedback on technical issues can change the agenda for classroom activities, for example, to facilitate deeper discussions.

The algorithm presented here is specific to the problem that was defined, yet still involves significant complexity. Providing the same quality of feedback on other educational tasks will require the education community to define and solve many other specialised feedback problems. The work here serves as an exemplary feedback algorithm due to the following characteristics:

- Infinitely many valid and invalid answers, including a wide variety of unique responses produced in practice.
- Independent of programming language and the learning platform on which it is deployed.
- Only one reference set is required to determine validity and identify reasons for invalidity (with appropriate feedback). This reduces the workload for the task author.

The motivation behind the algorithm we have presented is to enhance student learning through direct effects of timely and high-quality feedback, as well as by changing the agenda of contact time in the classroom by automating lower-level technical feedback. The focus here was on the algorithm and the data gathered by its deployment. We have presented evidence that enhanced feedback is possible and that students value the feedback that they receive. Future work should explore the student and teacher experience using automated feedback and evaluate outcomes for those stakeholders.

Task	Description	n	m	r	Reference set
1	The frequency of oscillation, f, of a guitar string appears to depend only upon its mass m, length l, and tension, T. If this observation is correct, what dimensionless power product will fully describe the oscillation?	4	3	3	$\frac{f^2 ml}{T}$
2	If the velocity, u, of a liquid leaving a nozzle depends only upon the pressure, p, and density, ρ, what dimensionless power product will fully describe the oscillation?	3	3	2	$\frac{\rho u^2}{p}$
3	The area, A, of a circle is related to its radius, r. What dimensionless power product describes this relationship?	2	1	1	$\frac{A}{r^2}$
4	A rigid mass is released with zero initial velocity at a height h above the ground. It accelerates under gravity, g, until it hits the ground after time, t. Suggest a dimensionless that group describes this relationship.	3	2	2	$\frac{g\,t^2}{h}$
5	The thrust, F, of a propeller can depends on the speed U, air density ρ, propeller size D, and rotation frequency ω. Tests on a model propeller in a wind tunnel (air density $\rho = 1.2\,\mathrm{kg/m^3}$) gave the following results for the thrust at a number of forward velocities. U (m/s): 0, 10, 15, 20, 30 F (N): 300, 278, 245, 211, 100 The propeller diameter was 0.8 m and it was spun at 2000 rpm. Using dimensional analysis find the non-dimensional parameters which govern this observed behaviour	5	3	3	$\frac{F}{D^4\rho\omega^2}, \frac{U}{D\omega}$
6	The resistance of a sea-going ship is due to wave-making and viscous drag, and can be expressed as $F_D = f(U, l, B, \rho, \mu, g)$, where F_D is the drag force, U is the ship speed, l is its length, B is its width, ρ and μ are the sea water density and viscosity, and g is the acceleration due to gravity. Find the non-dimensional power products that describe the problem.	7	3	3	$\frac{F}{\rho l^2 U^2}, \frac{B}{l}, \frac{gl}{U^2}, \frac{\mu}{\rho U l}$
7	Wind with speed U blows perpendicularly to a suspension bridge of length l. The bridge deck of thickness h disturbs the flow, leading to the periodic release of vortices, the frequency of which is called the vortex shedding frequency, and is denoted by f. The shedding frequency depends on the deck thickness, the bridge length, the wind velocity, the air density ρ and dynamic viscosity μ and the Reynolds number Re.	5	2	2	$\frac{fl}{U}, \frac{h}{l}, \frac{\mu}{\rho U l}$

Table 3. Tasks from a learning module in Fluid Mechanics. Here n is the number of quantities, r the rank of the dimensional matrix and m the number of fundamental physical dimensions.

	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7
Total responses	566	390	353	448	586	478	229
Unique (written)	382	218	101	268	442	414	165
Unique (mathematical)	100	56	45	79	179	224	62
Total correct responses	291	230	261	243	208	147	135
Unique correct responses	7	5	6	6	22	39	15
Count of three most common responses	239, 37', 34	92, 84, 31	234, 20, 16'	154, 78, 55'	97, 62, 35’	106, 27', 13	84, 21, 14
Case 0: Sufficient number of independent products	291 51.4 %	230 59.0 %	261 73.9 %	243 54.2 %	218 35.5 %	147 30.7 %	135 59.0 %
Case 1: ≥ 1 product not dimensionless	207 36.6 %	119 30.5 %	62 17.6 %	151 33.7 %	217 37.0 %	170 35.5 %	25 10.9 %
Case 2: Power products not independent	0 0 %	0 0 %	0 0 %	0 0 %	12 2.0 %	22 4.6 %	0 0 %
Case 3: Too few power products	0 0 %	0 0 %	0 0 %	0 0 %	37 6.3 %	77 16.1 %	16 6.9 %
Case 4: Too many power products	11 1.9 %	1 0.3 %	0 0 %	0 0 %	43 7.3 %	2 0.4 %	0 0 %
Case 5: Unknown symbol	34 6.0 %	24 6.1 %	18 5.1 %	32 7.1 %	102 17.4 %	59 12.3 %	46 20.1 %
Case 6: Invalid mathematical expression	34 6.0 %	17 4.4 %	12 3.4 %	22 4.9 %	12 2.0 %	25 5.2 %	7 3.0 %

Table 4. Combined results from two cohorts of ~190 students. Occurrences marked with ' indicate an invalid candidate set. Notes: mistakes are not mutually exclusive; Case 6 was prevented in the second cohort, see Section 3.3.

REFERENCES

Bridgman, P. W. (1922). *Dimensional Analysis*, Yale University Press.

Buckingham, E. (1914). On physically similar systems; illustrations of the use of dimensional equations, *Physical Review*, 4(4), pp. 345 – 376.

Birkhoff, G. (1960). *Hydrodynamics: A Study in Logic, Fact, and Similitude*, 2nd ed., Oxford University Press.

Deeva, G., Bogdanova, D., Serral, E., Snoeck, M. & De Weerdt J. (2021). A review of automated feedback systems for learners: Classification framework, challenges and opportunities, *Computers & Education,* 162(104094), 2021.

Hattie, J., & Timperley, H. (2007). The Power of Feedback, *Review of Educational Research*, 77(1), pp. 81 – 112.

Kreyzig, E. (2006). *Advanced Engineering Mathematics*, 9th ed., Wiley.

Langhaar, H. L. (1951). *Dimensional Analysis and Theory of Models*, John Wiley & Sons.

Lipnevich, A. A., & Panadero, E. (2021). A review of feedback models and theories: Descriptions, definitions, and conclusions, *Frontiers in Education*, 6 (Online).

Longo, S. G. (2021). *Principles and Applications of Dimensional Analysis and Similarity*, Springer Nature Switzerland.

Lundengård, K., Johnson P., Ramsden P. (2023). *Collected responses and reference implementation of algorithm described in "Automated feedback on student attempts to produce a set of dimensionless power products from a list of quantities that describe a physical problem"*. FigShare. Dataset. https://doi.org/10.6084/m9.figshare.24850131

Meurer, A., Smith, C. P., Paprocki, M., Čertík, O., Kirpichev, S. B., Rocklin, M., Kumar, A., Ivanov S., Moore, J. K., Singh, S., Rathnayake, T., Vig, S., Granger, B. E., Muller, R. P., Bonazzi, F., Gupta, H., Vats, S., Johansson, F., Pedregosa, F., Curry, M. J., Terrel, A. R., Roučka, Š., Saboo, A., Fernando, I., Kulal, S., Cimrman, R., & Scopatz, A. (2017). SymPy: symbolic computing in Python, *PeerJ Computer Science* 3:e103, https://doi.org/10.7717/peerj-cs.103

Panadero E. & Lipnevich, A. A. (2022). A review of feedback models and typologies: Towards an integrative model of feedback elements, *Educational Research Review*, 35 (100416 Online).

Polya, G. (2004). *How to solve it: A new aspect of mathematical method*. No. 246. Princeton University Press.

Shute, V. J. (2008). Focus on formative feedback, *Review of Educational Research*, 78(1):153 – 189.

White, F. M. (2016). *Fluid Mechanics*, 8th ed, McGraw-Hill.

BIOGRAPHICAL NOTES

Karl Lundengård is a Research Associate at Imperial College London (UK). Currently working on the Lambda Feedback project. MSc in Engineering Physics from Lund University (Sweden) and PhD in Applied Mathematics from Mälardalen University (Sweden).

Peter Johnson is a Principal Teaching Fellow at Imperial College London (UK). Peter teaches Fluid Mechanics theory and practice and is leading the Lambda Feedback project at Imperial College to provide automated feedback on self-study. MEng in Mechanical Engineering from the University of London (UK) and Columbia University (US); PhD in Fluid Dynamics from UCL (UK); MEd in Undergraduate Learning and Teaching from Imperial College London (UK).

Phil Ramsden is the Director of Cross-Curricular Mathematics Education at Imperial College London (UK). Phil teaches applied mathematics and computing and leads the development of e-learning materials in mathematics. BA in Mathematics from the University of Cambridge (UK); MSc in Applicable Mathematics from Queen Mary University of London (UK); PhD in dynamical systems from Queen Mary University of London (UK).