# Student misconceptions of the language of error

**Paper**

**Jane Tomlinson, Paul J Dyson and John Garratt**

*Department of Chemistry, University of York, Heslington, York, YO10 5DD*
*E-mail:* jlt7@york.ac.uk

We have collected student responses to questions designed to establish their understanding of twelve terms used regularly when working with error and uncertainty in quantitative data. The terms are 'reproducible', 'precise', 'accurate', 'sensitive', 'random error', 'systematic error', 'negligible', 'significant difference', 'qualitative procedure', 'quantitative procedure', 'correlated data', and 'transforming data'. In most cases less than 50% of the sample of first year chemistry students provided evidence of 'some or good understanding'. We suggest that their misconceptions are most likely to be rectified by persistently challenging the students to make explicit use of key words and concepts such as these whenever they present reports of quantitative data collected during practical work.

## Introduction

One of the skills that chemistry graduates are expected to acquire is the *"ability to interpret data derived from laboratory observations and measurements in terms of their significance and the theory underlying them"*.[1] Effective data interpretation involves dealing with errors and uncertainty in the measurement of physical quantities. This is an area that requires a particular use of language. Even the word 'error' is a source of confusion to many students since they commonly regard 'errors' as personal mistakes[2] rather than recognising that *"every physical measurement is subject to a host of uncertainties that lead to a scatter of results"*.[3] Another simple example is that many dictionaries give 'accuracy' as one meaning of 'precision'[4] whereas these two words have distinctly different meanings in the context of scientific measurement.

We recently published evidence that first year chemistry students have not developed an effective use of the language used to handle error and uncertainty.[5] The majority of a sample of 65 students believed that the **accuracy** of their results would be improved by using an objective rather than a subjective method to fit a line to their experimental data, and that rather more than half of them showed confusion over the meaning of the term 95% confidence limits. In retrospect, it seemed to us that these misconceptions were easy to reconcile with their previous experiences; they are accustomed to teachers looking for a high level of 'accuracy' and consequently they readily assume that a 'line of best fit' means that the line gives the most accurate result; their use of 'confidence' in every day language is sufficiently different from its technical usage in error analysis that this can easily be a source of confusion.

The constructivist view of learning[6] leads to the expectation that it is not easy to bring about a reconstruction of a misunderstood concept already embedded in the mind.[7] This may help to explain the conclusion that *"students find it difficult to grasp the value and purpose of statistical procedures"*.[8] We therefore decided to explore student understanding of key words and concepts used in dealing with error and uncertainty in measurement. We hoped this would lead to better understanding of the concepts held by first year students, and that this would help us to devise opportunities for learning that would lead to better understanding of those issues generally considered to be important. We had the opportunity to include a set of questions as part of a first year laboratory course on Analysis taken by the same cohort of chemistry students who, in the previous term, alerted us to the possible problem.[5] At this stage they had received virtually no relevant instruction in this area, at least not since they left school. Our intention was therefore not to evaluate the course itself, but to gain a better understanding of the concepts related to errors and uncertainty, which these students brought to the course. We expected that this would help us to devise opportunities for learning that would address specifically any widely held misconceptions. We report here on the responses received to this set of questions.

## Methods

In devising our questions we first attempted to obtain (through discussion with a number of concerned colleagues) a consensus view on the vocabulary with which first year students are expected to be familiar. Some of these (such as 'accuracy', 'precision', 'systematic and random error') are routinely defined or described in many standard treatments of error (e.g. ref.-s 3 and 12), others (such as 'sensitivity', 'negligible') are rarely dealt with in such texts though they are routinely used by scientists and have a special meaning in context.

We attempted to word the questions in a way that would distinguish between ability to provide a definition (declarative knowledge) and an ability to use words and concepts (procedural knowledge). Several drafts were required before we were satisfied with the set of questions. An important requirement was that we could ourselves prepare, for each question, a short answer that we regarded as demonstrating an adequate understanding of word usage in the context of analysis and error. This requirement caused us to make significant revisions to the wording of our first draft.

Figure 1 shows the final version of the set of five questions; these cover 12 key words or concepts (counting ‘significant difference’ and ‘insignificant difference’ as a single concept). Also shown is the explanatory preamble which draws attention both to the possible differences between the technical and every day meaning of some words, and to the fact that we were looking for each student’s view of how to **use** the words, given that the meaning may vary with the context. The whole fits conveniently on to one side of a sheet of A4 paper.

The students in our survey were in the second term of their chemistry degree course. Each of the 103 students of the first year cohort received their own copy of the question sheet at the beginning of the six-week course on Analysis that started in the first week of the spring term of 2000. The question sheets were handed out by the Course Organiser, who gave a short verbal explanation to reinforce the points made in the preamble. No responses were received before the end

**Questions set to first year chemistry students**

**The Language of Analysis and Error**

Analysis usually involves measuring quantitatively or qualitatively one or more constituent of something. In order to communicate analytical results (including information about the effect of experimental error on their reliability), chemists use words which have technical, or specialist meanings. Many of these words are also in ordinary use; for example *accuracy, precision, random, systematic, significant.* The purpose of this exercise is to give you an opportunity to think about and explain or describe how **you** would use, in a scientific context, some words which have both a technical and a general meaning.

There are five questions for you to answer. Write your answers clearly and unambiguously so that your reader knows exactly what you think. Remember that the questions are asking what **you** think; they are not asking for the ‘correct’ answer; (in a sense there is **no** single correct answer, since the meaning varies with the context). Later in the course, when your answers have been handed in, you will be provided with our answers, so that you know how **we** think the words should be used. You should compare your answers with ours, and reflect on any differences.

Be concise. Each question should be answered in a few lines.

**Questions**

1. An analytical procedure needs to be *reproducible, precise, accurate, sensitive*.
How would you investigate how well a procedure meets these criteria?

2. Explain why *systematic error* is harder to detect than *random error.*
3. Under what circumstances would you describe an error as *negligible*, and a difference between two values as *significant* or *insignificant*?

4. Can a *qualitative* procedure prove that a constituent is absent from a substance? And can a *quantitative* method be used to determine exactly how much of a constituent is present?

5. How would you decide whether data are *correlated* and when would you consider *transforming* data?

Notes
* in order to be *correlated,* data must have an x value and a y value; the question here is: under what circumstances are the two values correlated with each other?
* if you measure something (e.g. temperature), you may sometimes wish to *transform* it (for example to 1/T)

*You may use examples to illustrate your answers, if you find this appropriate.*

**Figure 1**

of the course and any handed in before the end of term were accepted for evaluation. We do not think the responses were affected either by the content of the course or by the four-lecture course on Analytical Procedures which coincided with the beginning of the laboratory course, since neither were designed to deal with the kinds of question we asked in the questionnaire. In order to encourage students to provide answers that reflected their current understanding, and to discourage them from seeking textbook answers, we made it clear that the exercise was voluntary and that answers would not contribute towards the mark for their laboratory work.

We did not attempt to analyse the overall response of each individual since our intention was not to attempt to map individual understanding of errors but to gain an overview of the sorts of ideas and misconceptions that students in general have about errors. We were concerned not only to discern how much understanding the students have, but also the nature of any misunderstanding. Accordingly, we evaluated how well each response answered the question and we also looked for answers that demonstrated some understanding of the issues even though the wording was more indicative of declarative knowledge than of procedural understanding. Before attempting to evaluate the student responses, we drew up a table giving a short (one sentence) acceptable answer to questions about each of the thirteen words or concepts. This defined the key points we looked for and helped us to be consistent in our evaluation.

## Results

A total of 33 responses were received. This rather low response rate (32%) is almost certainly a consequence of the explicitly voluntary nature of the exercise. However, the average A-level score of the respondents was 22 points (three subjects, excluding General Studies, equivalent to BBC) compared with 24 points (equivalent to BBB) for the whole cohort. Thus, on the basis of the only criterion available, the respondents are a reasonable cross section of the whole cohort.

We classified all the responses as showing either ‘some or good’ understanding, or ‘little or no’ understanding. Our attempts to attain greater precision, for example by classifying under four rather than two headings, were unsatisfactory. It therefore seemed better to present a crude summary followed by a more detailed discussion of the student responses to each of the five questions in turn. Figure 2 shows the summary of our findings. Even though we were generous in attributing ‘some understanding’ to some responses, less than 50% of the respondents were judged to show ‘some or good understanding’ of most of the terms. In the discussion which follows we first enumerate and describe those responses which show ‘some or good understanding’, and then those which show ‘little or none’.

Summary of responses from 33 students

| Word or concept | Good or some understanding | Little or no understanding |
|---|---|---|
| Reproducible | 18 | 15 |
| Precise | 13 | 20 |
| Accuracy | 14 | 19 |
| Sensitive | 6 | 27 |
| Random error | 13 | 20 |
| Systematic error | 26 | 7 |
| Negligible | 16 | 17 |
| Significant difference | 20 | 13 |
| Qualitative | 6 | 27 |
| Quantitative | 7 | 26 |
| Correlated | 31 | 2 |
| Transforming data | 31 | 2 |

**Figure 2**

**Question 1 An analytical procedure needs to be *reproducible, precise, accurate and sensitive.* How would you investigate how well a procedure meets these criteria?**

Specimen answer:

- Reproducible: Make multiple measurements on same sample using same procedure.
- Precise: Determine by observation (of replicate measurements) how many significant figures are justifiable.
- Accurate: Use procedure on a standard sample to check closeness to correct value.
- Sensitive: Use decreasing quantities or concentrations until signal cannot be distinguished from noise.

The question of investigating **reproducibility** was significantly better answered than the other three concepts; eighteen of the respondents based their answers on **replicate measurements,** thus showing ‘good understanding’. Five appeared to have confused the reproducibility of results with the opportunity to repeat an experiment (obviously a prerequisite for determining reproducibility, and one which must surely be understood to apply to any analytical procedure). A typical example of this misconception is “*For an experiment to be reproducible it should be easy and affordable to recreate the experiment and experimental conditions*”. The remaining ten students (30% of the sample) are judged to have no useful understanding of any of the four terms dealt with in question 1 because they either made no attempt to differentiate between them or they dealt specifically with reproducibility but did not distinguish between precision, accuracy, and sensitivity. Examples of these responses are “*Repeat the procedure a number of times to see if there is a large error in the accuracy etc. in which case the results would be significantly different*” and “*In order for an experiment to be reproducible all measurable* *external stimuli should be measured and taken into account. Precision, accuracy and sensitivity can be obtained using a suitable instrumentation giving an acceptable degree of accuracy and a method of measurement which reacts quickly enough to observe significant changes."*

In describing how to investigate the **precision** of a measurement, only seven respondents gave some indication that the key indicator is the number of significant figures that can be justified. A lower level of understanding was demonstrated by six students who made some reference to repeating the procedure several times to obtain the precision (a point already made by three of them in connection with reproducibility); the weakness of these answers was that none gave any indication of how they would judge the precision, or indeed that they understood its meaning in the context of analysis. In addition to the ten already identified as showing no understanding, a further nine showed little understanding, and one gave no response to this part of the question. Four of the nine suggested that precision is related to the care taken with experimental procedures; undoubtedly this is in a sense correct, but it is not a response which engenders confidence that these respondents have a clear understanding of the meaning of precision in this context and it may be related to an assumption that variation is the result of their mistakes.[2] Two confused precision and accuracy and three respondents gave answers, which defied any attempt at classification.

Nine responses on **accuracy** showed good understanding by referring to the use of known or standard samples and comparing experimental results with these. Within this group of nine, four were clearly referring to investigating the accuracy of the procedure using some specific standard sample independently of making an experimental measurement (one of these also offered as an alternative the possibility of making the same measurement using a variety of procedures) and five referred to comparing their results with a literature or text book value (thus illustrating that they are thinking only in terms of analytical exercises to which the answer is already known). Five showed a lower level of understanding by making simple statements about 'closeness to the correct or true value' without giving any indication of how the true value might be known. Nine respondents showed two different sorts of misconception. Five referred to the need to take care either with the procedure or with the choice of equipment, but showed no awareness of the need for calibration or standardisation. The other four confused accuracy with precision (one of them explicitly stating that both accuracy and precision are determined *"from the range of values over which repeats lie"*). Ten showed no understanding, as described above.

Six respondents showed that they understood **sensitivity** to be concerned with the ability of an analytical procedure to detect small quantities or low concentrations of the analyte, though (perhaps not surprisingly) none of these referred to the signal-to-noise ratio as a criterion for judging detection limits. The remaining twenty-seven either showed misconceptions (fifteen) or showed no useful understanding (ten referred to above) or failed to provide an answer (two). Eleven of the fifteen responses with identifiable misconceptions were concerned with the ability of a procedure to detect small **differences** or with the effect on the result obtained of **small changes** in conditions. This is a common meaning of sensitivity in everyday language; however our view is that, in the context of chemical analysis, it is the **precision** of a procedure that determines whether small differences can be detected, and that **sensitivity** is properly reserved to refer to the lower limit of detection. Four responses showed neither useful understanding nor any clear misconception (examples are "*ensuring that all likely changes are measured*" and "*if the reactant is not sensitive to a test, the results will be hard to obtain*").

**Question 2. Explain why *systematic error* is harder to detect than *random error.***

Specimen answer:
Random error: Easy to detect from variability of replicate measurements.
Systematic error: Difficult to detect unless you have a reason for supposing the result is incorrect.

The majority of students' responses (twenty-six) showed an understanding that, when systematic errors are present, they occur in all measurements; twenty-one of these linked this with the difficulty in detecting systematic error. Six respondents submitted statements that could not be interpreted as showing any understanding of the use of either 'random' or 'systematic' error. One response included no reference to systematic error.

The responses showed a much lower appreciation of the meaning of random error. Thus only five respondents (a mere 15% of the sample) made clear statements about random error causing variation in readings. An example of these responses is *"random errors are generated by the finite precision of measurement. They affect different readings by different amounts"*. A further eight could (with more generosity) be interpreted as showing some understanding of the term. Four of these stated that random error can be removed by repetition as in *"random error can be worked out of the experiment by repeating it several times"*; these statements could be interpreted as showing that the respondents understand that replicate readings vary as a result of random error, and that the mean value is likely to approximate to the value that would be obtained in the absence of random error. The other four were more difficult to interpret, as in *" random error only occurs in one mode on one variable at one time, and its nature* ***must*** *change on repetition so its location and magnitude can easily be determined"*. Fourteen made the mistake of assuming that random error occurs in a single or a small number of results and one of these explicitly described random error as "*a human mistake*". An example of this style of response is "*random error will not affect all results so a result due to random error will look out of place"*. As stated above, the remaining six showed no understanding of the term.

**Question 3. Under what circumstances would you describe an error as *negligible*, and a difference between two values as *significant* or *insignificant*?**
Specimen answer:
Negligible: When the error is so small compared with the value of the measurement in question that it does not affect the final result (enough for you to care about).
Significant difference: When a statistical (mathematical, objective) test shows that there is only a small chance that the difference between two values arose by chance.

Sixteen responses showed understanding by making some comment to the effect that an error is negligible when it does not have a (noticeable) effect on the overall result (this includes two who related this to whether the overall result is "*close to the expected answer"*, thus drawing attention to the view of many students that analysis involves looking for an already known answer). Three of these sixteen specified a percentage error that would qualify as negligible, but most made little attempt to describe how they would make their judgement as in "*if it does not affect the final result too much"*. Eight other responses specified a percentage error that would be regarded as negligible, but gave no indication of why this was negligible and were therefore judged to be too simplistic to qualify as showing useful understanding. Interestingly the estimates of what might be considered negligible varied from "*several orders of magnitude less than the value"* to "*10%"*, with the most common suggestions lying around 1%. One respondent firmly stated that "*no error is negligible and should always be stated*" – a belief with which we have some sympathy and are inclined to applaud, but it is a very inflexible attitude to apply to the real world of experimentation. One suggested that an error is negligible if it only affects a small proportion of the results – reinforcing that some students regard errors as occasional events rather than as an inevitable feature of measurement. Three more used suspiciously similar wording to state that an error is negligible "*when the result is unaffected whether or not it is included*" (presumably thinking, as the previous response indicated, that the 'error' occurred in one measurement out of a number). The remaining four were so confused as to defy analysis.

In this question we linked the concepts of 'negligible' and 'significance' with the intention of drawing out the point that the latter is almost always concerned with a **difference** between two values (normally mean values), whereas the former (at least in this context) applies to the error in an **individual value**. In practice twenty respondents addressed the question of significant difference in a meaningful way but only five of these showed real understanding of this concept, as in the statement "*when a difference between two values can be explained by error, then it can be regarded as insignificant"*. Fourteen of them stated that a difference would be significant (or insignificant) if it was greater (or less) than a specified percentage of (one of) the values in question. The percentages suggested as a measure of a **significant difference** varied from 1% to 5% (except for one student who did not give a general rule of thumb, but gave as an example that *"the difference between 11032.06 and 11032.91 is insignificant, but that between 1.123 and 1.921 is significant"*). One of them simply stated that a difference is insignificant "*when it won't affect the results*"; we find it hard to assess the level of understanding that this represents. The weakness of all twenty of these answers, which we classified as showing some understanding, is that **none** of them showed any awareness that the amount of variation in (or precision of) data is crucial in deciding whether a difference is significant. Eleven respondents demonstrated that they had not grasped the key point by failing to refer to a difference, and eight of these specifically referred either to a significant **value** or to a significant **error**. One made no attempt to answer this part of the question, and another stated (as an example) that the arithmetic difference between an atomic number and a molar mass is insignificant *"because it has no meaning"* thus demonstrating at least some concern for sensible handling of units!

**Question 4. Can a *qualitative* procedure prove that a constituent is absent from a substance? And can a *quantitative* method be used to determine exactly how much of a constituent is present?**
Specimen answer:
Qualitative procedure: Can only prove that something is below a certain level (determined by the sensitivity of the method).
Quantitative procedure: Can only determine the quantity within the limits determined by the precision of the method.

Only three students recognised the limitation that very small amounts (below the sensitivity of the procedure) cannot be detected by qualitative methods, and that experimental error prevents exact measurement. In addition the first limitation was recognised by four and the by second three. This left twenty-three respondents who replied affirmatively to both questions. It is true that the wording of many responses might be regarded as ambiguous in a court of law; thus one of the respondents did not state explicitly that a qualitative method can **prove** the absence of a substance, and only nine explicitly stated that a quantitative method could determine **exactly** how much is present. However, wording such as *"a qualitative procedure is used to determine **what** is in a substance"*, or *" a quantitative test can be used to determine the amount of a substance"* does not lead us to believe that the respondent was trying to suggest that the determination is in doubt. Indeed we suggest that the reason why more students specified 'proof' for qualitative procedures but not 'exactly (how much)' for quantitative was that their sentence structure did not require them to add the word 'exactly' to their description of a quantitative procedure, whereas it was harder for them to avoid a word such as 'proof' when describing the testing of absence.

### Question 5. How would you decide whether data are *correlated* and when would you consider *transforming* data?

Specimen answer:

Correlated data: Examine a graph of x vs. y to see whether there is evidence of a relationship (correlation).

Transforming data: When the transformation converts a non-linear relationship into a linear one.

Twenty respondents indicated that correlations could be detected from a graph of the data, and a further six said that they would be regarded as correlated if two variables showed some kind of relationship (without specifying how they would detect the relationship). Five said that data are correlated when they fit a mathematical relationship. All of these show some understanding of correlation and only two did not answer this part of the question. Of the twenty who referred to graphical representation of the data, only eight actually answered the question as set by stating that their decision would involve inspecting a graph of the relevant data. The other twelve wrote their answer more in the form of a definition or a theoretical description of correlation. One respondent accepted data to be correlated "*when they show a pattern of increase or decrease or **constancy**"* (our emphasis) and four explicitly restricted the meaning of correlation to linear relationships, (though there is evidence that others shared this misunderstanding even though they did not make it explicit). The wording of the responses (especially the unwillingness to deal with the question "*how would you…*") gave the impression that most students were more familiar with the collection of data known to be correlated than with the concept of investigating whether data are correlated or not.

The question of transforming data gave a similar indication of the majority showing some understanding. Thirty-one referred in a wide variety of ways to the wish to show a relationship more clearly. Many made it clear that the main (or only) purpose was to create a linear relationship from a non-linear one. However the range of answers included a number that indicated more confused objectives almost certainly based on misconceptions. For example three respondents seem to believe that transformation can reveal a correlation which does not exist in the raw data (e.g. *"if there is no correlation, consider transforming the data to determine whether there is any correlation there*"); it may be that these students belong to the group who believe that correlation implies a linear relationship, and several other forms of words suggest that this is the case for a number of others. Many students appear to believe that correlated data can only be analysed when the relationship is linear (*"a plot of T vs. P is useless, but a plot of lnT vs. 1/P is useful"*, or *"data should be transformed to give meaningful graphical data"*), and this includes two who explicitly stated their assumption that the accuracy of their results will be improved by working with a linear relationship. As with the first part of this question, students were apparently reluctant to personalise their answers by answering *"when would you consider…"*. Only two respondents gave no answer to the question about deciding to transform data.

## Discussion

Our survey is based on a relatively small sample of students. Nevertheless we believe that our sample is sufficiently representative for us to draw useful conclusions. This confidence is based largely on our comparison of the A-level grades of the respondents and the cohort as a whole. For two of the concepts we tested we are also able to compare our data with the conclusions of Davidowitz et al.[9] These authors analysed the reasons given by 135 second year chemical engineering and science students for making repeat measurements. They concluded that 45% of their sample perceive the purpose to be either to identify a recurring (correct) value (20%) or to perfect measuring skills (25%). This is in broad agreement with our finding that students are more inclined to link variation in measurements with 'mistakes' than with random error (in the sense used in quantitative measurement). The same authors found that more than half of their sample of students, when asked to compare two sets of data, regarded the mean value as of much greater significance than the spread. This is also consistent with our finding that none of our sample of first year students made reference to the spread of data when discussing significant differences.

Our data indicate that first year chemistry students would benefit from a considerably better understanding of the language used to deal with error in quantitative measurement. We believe that this is a matter for concern since the handling of quantitative data is of crucial importance to the procedural understanding of science. We suggest that teaching in this area needs to be radically rethought and restructured because the problem is not simply one of impressing correct ideas about errors on to a blank sheet of a student's mind, but to reconstruct their misconceptions into mature understanding. The first step must be a careful analysis of the key concepts, which students need to understand. We do not claim that our list of five questions covers all of these key concepts. For example, with the benefit of hindsight, we recognise the value of directly probing the students' perception of the origin of variation in measurement. However our study illustrates the value of covering qualitative aspects of the use of language (such as 'negligible') as well as rigorously defined terms (such as 'random error'). Furthermore, the wording of our questions indicates the value of giving meanings and understanding in operational terms like "how would you investigate…" and "explain why…".

It seems unlikely that students will improve their understanding through textbooks, even if they could be motivated to read them, Our pessimism is based on our inability to find guidance to our questions 3 – 5 in most undergraduate texts. The Open University text on this subject[10] provides a rare example of dealing with the inappropriateness of judging the significance of differences between values by reference only to a mean value and of the benefit of preferring a graph to a table of data in order to discern a correlation between variables. Even for our questions 1 and 2 the guidance given is often contradictory. For example Skoog *et al.*[3] define 'precision' as *"the agreement between two or more measurements that have been carried out in exactly the same fashion"* thus suggesting that a procedure capable of determining a value to only two significant figures is very 'precise' because it lacks the precision needed to detect significant variation. In comparison Atkins[11] defines 'precise measurements' as having *"small random error"* making the cardinal mistake of failing to specify that the error must be small compared with the size of the measurement. Hanson *et al.*[12] state that *"the standard deviation is a measure of the precision of the measurement"*. They go on to use as an example a measure of the boiling point of water for which they quote a mean value of 400.00K and a standard deviation of 0.0126. In contrast the Open University[10] maintains that it is *"rather silly"* to quote a standard deviation to so many significant figures, and there certainly seems little justification for quoting a greater number of decimal points for the standard deviation than for the mean.

Given the lack of clarity and consistency in the textbooks, most chemistry students will necessarily rely on their course work for information about errors and their treatment. Meester and Maskill reported that most laboratory manuals for first year chemistry courses included some information about error analysis but concluded that *"although this indicates the great importance attached to it, generally speaking, error analysis was not a central feature of the courses"*.[13] We do not think the situation has changed significantly in the ten years since the survey was conducted. Furthermore, we have no reason to suppose that the sections on errors in the laboratory manuals are any more likely than the text book accounts to lead to effective learning; our scepticism is based on anecdotal evidence suggesting that there is no consensus amongst academics either about the correct usage of words and concepts used to describe uncertainty in data or in the best procedures available for interpreting experimental data. We are thus led to the conclusion that there is a need for much careful thought about the best ways to meet the Benchmark objective relating to data interpretation.[1] We do not believe that lecturing is likely to be effective because the misconceptions we have documented are unlikely to be corrected by an account (however authoritative) of received wisdom; such accounts rarely involve active participation of the students. Even workshop activities often do little more than introduce the students to routine exercises, which do not really engage their minds. The student perceptions need to be actively challenged in such a way that they reconstruct their own understanding.[7]

We suggest that these operational learning outcomes are most likely to be achieved by regularly and persistently challenging students, through their laboratory work, to discuss their data in terms of the desired outcomes. Thus they could be required to give evidence of the random error in their data, to indicate precautions they have taken to avoid systematic error, to comment on the comparability of data collected by different individuals (and whether differences are significant), and so on. Our proposal is that these regular challenges should be fully integrated into the laboratory course; there is little advantage in paying lip service to the idea by requiring a bolt-on statement about error at the end of a laboratory report. The familiarity with the terms and concepts gained through such regular usage is likely to lead the students to revise their own understanding through a deep learning process. We do not doubt that such learning could be usefully reinforced by structured class discussions and interactive workshops or even by well-constructed lectures, but we suggest that the primary route for learning should be frequent and explicit challenges to use the relevant words and concepts.

## Acknowledgements

We acknowledge financial support from the Royal Society for a University Research Fellowship to PJD, the LTSN Physical Sciences Centre for a project grant to JLT, and Glaxo Wellcome plc for support for CJG.

**References**

---

1 *General guidelines for the academic review of Bachelors Honours Degree Courses in Chemistry*, Quality Assurance Agency, Gloucester, 1998.

2 M.J. Goedhart and A.H. Verdonk, *J.Chem.Ed.*, 1991, **68,** 1005.

3 D.A. Skoog, D.M. West and F.J. Holler, *Fundamentals of Analytical Chemistry,* 5th ed., Saunders College Publishing, Fort Worth, 1996.

4 J.M. Hawkins, J. Weston and J.C. Swannell, *The Oxford Study Dictionary,* Oxford University Press, Oxford, 1992.

5 J. Garratt, A. Horn and J. Tomlinson, *U.Chem.Ed.*, 2000, **4,** 54.

6 G.M. Bodner, *J.Chem.Ed.*, 1986, **63,** 873.

7 K.S. Taber, *U.Chem.Ed.*, 2000, **4,** 63.

8 J. Garratt, D. Clow, A. Hodgson and J. Tomlinson, *Chemistry Education Review*, 1999, **14,** 51.

9 B. Davidowitz, F. Lubben and M. Rollnick, *J.Chem.Ed.,* 2001, **78**, 247.

10 The science foundation team, *The handling of experimental data,* Open University Press, Milton Keynes, 1982.

11 P.W. Atkins and J.A. Beran, *General Chemistry,* W H Freeman, New York, 1992.

12 J.R. Hanson, J.I. Hoppe and W.H. Prichard, *A guide to the reporting of practical work and projects in chemistry,* Royal Society of Chemistry, London, 1995.

13 M.A.M. Meester and R. Maskill, *First year practical classes in undergraduate chemistry courses in England and Wales*, Royal Society of Chemistry, London, 1993.