SUPPORTING INFORMATION

Measuring & Reducing Chemical Spills by Students: a Randomised Controlled Trial of Providing Feedback

Aimilia M. Tsokou, Alix Howells, and Moray S. Stark*

Department of Chemistry, University of York, Heslington, York, YO10 5DD, UK

Hodges Lehmann Estimator analysis and histograms for *volume* and *log₁₀(volume)* differences

Two approaches were used in determining the Hodges-Lehmann estimator in the study; in one the analysis was performed using absolute differences of volume spilled between control and intervention groups, as described by equation 1, where the volume spilled by the N=67 students receiving the intervention is $V_{n(intervention)}$, and the volume spilled by the M=76 control students by $V_{m(control)}$, with the Hodges Lehmann Estimator for the difference in volumes expressed as $\overline{\Delta_{HL}(V)}$.

$$\overline{\Delta_{HL}(V)} = \text{median of differences} \begin{cases} m = 1,M \\ n = 1,N \end{cases} (V_{n(intervention)} - V_{m(control)}) \qquad \text{equation 1}$$

The second analysis was performed using differences of the $\log_{10}$(volume), as described by equation 2, where and the Hodges Lehmann Estimator for the difference in log of the volumes is expressed as $\overline{\Delta_{HL}(log_{10}V)}$.

$$\overline{\Delta_{HL}(log_{10}V)} = \text{median of differences} \begin{cases} m = 1,M \\ n = 1,N \end{cases} [\log_{10}(V_{n(intervention)}) - \log_{10}(V_{m(control)})] \qquad \text{equation 2}$$

Both volume and log volume analyses are presented where there is a noticeable difference between the two. It is worth noting that the analysis performed using differences of the $\log_{10}$(volume) spilled is equivalent to the ratio of the amount spilled by the control group over the intervention group, so is easily expressible as a percentage change in amount spilled by giving the feedback intervention, as indicated by equation 3.

$$\overline{\Delta_{HL}(log_{10}V)} = \text{median of set of differences} \begin{cases} m = 1,M \\ n = 1,N \end{cases} \left[\log_{10}\left(\frac{V_{n(intervention)}}{V_{m(control)}}\right)\right] \qquad \text{equation 3}$$

Figure 6 gives a histogram for the 5092 (= 67 x 76) differences in absolute volumes, which gives a non-symmetric, far from bell-curve distribution as the data are highly skewed.

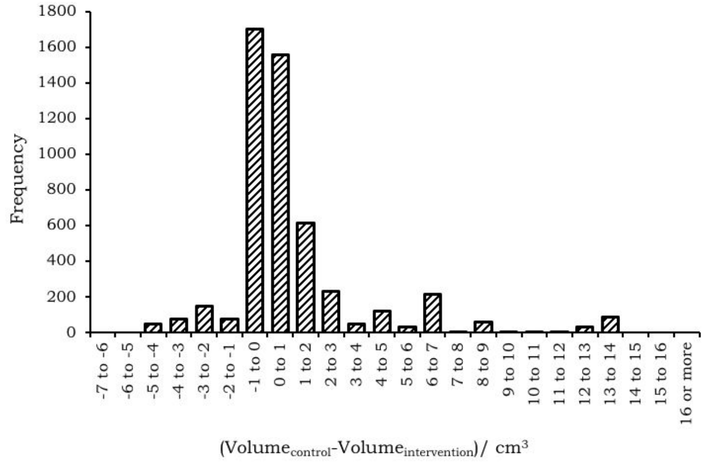


Figure 6. Difference of control and intervention group to obtain the Hodges-Lehmann median and limits

Figure 7 gives the corresponding histogram for the differences in the $\log_{10}$ of the volumes spilled, which by contrast is more symmetrical bell-shaped distribution. The Hodges-Lehmann estimator assumes that the distribution of the differences approximates to the normal distribution, so the log transformed data represented in figure 7 has higher validity and is given preference in the statistical analysis.

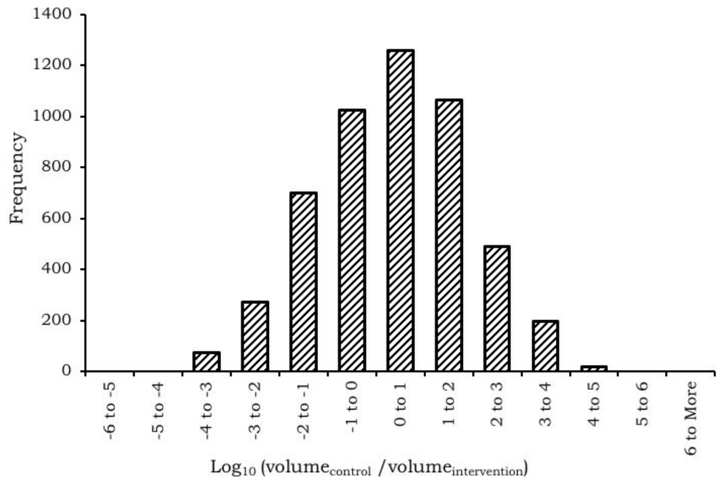


Figure 7. Logged difference of control and intervention group to obtain the Hodges-Lehmann median and limits

In these data sets there are a significant number of zero entries (363), so these are split between the *-1 to 0* and *0 to* 1 bins in the histograms given in figures 6 and 7.

The Hodges-Lehmann estimator for the median difference of the logged volumes is –0.301, which is equivalent to those receiving the intervention spilling approximately half the volume spilled by the control students, a 50% decrease in chemical spillage due to receiving feedback (equation 4). The 95 % confidence interval for the Hodges-Lehmann estimator of the median difference for the logged data was 0 to -0.699, which is equivalent to those receiving the intervention spilling between the same volume spilled by the control students and 0.2 of that spilled by the control students, which alternatively represents a decrease in chemical spillage due to receiving feedback of between zero and a factor of 5.

$$10^{\overline{\Delta_{HL}(log_{10}V)}} = 10^{-0.301} = 0.50 \quad \begin{cases} \text{upper 95 \% confidence limit} = 10^{-0.699} = 0.20 \\ \text{lower 95 \% confidence limit} = 10^{0} = 1.0. \end{cases} \qquad \text{equation 4}$$

Feedback tool for demonstrators on chemical spills by students

> This spillage feedback tool for demonstrators was lodged with a colleague at the University of York not otherwise involved in the project in advance of the start of data collection and is available directly from them at julia.sarju@york.ac.uk.

For experiment C06, the students will measure the concentration of two copper compounds, one in the morning and one in the afternoon. As an addition to C06 this year, demonstrators are asked to:

- Measure the area of chemicals spilled onto a paper liner by year 1 undergraduate students carrying out this experiment
- Give feedback to the students on what the safety consequences would be if this amount of chemicals were routinely spilled
- For each student and for both morning and afternoon sessions, record onto the paper liner the:
    - Student's name and email (prefilled)
    - Student's group (prefilled)
    - AM or PM (please circle)
    - The copper compound used by the students (please circle)
    - The measured concentrations for the copper compound
    - Area of chemicals spilled (number of squares on grid covered by spilled chemicals)
    - Whether feedback on spillage of chemicals was provided

Measuring the volume of chemicals spilled

Students will carry out their handling of chemicals for this experiment on a paper liner. The copper compounds they use are sufficiently coloured to show up as a stain (the low concentrations are not very distinct but please count every spill that is visible to you). You will be given an A3 transparent plastic sheet with a 3cm x 3cm grid drawn on it. Place this over the spilled areas and count the squares.

For minor spills, four of the squares of the 3x3 grid are sub-divided into quarters (marked in red or green), so use these to measure small spillage areas (lower than *four* 3x3cm squares) more accurately. Please indicate that you have used the smaller grid by recording the number of quarter squares counted as "*X/4*". To count the spill area, first line up the top and one side of a major area of spill with horizontal and vertical grid lines (instead of centring it). If a square only partially contains spilled chemical, still count it as a whole square.

Record the number of squares of spillage on the sticky label provided, attaching to the paper liners in a corner, then use this number to choose which feedback description to give to the student.

Feedback to students

The areas of spill you measure have been translated into a short description to provide to the student; choose the appropriate phrase from the table below to use for each student.

Please note, only approximately 40 % of students are provided feedback for the morning sessions, while all students should be provided feedback for the afternoon's experiments (see below). The feedback should be provided promptly, in the lab, as soon as the student indicates they have completed the experiment. For the morning experiment, for those students receiving the feedback, this should be given *before* they leave for lunch after the morning's experiment. The list of students receiving feedback or not receiving feedback is provided.

These threshold areas have been chosen by establishing a safe level of spillage for a number of common chemicals with a wide range of toxicities, from high to low: potassium cyanide; hexane; ethyl acetate; & ethanol. A descriptive grade, from *A* to *E*, can also be given to the student, so they have a more easily understandable description of their performance at handling chemicals safely.

During your morning briefing please tell the students that the paper is there to just absorb any spills and that they shouldn't try to clean up any chemical they spill on the paper. Please also let them know that the experiment is not assessed and the amount of chemicals they spilled does not affect any mark.

Feedback comments to give to students

No. of (3 x 3cm) squares with chemical spilled	**Feedback** ***"If you spilled this volume of chemicals routinely, then you would:"***	**Descriptive Grade**
<2	*"be able to handle high hazard chemicals safely, such as 1M potassium cyanide"*	A
2-10	*"be able to handle higher hazard chemicals, such as hexane, safely but not more hazardous chemicals, such as cyanides"*	B
10-30	*"be able to handle routinely hazardous chemicals safely, such as ethyl acetate, but not more hazardous chemicals, such as hexane or cyanides"*	C
30-150	*"be able to handle lower hazard chemicals safely, such as ethanol, but not more hazardous chemicals"*	D
>150	*"not be able to handle low hazard chemicals safely, such as ethanol"*	E

Guidance to Demonstrators on Randomised Controlled Trial of Effect of Giving Feedback to Students

> This guidance for demonstrators was lodged with a colleague at the University of York not otherwise involved in the project in advance of the start of data collection and is available directly from them at julia.sarju@york.ac.uk.

The students will measure the concentration for two copper compounds, one in the morning and one in the afternoon. Please measure the area of chemicals spilled for each compound separately – the paper liner will be changed between morning and afternoon session – if you have time at the end of the morning session please assist the lab staff to put out fresh paper (absorbent side up, shiny side down).

The students are allocated to one of two groups, an intervention group, and a control group, with approximate half in each.

For the morning session, only the intervention group will be provided with feedback by you; the control group will receive no feedback in the morning session.

For the afternoon session, all the students will receive feedback. The demonstrator for a student during the afternoon session should be different to that from the morning session – during the afternoon session please do not seek to find out if a student has received feedback or not during the morning session.

Please record all details requested above on each sheet, for both groups, for both morning and afternoon sessions, this includes the chemical used and the concentrations measured. The information you record will allow the effect to be established of providing feedback on the students' chemical handling skills during the morning experiments on their chemical handling skills during the afternoon experiments.

Details of year 1 experiment to develop volumetric handling skills

The following details are given to year 1 chemistry undergraduate students to aid them in carrying out this day-long experiment designed to help them develop their volumetric handling and UV-vis spectroscopy skills, along with a briefing from demonstrators at the beginning of the lab.

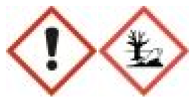
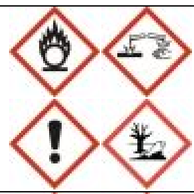
Volumetric solutions and UV-Vis spectroscopy

Introduction

Being able to prepare solutions which are of known concentrations is an important skill for use in quantitative work. The most common way to achieve this is through the preparation of volumetric solutions and accurate dilutions using volumetric glassware. This glassware is manufactured and labelled with its specific precision, which then allows you to accurately prepare solutions with a known precision.

In this experiment you will use volumetric glassware to determine the molar absorption coefficient (ε) for two different copper salts (one in the morning and one in the afternoon), and use this to find out the concentration of some solutions of the same copper salts at unknown concentration.

Risk assessment

Copper(II) sulfate pentahydrate	H302 Harmful if swallowed. H315 Causes skin irritation. H319 Causes serious eye irritation. H410 Very toxic to aquatic life with long lasting effects.	
Copper(II) chloride dihydrate	H290 May be corrosive to metals. H302 + H312 Harmful if swallowed or in contact with skin. H315 Causes skin irritation. H318 Causes serious eye damage. H410 Very toxic to aquatic life with long lasting effects.	
Copper(II) nitrate trihydrate	H272 May intensify fire; oxidiser. H302 Harmful if swallowed. H315 Causes skin irritation. H318 Causes serious eye damage. H400 Very toxic to aquatic life.	
Copper(II) acetate monohydrate	H302 Harmful if swallowed. H314 Causes severe skin burns and eye damage. H410 Very toxic to aquatic life with long lasting effects.	
Physical hazards	Take particular care not to use too much force when inserting pipettes into pipette bulbs. These only need to be inserted far enough to form a seal, and should be easy to remove.	
Precautions	Wear lab coat and safety specs at all times. Wear gloves while handling chemicals. Weighing should be conducted on the balances provided. Instrumentation work should be carried out in the instrument room. All other aspects of the procedure should be carried out in the fume hood	

Reagent quantities

Compound	Formula	RMM / g mol^{-1}	Approx. concentration of stock solution / M
Copper(II) sulfate	$CuSO_4{\cdot}5H_2O$	249.69	0.1
Copper(II) chloride	$CuCl_2{\cdot}2H_2O$	170.48	0.1
Copper(II) nitrate	$Cu(NO_3)_2{\cdot}2½(H_2O)$	232.59	0.1
Copper(II) acetate	$Cu(CO_2CH_3)_2{\cdot}H_2O$	199.65	0.05

Procedure

Prepare 50 mL of a stock solution at the approximate concentration shown for one copper compound and determine the wavelength of local maximum absorbance (λ_{max}) using a scanning UV-Vis spectrometer (1000-350 nm).

Perform serial dilutions of your stock solution to give the concentrations shown in the table below.

Accurate concentration /mol dm^{-3}	
$SO_4/Cl_2/NO_3$	$(OAc)_2$
0.05	0.025
0.02	0.01
0.01	0.005
0.005	0.0025

Record the absorbance at λ_{max} for your stock solution and each of the dilutions using a single wavelength spectrometer.

Determination of ε

Plot a graph of absorbance versus concentration and determine the molar absorption coefficient (ε) in units of dm^3 mol^{-1} cm^{-1}.

Determining the concentration of an unknown

Determine the concentration for one of the solutions of your salt which is at an unknown concentration. You will need to record the absorbance at your determined λ_{max}, ensuring that the value is on scale. This may require diluting a sample of the solution of unknown concentration.

Protocol for using an RCT to Assess the Effect of Feedback on Safe Chemical Handling Skills in Practicals

> This trial protocol, including details on consent, was lodged with a colleague at the University of York not otherwise involved in the project in advance of the start of data collection and is available directly from them at julia.sarju@york.ac.uk.

Investigators

Principal Investigator & Guarantor:	Moray Stark
MChem Project Student:	Aimilia Tsokou

Timelines for RCT

- Date of Ethics Approval: 8/10/18
- Dates of Consent being sought: 9/10/18 – 15/10/18
- Date of archiving Protocol for RCT: 19/10/18

- Date of start of RCT: 22/10/18
- Date of end of RCT: 6/11/18
- Planned date for report write up: May 2019
- Planned date for submission for publication of RCT: May 2019

Aims

The aims of the project are:

- To develop and implement a tool for demonstrators for providing formative feedback to students on their skills at handling chemicals safely
- To test the effectiveness of providing this feedback in improving (or otherwise) the students' skills at handling chemicals safely in a randomised controlled trial

Background

This project is a development of a project in 2017-18, which examined the volumes of chemicals spilled by students during a chemistry practical, and also how this correlated with other lab skills, described previously under the ethics application dated 9/11/17 (see *Addendum to App Form Stark300916 Chemistry Practicals Investigations 9 11 17.pdf*).

For this project, a tool for lab demonstrators to provide formative feedback to students will be developed, and then tested in a randomised controlled trial to determine whether it has a measurable effect on the students' subsequent spillage of chemicals.

The chemicals used in the experiments described here are of low hazard in comparison with the higher hazard chemicals the students will use later in their studies and subsequent careers; the overall aim of this project is to assess whether feedback helps the students to improve their safe handling of hazardous chemicals, while they are handling less hazardous chemicals.

A preliminary literature search indicates that there are no previous RCT investigations into the effectiveness of methods of teaching in undergraduate chemistry laboratory practicals, and this trial has the capability to provide novel information on this important topic.

Ethics Approval for RCT

Ethics approval for this RCT is sought in advance from the Faculty of Physical Sciences of the University of York

- Application file name: Ethics App Form Stark Chemistry Practicals RCT 17 9 18 update.pdf
- Approval date: 8/10/18

Outline of Procedure

- Students will carry out experiments as part of a year 1 undergraduate practical course designed to develop lab skills. These experiments have been run for a number of years, and have not been developed for this investigation.
- During two of these experiments, the volume of chemicals spilled will be measured by lab demonstrators, using an approach developed for this project.
- Formative feedback will be provided to ca. half the (consenting) students after the first experiment, (random allocation of students to groups).
- The assessment of volume of chemical spillage during the second experiment allows an examination of whether the provision of the feedback has an effect on the handling skills of students.
- All students receive feedback after the second experiment.

Recruitment, Cohort & Timing

All year 1 undergraduate chemistry students in the Autumn 2018 cohort will be invited to participate in the RCT. Those giving consent will be recruited.

- Year 1 undergraduate chemistry students at the University of York (185) will carry out an experiment where the amount of chemicals they spill will be recorded.
- The experiments for this RCT will be carried during year 1 practicals during weeks 5, 6 and 7 of the Autumn term 2018 (22/10/18 - 6/11/18)
- All of the cohort are invited to take part in the trial, and only those giving consent are included allocated to a control or intervention group, and are included in the data analysis for the trial.

Consent

Consent of the year 1 Chemistry undergraduate students will be sought for permission to be randomly allocated to a control and intervention group, which receive feedback at different times of the same day. The following statement and question are included in a pre-existing online test the students complete as part of their course at the beginning of term. They are given the option of not completing the self-assessment question at no penalty.

> ***Initial question on practical abilities***
>
> > *Before attempting the questions of the chemical safety quiz, please consider an initial question that is asked as part of an academic study which has the overall aim of understanding and improving practical abilities in science students.*

You do not have to answer the question if you do not wish, and answering or not answering this question will have no effect on any practical mark.

By answering this question, you agree to take part later in the term in an academic study of the effect of providing feedback on performance in the lab, and to be randomly allocated to a group receiving feedback at different times of the day.

Your answer and selected experimental results from year 1 Teaching Labs will be used in an anonymised and aggregated manner as part of this academic study.

Please estimate your ability in the subject of practical chemistry, relative to the other students in your year by clicking on a percentile ranking

(for example, click on 0-10 if you think you are in the lowest 10 % of your year at practical work, or 90-100 if you think you are in the highest 10 % of your year at practical work)

Study design

Type of trial
This is a randomised controlled trial (specifically, a pragmatic, waiting list, parallel group, randomised controlled trial) of providing formative feedback to students on the volume of chemicals they spill during an experiment.

It is a pragmatic study, as no factors will be controlled during this trial, the feedback will be provided by and to the type of people (demonstrators & students) and in a manner that would be used in practice in a normal teaching lab setting.

Spillage will be measured after the first and second experiments for both groups. One group will be provided formative feedback after the first experiment, and the remainder after the second experiment. This is therefore a waiting list trial, as both intervention and control group will receive feedback, with the control group being delayed by approx. half a day, to reduce any perceived unfairness through providing a (slightly) differing educational experience.

It is a parallel group study as both the control and intervention groups carry out this experiment concurrently.

Randomisation
Randomisation of students into control and intervention groups will be carried out after consent has been given (carried out on 16/10/18).

The students who carry out the experiment involved in this trial are allocated by the practical organisers to 12 groups, two of which carry out the experiment used for this trial on the same day.

After removing those who did not give consent, the class list for each group are allocated a random number between 0 and 1 using the Excel function, =RAND(), and the student within each group are then ranked by the random number, lowest to highest. The lowest half (below the median random number) will be allocated to the intervention group, and the highest half will be allocated to the control group.

If a practical group contains an odd number of people and hence a larger number in either the control or intervention group, then subsequent practicals groups to contain an odd number will have the threshold adjusted to ensure that the extra person is in the intervention group, if the previous practical group had one more in the control group, and visa versa (with each subsequent group containing an odd number will alternate between the extra person being allocated to the intervention group and control group).

To test the effect (if any) of providing formative feedback, the volume of chemicals spilled during the second experiment by students in the intervention (feedback) group will be compared with the amount spilled during the second experiment by the control group. This allows an objective test of the efficacy of providing formative feedback on the chemical handling skills of the students, and whether this can lead to a (statistically significant) change in chemical handling skills.

Treatment of "Drop out"
If spills data is not collected for any reason for individual students (e.g. missing labs due to illness, or result not recorded) giving consent then this dropout number will be reported for each group, and then they are ignored in subsequent data analysis.

Reporting results of trial

Reporting standard
The trial will be designed for the results to be reportable consistent with the CONSORT standard (Consolidated Standards of Reporting Trials, Schulz et al, 2010), and it is intended to prepare the trial for publication in an appropriate peer-reviewed primary research journal.

Graphical representation of data
Data for the control and intervention groups, and the combined control+intervention group (indicated with different symbols in the combined figure), will be reported graphically. At a minimum:

- Volume of chemical spill by each student, ranked lowest to highest, with logarithmic vertical axis
- Volume of chemical spilled, normalised to the total spilled.
- Integrated (cumulative) volume spilled by each student ranked highest to lowest
- Integrated (cumulative) volume spilled, normalised on both vertical and horizontal axes
- Volume spilled by the two groups, ranked separately, on the same graph using separate symbols for each group

The data presented will available either from the PI, or in journal supplemental information.

Primary Descriptive statistics
The primary summary, descriptive statistic reported will be the Hodges–Lehmann estimator (Hodges and Lehmann, 1963) a measure of the median difference between the two groups.

Primary Inferential statistics
The primary inferential statistic reported will use the Mann-Whitney U test (Mann & Whitney, 1947) a non-parametric, two-sided test of significance for independent observations to examine whether providing the feedback has a statistically significant effect. This non-parametric approach is used here as the distribution of spilled volumes is far from being Gaussian (it is approximately exponential).

The null hypothesis investigated is that providing feedback has no effect, and the alternative hypothesis is that it has an effect. The Mann-Whitney U test is typically used for ordinal outputs, so is used here with rankings of volume spilled by the students. This approach also allows for confidence intervals to be evaluated (e.g. Bonett & Price, 2002). The pre-specified confidence limit (alpha) for this hypothesis test is 0.05 (5%).

Treatment of students spilling equal quantities is by giving them a joint, averaged (arithmetic mean) ranking for the purposes of calculating the Hodges–Lehmann estimator.

No other hypotheses will be tested during this RCT to avoid the need to reduce the confidence level used to adjust for the problem of multiple comparisons, and hence maximise the possibility of a statistically significant result being observed. (Dunn, 1958)

Secondary Descriptive statistics
Descriptive statistics for both the intervention and control groups, and the combined intervention+control group, will also be reported without further detailed comment:
- The arithmetic mean
- The median
- The interquartile ranges
- The range of the individual groups

Other non-RCT data analysis

The data collected during this RCT will also be analysed to provide insight into:
- *Time dependence of volume of chemicals spills*
- *Correlation between accuracy of measurements by students & volume of chemicals spilled*
- *Correlation between self perception of experimental ability and accuracy of results*
- *Correlation between self perception of experimental ability and volume of chemicals spilled*

References

D. G. Bonett, R. M. Price, *Statistical inference for a linear function of medians: Confidence intervals, hypothesis testing, and sample size requirements* **Psychological Methods**, 2002, Vol. 7, No. 3, 370–383

Dunn, O. J., *Estimation of the Means for Dependent Variables*, **Annals of Mathematical Statistics** (1958). 29 (4): 1095–1111. doi:10.1214/aoms/1177706374

J. L. Hodges and E. L. Lehmann, *Estimates of Location Based on Rank Tests*, **Ann. Math. Statist**.
Volume 34, Number 2 (1963), 598-611.

Mann, H. B.; Whitney, D. R.*, On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other.* **Ann. Math. Statist.** *18 (1947), no. 1, 50--60. doi:10.1214/aoms/1177730491.* https://projecteuclid.org/euclid.aoms/1177730491

K. F. Schulz, D. G. Altman, D. Moher, the CONSORT Group, *CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials*, **Trials** 2010, 11:32 http://www.trialsjournal.com/content/11/1/32