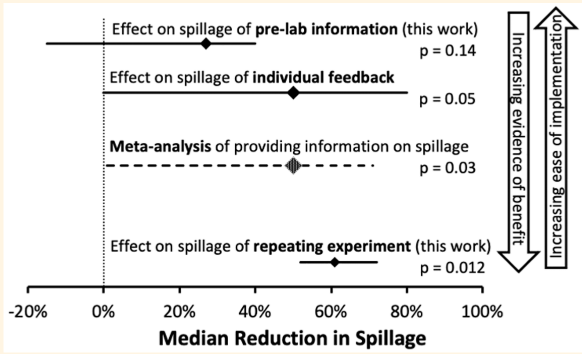
ACS CHEMICAL HEALTH & SAFETY

Effect of Pre-lab Information on Chemical Spillage on Volume Subsequently Spilled: A Randomized Controlled Trial, Meta-analysis, and Comparison with Improvement through Repetition

Moray S. Stark,* Stephen A. Robertson, and Aimilia M. Tsokou

Cite This: *ACS Chem. Health Saf.* 2023, 30, 333–339

ABSTRACT: Understanding methods to improve the safe handling of hazardous chemicals is important to improve laboratory safety. In this work, a simple online resource with contextual safety information on chemical spillage was developed and provided to year 1 undergraduate students prior to undertaking a laboratory practical. The effects of this safety information on amount of chemical subsequently spilled was examined using a randomized controlled trial, with a median effect size for the reduction in spillage of 37%, in comparison with those not receiving this information (95% confidence interval: −18% to 68% reduction and $p = 0.14$). To improve the robustness of this finding, a pretrial protocol for this randomized controlled trial was published on an open platform in a frozen document prior to data collection commencing. The effects of this pre-experiment, nonindividualized safety information was combined, using meta-analysis methodology, with results from a previous study which provided safety information postexperiment based on spillage by individual students; the effect of contextual safety information on chemical spillage gave a median reduction in spillage of 50% (95% confidence interval of 0% to 71% reduction, and $p = 0.034$). Any improvement through repeating the experiment was also investigated with spillage reduced by a median of 61% (95% confidence interval of 52% to 72% reduction, and $p = 0.012$). These three methods for reducing chemical spillage are compared using an implementation science perspective, highlighting that for the three methods discussed there is the trade-off in that the higher the evidence of benefit, the lower the ease, and hence likelihood, of implementation.



KEYWORDS: *laboratory safety, chemical safety, chemical safety education, laboratory safety perception*

■ INTRODUCTION

There is a risk of harm when hazardous chemicals are unintentionally spilled while being handled,[1,2] and there have been a number of serious chemical accidents in university laboratories.[3] Consequently, it is recognized that chemical safety education and training in the undergraduate curriculum should be improved.[4]

While wearing personal protective equipment (PPE) such as appropriate chemical resistant gloves is a necessary aspect when handling hazardous chemicals,[5] the use of PPE is lowest on the hierarchy of risk controls, and providing improved training on chemical handling to reduce spillage would, as an administrative control, be higher on the hierarchy of risk controls, with the ideal always being to minimize chemical spills in the first place.[6]

Consequently, previous work has investigated whether the handling of chemicals by students learning this skill could be improved by providing feedback on the volume they have spilled during previous experiments, along with some context on the risks this could entail if this level of spill was routine.[7] This previous work using individualized, postexperiment feedback on spillage had a potential limitation of requiring a level of commitment of resources, in terms of demonstrator time needed to measure the spillage by each individual student and then provide personalized feedback to the student, which could reduce the likelihood of widespread implementation.

Therefore, in the work reported here, a simpler intervention to help the students improve their chemical handling skills by reducing spillage was developed to provide the students with information on the risks from routinely spilling different amounts of chemical, and to simplify implementation, the students all received the same information online prior to the lab experiment; i.e., it was not individualized, and so would not require any time commitment from laboratory demonstrators.

Received: June 2, 2023
Revised: June 30, 2023
Published: August 16, 2023

https://doi.org/10.1021/acs.chas.3c00054
ACS Chem. Health Saf. 2023, 30, 333–339

This approach aims to go beyond simply providing the students with standard safety information (hazard phrases and symbols)[8] for chemicals, which has been shown to have limitations in promoting understanding of the risk.[9]

The efficacy of implementing this pre-experiment, non-individualized spillage information approach was examined using a randomized controlled trial approach.

It is also investigated here whether students repeating an experiment spill lower amounts of chemical during the second experiment, as it is recognized that repetition of a practical skill can lead to improvement.[10]

These three approaches to affecting spillage of chemicals (repetition of experiment; pre-experiment, nonindividualized context on chemical spillage; and postexperiment, individualized context on spillage) are compared and reviewed from the perspective of implementation science, which aims to consider the barriers to widespread adoption of new approaches to improve education and in other sectors.[11]

■ RANDOMIZED CONTROLLED TRIAL OF EFFECT OF PRE-LAB INFORMATION ON CHEMICAL SPILLAGE

Undergraduate students undertook a spectroscopy experiment, described in detail previously,[7] that involved handling highly colored copper(II) compounds, and any spillages by the students were evident as stains on paper liners that covered the fume cupboards where each student worked individually.

Feedback Tool to Provide Contextual Information on Spillage of Chemicals. Previous work[7] reported nominal acceptable exposure limits for spillage (volume per person per day) for four chemicals ranging in toxicity from 1 M potassium cyanide down to ethanol, which were calculated using Derived No Effect Levels (DNELs) for dermal exposure[12] (assuming an average of a 66 kg person[13]). DNELs are calculated limits undertaken by industry bodies as part of the European Union REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals) Regulation application process to determine the level of exposure for a chemical that should not be exceeded for humans for there to be no effect due to the exposure to the substance.[12,14]

DNEL volumes were converted into areas for spillage onto a paper liner of a type used to cover lab surfaces (Benchguard BG-50E extra absorbent paper). In this previous work (described here as cohort 1), the amount of chemicals spilled by individual students was measured, and comparison with these threshold areas of spillage allowed the students to be given individualized contextual feedback on the potential safety consequences if the students routinely spilled this mount of chemical in future work.

As it was identified that measuring spillage for individual students required a non-negligible amount of time by the lab demonstrators and staff, a motivation for the work reported here was to develop a means to provide similar context to students of the potential safety consequences of spilling chemicals but which would require minimal investment in time (or money). The method developed to achieve this was to provide the students with similar safety information as previously reported, but *prior* to their laboratory experiments (as opposed to after). To help the students visualize these nominally acceptable threshold areas of spillage of the four chosen chemicals[7] they were expressed in terms of objects that had similar areas to the thresholds and would be familiar to the students. This information is summarized in Table 1, and the full description (including an introduction to the topic) is given in the pretrial protocol, published before the start of data collection.[15]

Table 1. Summary of Online Contextual Safety Information on Spillage Given to Students in the Intervention Group before the Practical Session

If you routinely spilt chemicals with the area of a 20p coin [≈ 2 cm diameter] each day then you would be able to handle high hazard chemicals safely, such as cyanide solutions.
If you routinely spilt chemicals with the area of a hand each day then you would be able to handle higher hazard chemicals, such as hexane, safely but not more hazardous chemicals, such as cyanides.
If you routinely spilt chemicals with the area of your lab book (A4) each day then you would be able to handle routinely hazardous chemicals safely, such as ethyl acetate, but not more hazardous chemicals, such as hexane or cyanides.
If you routinely spilt chemicals with the area of half a fumehood each day then you would be able to handle low hazard chemicals safely only, such as ethanol, but not more hazardous chemicals.

An advantage of this approach was that this was a noticeable improvement in the ease of implementation, but with the potential downside that the information to students on spillage was not based on their own spillage, so was not individualized and so could, potentially, be less effective in reducing subsequent spillage of chemicals.

RCT Methodology. The effect of providing pre-lab information on chemical spillage on the amount subsequently spilled by students during an experiment was investigated using a randomized controlled trial (RCT). The RCT methodology, including statistical analysis, is generally as reported for the previous study on the effects of individualized feedback on spillage,[7] excepting any differences highlighted here.

Eligibility criteria for participants were solely that they were first year undergraduate chemistry students at the major UK university where this study was undertaken (described here as cohort 2) and who gave consent.

As a standard part of their practical course, the students were required to complete a pre-lab online test prior to the experiment they were to undertake, to help ensure they had read about and understood the key aspects of the experiment, including standard safety information (GHS (Globally Harmonized System)[8] hazard phrases and symbols) on the chemicals. To implement both the consent request, randomization, and the intervention itself, this pre-lab online test was modified to include two additional questions. The first asked whether the students consented to take part in this randomized controlled trial, and those declining or not answering were not included in the RCT, with no penalties or rewards for each of the responses.

Those who did consent to take part were then randomly allocated to either continue directly to the pre-lab test (the control group), or to receive the pre-experiment chemical spills information intervention being examined by this RCT, summarized in Table 1 and available in full in the supporting literature (the intervention group). The primary outcome of this RCT was the volume spilled by each student during their experiments.

This was a parallel group randomized controlled trial, with nominal 1:1 allocation to intervention group (who received the pre-lab information on spillage) and the control group (who did not receive this pre-lab information on spillage).

This RCT can be categorized as a waiting list (also known as a waitlist) randomized controlled trial,[16] as all students

https://doi.org/10.1021/acs.chas.3c00054
ACS Chem. Health Saf. 2023, 30, 333−339

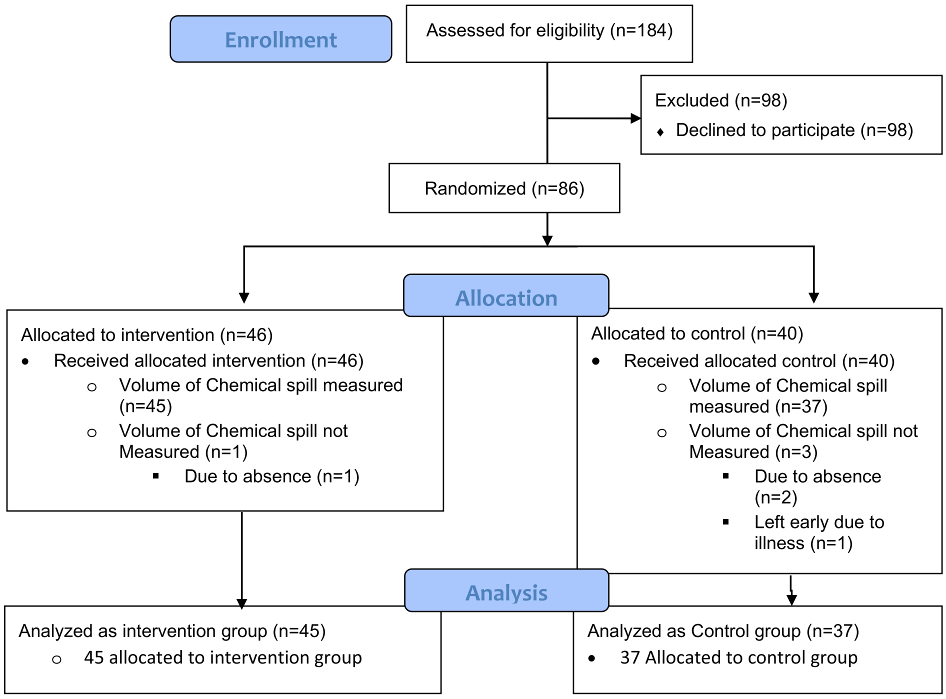


Figure 1. CONSORT flow diagram for the transparent reporting of trials (CONsolidated Standards Of Reporting Trials).[27]

received the spillage information summarized in Table 1 in the end, with the control group (as well as those declining to participate in the RCT) getting this information after the intervention for the RCT had been completed. As no other aspects of the students' lab experience were altered for this trial, apart from providing this short additional information on chemical spillage, this trail can also be viewed as a pragmatic RCT.[17]

Measurement of volume spilled was by an author (SAR) and was undertaken prior to disclosure of allocation to intervention or control groups; therefore, assessment for this RCT was blinded.

For the primary outcome (volume spilled), two nonparametric statistical tools were prespecified in the trial protocol, available in the Supporting Information, and published prior to the RCT taking place at the Centre for Open Science as a pretrial protocol.[15]

The prespecified summary statistic for the effect size was the Hodges–Lehmann estimator[18,19] and associated 95% confidence intervals,[20] with the primary inferential statistic being the Mann–Whitney U test,[21] with a prespecified confidence level of $p = 0.05$, and a null hypothesis that providing pre-experiment information on spillage has no effect on the volume of chemical subsequently spilled by the students.[15] The Mann–Whitney U test is most appropriate for ordinal outputs,[22] so was used here with rankings of volume spilled by the students.

No other hypotheses were prespecified or tested during this trial to avoid the need to reduce the confidence level used to adjust for the problem of multiple comparisons.[23] There were no deviations or alternations between the statistical methods proposed to be used in the pretrial protocol[15] and those implemented in the trial. All calculations were performed both in Excel and SPSS,[24] and for this study statistical analysis used a *per protocol* approach[25] (analysis based on what the students actually received), which gave identical results to those using an *intention to treat* analysis (analysis based on which group the student were allocated to).[26]

Pretrial Protocol of the Randomized Controlled Trial. The protocol for the randomized control trial into the effects of pre-lab information on chemical spillage was published, prior to data collection starting, in a public registration in a frozen, noneditable Pre-Trial Protocol at the Center for Open Science, with the protocol submitted, registered, and frozen on October 25, 2019.[15]

Results of RCT of Effect of Feedback on Chemical Spillage. Of the 184 students in cohort 2, 86 consented to take part in this trial (47% of cohort). As allocation to intervention or control groups was determined for each individual randomly during their interaction with online pre-lab information and questions, the two groups were not equal from the outset, with 40 allocated to the control group and 46 to the intervention group. Data was not collected for four individuals (3 in control, 1 in intervention, due to absence or leaving lab early) and were not included in the statistical analysis as per the trial protocol.

This is shown in the CONSORT (CONsolidated Standards Of Reporting Trials) flow diagram, Figure 1, which are a set of benchmark guidelines designed to allow for more straightforward replication and subsequent synthesis with previous or future findings into a combined result using a meta-analysis of more than one study on a topic.[27]

Spillage by the two groups of students (intervention group who received safety information prior to the experiment and the control group who did not receive this prior to carrying out the experiment) is shown in Figure 2, with the spillages ordered by lowest at the left and highest at the right. The spillage volumes are shown on a logarithmic scale due to the very wide range spilled of almost 3 orders of magnitude, and, visually, spillages by those receiving the intervention were generally lower than those who did not receive the intervention.

https://doi.org/10.1021/acs.chas.3c00054
ACS Chem. Health Saf. 2023, 30, 333−339

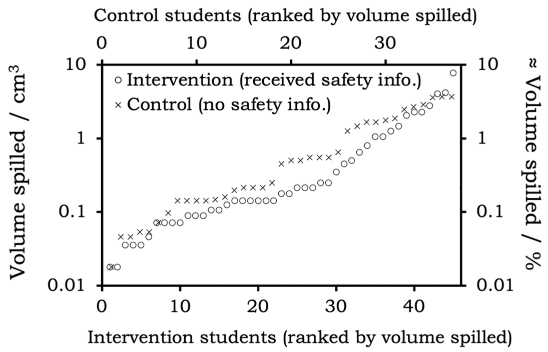


Figure 2. Volumes of solution, displayed with logarithmic vertical axes, spilled by 45 students receiving the feedback intervention and the 37 students not receiving this feedback (ordered left to right, lowest to highest).

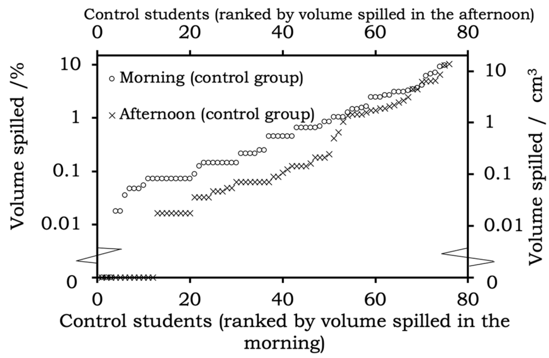


Figure 3. Volumes of solution spilled by 75 students in cohort 1 belonging to the control group in the morning and the same 75 students' volume spilled when repeating the experiment in the afternoon (ordered left to right, lowest to highest).

To quantify the effect of providing this pre-lab safety information on chemical spillage this was estimated using the Hodges–Lehmann estimator for effect size, which is a measure of the median difference between all the different possible pairings between control and intervention groups,[28] using differences of the $\log_{10}$(volume) as described previously.[7] Providing pre-lab contextual safety information on spillage was found to give a median effect size for the reduction in spillage of 37%, in comparison with those not receiving this information (with a 95% confidence interval of −18% to 68% reduction and a p value for accepting or rejecting the null hypothesis of $p = 0.14$ (using Mann–Witney analysis).[21,29,30] There were no harms or unintended effects noted in either group for this trial.

EFFECT OF REPEATING EXPERIMENT ON CHEMICAL SPILLAGE

A common paradigm is that an individual repeatedly practicing a skill will lead to improvements in that skill.[10] However, there is not a great deal of quantified evidence for this in the chemistry sector and, more specifically, none on how students' abilities to handle chemicals without spillage can improve with practice. Data collected for a previously report randomized control trial[7] on methods to improve chemical spillage (cohort 1) allows a quantification of this effect to be made. In this previous study, a control and an intervention group carried out the same experiment twice: once in the morning and once in the afternoon. The intervention group received feedback on the amount of chemical that was spilled during the morning experiments, and the difference in spillage for the afternoon experiments between the control group (who received no feedback) and the intervention group was reported.

As spillage for the control group of this cohort was also recorded for both morning and afternoon experiments, this allows a quantification of the effect on safe chemical handling through repeating an experiment (the data for the intervention group was not used in this morning/afternoon comparison, as the intervention could confound any effect due to just repeating the experiment).[31]

The volume of chemicals spilled by the control group of cohort 1 in the morning was compared with that spilled by the control group in the afternoon, and this is shown in Figure 3; it is evident that across almost the whole range of spillages that during the afternoon experiments a noticeably lower proportion of chemical was spilled in comparison with the their first attempt at the experiment in the morning.

As the same students were performing the experiment twice, once in the morning then repeating in the afternoon, it could perhaps be assumed that there would be a strong correlation between spillages by individuals, and hence statistical tests that rely on paired data (such as the Wilcoxon Signed Rank Test for significance[32]) would give a higher statistical power.[33] However, it has previously been demonstrated that there is in fact a very low correlation between spillage by individual students during the morning and afternoon experiments;[34] therefore statistical tests that do not rely on correlation are used in preference. This unusual phenomenon of being able to observe statistically meaningful changes at the group level while simultaneously having little predictive understanding at the level of the individual is termed the *Group to Individual Inference problem* and has been recognized previously.[34,35]

The effect size was therefore calculated using the Hodges–Lehmann estimator (appropriate for unpaired/noncorrelated non-Gaussian distributed data),[19] with spillage found to be reduced through repeating the experiment by a median of 61%, with a 95% confidence interval of a 52% to 72% reduction in spillage, with a significance of $p = 0.012$ (calculated using the Mann–Whitney U test).[21]

DISCUSSION

Comparison and meta-analysis of Pre- and Postlab Information on Spillage. The work reported here on the effect of pre-lab, nonindividualized information on the effects of chemical spillage can be compared with the previously published effect of individualized, postlab information on chemical spillage,[7] with the latter having a slightly larger effect size and lower p value. The p value is above the prespecified threshold of $p = 0.05$; however the result is still of interest as a presentation of evidence on this technique and to allow these results to be synthesized with other work in meta-analysis—the limitations of using an arbitrary threshold for p-values for deciding significance has also been noted.[36]

These two effect sizes are compared graphically in a Forest plot[37] in Figure 4. Also included is a meta-analysis which combines data from these two studies, which was undertaken to assess the averaged effect of contextual safety information (provided in any manner) on chemical spillage by students, using individual participant data.[38] The effect size for the median change in spillage volume for this meta-analysis on

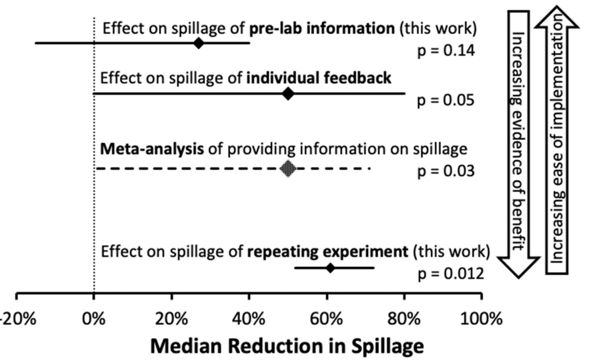


Figure 4. Forest plot comparing effects on chemical spillage of pre-lab contextual information on chemical spillage; individual feedback on previous spillage; a meta-analysis of pre-lab information and individual feedback to give a general effect of proving information on spillage; and the effect of repeating experiments (also shown is an indication of ranking by evidence and ranking by ease of implementation).

receiving information was a 50% reduction (95% confidence interval of 0% to 71% reduction (using Hodges−Lehmann analysis) and a p value of 0.034 (using the Mann−Whitney U test). This meta-analysis result therefore can give increased confidence that providing students with information, in some form, on the possible safety consequences of spillage can help reduce the amount of chemical they subsequently spill.

Implementation Science Approach to Evaluating Methods to Improve Safe Chemical Handling. An important (if often implicit) aim for researchers is for their published work on, for example, new educational interventions to have an impact by being widely taken up and implemented by the sector. However, what affects whether reported interventions actually get taken up?

Evidence of effectiveness of a new technique or approach that could convince practitioners to adopt it may include quoting p (significance) values (sometimes taken, crudely, as a measure of whether the intervention has a real effect[39]). Further evidence in published work that may convince the audience would be to include an effect size (with corresponding confidence intervals), which allows a judgment to be made on whether the technique can produce a worthwhile effect (and not just a very small but possibly statistically significant, effect).

Of the three techniques for reducing chemical spillage reported or reviewed here (pre-experiment nonindividualized information on spillage; postexperiment individualized information on spillage; and repeating the experiment), all have a reasonable effect size, which could support implementation if all other factors were positive. However, on statistical significance (p values), the three approaches differ noticeably: pre-lab information has a p value greater than the predetermined threshold of 0.05 specified in the trial protocols for all these studies; postlab information has a p = 0.05, on the threshold of acceptance of rejecting the null hypothesis of no effect; while repeating the experiment has a p value below the threshold at.034. These are indicated on Figure 4 as increasing evidence of benefit.

However, there are other factors that can affect whether a new educational technique will be widely adopted, and the analysis of these factors is often described as implementation science. This is a well-developed subject in the field of medicine and health where it receives a comparatively high level of scrutiny, due to the often very high costs of developing new treatments and techniques and the risks to human wellbeing when implementing them.[40,41] In contrast, implementation science in the education sector has not been so commonly discussed.[11]

Factors, in addition to statistical evidence, that may affect the wider uptake of a new technique could include, for example: ease of implementation, including cost and time for implementation; opportunity cost; and ease of reversal of implementation, should new evidence become available in the future supporting lack of effectiveness (i.e., if a timely reversal of implementation is difficult, then this inevitably creates an increased "activation energy" for implementation which is only likely to be overcome by requiring a higher level of confidence and evidence of positive effects for the technique).

Of the three approaches compared here: pre-lab information on spillage is a very low time commitment to both staff and students and hence very low cost, and it is straightforward to reverse; postlab individualized information on spillage is a somewhat greater commitment for staff and demonstrator time and, hence, cost (estimated at 1 $/£/€ and 1 min per student,[7] which could be viewed as nontrivial when there may be 100+ students per lab session) but is comparatively straightforward to reverse; and repeating the experiment is a much greater commitment in time for the students and staff and, due to this and the high costs of lab space, is noticeably more expensive (and as lab work is often planned and timetabled a year or more in advance, repeating an experiment is difficult to reverse quickly, once implemented). Repeating experiments also carries a much greater opportunity cost, as repeating an experiment would usually necessitate something else being displaced from the students' lab activities, though if different experiments used and developed the same chemical handling skills, then it may be expected that improvement in skills could be similar.

A (comparatively subjective) ordinal ranking[22] can be produced reflecting the ease of implementation of these three techniques, which is also indicated in Figure 4. This highlights that the technique with the least supporting evidence (i.e., the highest p value), providing pre-lab information on chemical spillage, is the most easily implementable, and conversely, the most statistically robust approach (to have students repeat the experiment) is the least implementable approach, being (comparatively) costly in terms of both time and money. This highlights therefore that a simple comparison of statistical evidence on how to improve safe chemical handling may be insufficient for practical decision making, and that, for instance, adoption of the pre-lab contextual information on chemical spillage could be the most readily implemented (including for repeating the trial reported here to firm up the evidence through meta-analysis).

Publishing of Pretrial Protocols for Investigations. As highlighted, an important factor that can influence practitioners on whether to adopt a new technique is how robust they view any evidence reported in the literature. The robustness of published research was highlighted in the medicine and health research sector by Ioannidis in a landmark essay with the provocative title of, *Why Most Published Research Findings Are False*,[42] and subsequently the field of psychology research has also discussed the so-called "reproducibility crisis" and has started a widespread re-evaluation of landmark published work to check if key findings are in fact

https://doi.org/10.1021/acs.chas.3c00054
ACS Chem. Health Saf. 2023, 30, 333−339

replicable.[43,44] This issue has also been discussed in the context of education[45,46] and in the wider field of science.[47]

One measure that project designers can use that can help to improve the robustness, and hence believability, of studies is to say publicly, in advance of the start of data collection, what is going to be investigated, the methods for achieving this, and, crucially, the statistical and other analysis techniques that are to be employed to draw conclusions from the data collected. The reasons for doing this is that it can reduce or remove the possibility of selective publication from large data sets or of multiple statistical analyses being carried out on data sets, through only publishing those that pass a certain significance test.[23] For this reason, the project plan for the randomized controlled trial reported here was published as a pretrial protocol.[15]

■ CONCLUSIONS

The effectiveness of two methods to improve chemical safety by reducing spillage is reported here; by providing non-individualized pre-lab information on the possible safety consequences of spilling chemicals, and by repeating the experiment. The reduction in spillage noted for these two approaches is compared with previous reported work where the approach was to provide individualized postexperiment information to each student on what volume they had spilled and the safety implications of this.

These three methods are examined from an implementation science perspective, which highlights that the technique with the least experimental evidence (providing nonindividualized pre-lab information on the possible consequences of spilling chemicals) is in fact the most easily implementable, as it would require the least commitment of time and resources and is easily reversible.

Results for two randomized controlled trials on the two techniques for improving chemical handling (pre-lab non-individualized information, and postexperiment individualized information on the consequences of spillage) were combined using a meta-analysis approach, which provided stronger evidence ($p = 0.03$) that giving students information on the safety consequences of chemical spillage in some form can reduce subsequent spillage.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.chas.3c00054.

Full text of the online contextual safety information on spillage summarized in Table 1 and given to students in the intervention group before the practical session (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Moray S. Stark − *Department of Chemistry, University of York, Heslington, York YO10 5DD, U.K.; Strathclyde Centre for Occupational Safety & Health, University of Strathclyde, Glasgow G1 1QE, U.K.;* orcid.org/0000-0002-2175-2055; Email: moray.stark@strath.ac.uk

Authors

Stephen A. Robertson − *Department of Chemistry, University of York, Heslington, York YO10 5DD, U.K.*

Aimilia M. Tsokou − *Department of Chemistry, University of York, Heslington, York YO10 5DD, U.K.*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.chas.3c00054

Author Contributions

CRediT: **Moray S. Stark** conceptualization (lead), data curation (lead), formal analysis (equal), investigation (equal), methodology (lead), project administration (lead), supervision (lead), visualization (equal), writing-original draft (lead), writing-review & editing (lead); **Stephen A. Robertson** formal analysis (equal), investigation (equal), visualization (equal), writing-review & editing (supporting); **Aimilia M. Tsokou** formal analysis (equal), investigation (equal), methodology (supporting), visualization (equal).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Helen Burrell, Liza Binnington, Phil Helliwell, & Scott Hicks for support during these undergraduate lab practicals, and David Pugh and Nick Wood for developing and running the spectroscopy experiment undertaken by the students. Information on RCT Preregistration of Trial Protocol was published on the Center for Open Science platform in a frozen document in advance of the start of data collection.[15] The trial design was submitted to the University of York Faculty of Science Ethics Committee and approved prior to the start of data collection. Participants agreed to be randomized to take part in this RCT through giving consent as part of a pre-lab quiz taken routinely before their laboratory experiment. This work was unfunded, except for a small consumables budget provided by the University of York

■ REFERENCES

(1) Brückner, S.; Marendaz, J.-L.; Meyer, T. Using very toxic or especially hazardous chemical substances in a research and teaching institution. *Safety Science* **2016**, *88*, 1−15.

(2) Mannan, M. S.; O'Connor, T. M.; Keren, N. Patterns and trends in injuries due to chemicals based on OSHA occupational injury and illness statistics. *Journal of Hazardous Materials* **2009**, *163* (1), 349−356.

(3) Ménard, A. D.; Trant, J. F. A review and critique of academic lab safety research. *Nat. Chem.* **2020**, *12* (1), 17−25.

(4) Fivizzani, K. P. Where are we with lab safety education: Who, what, when, where, and how? *Journal of Chemical Health & Safety* **2016**, *23* (5), 18−20.

(5) Blayney, M. B. The need for empirically derived permeation data for personal protective equipment: the death of Dr. Karen E. Wetterhahn. *Appl. Occup. Environ. Hyg.* **2001**, *16* (2), 233−6.

(6) Manuele, F. A. Risk assessment & hierarchies of control. *Professional Safety* **2005**, *50* (5), 33−39.

(7) Tsokou, A. M.; Howells, A.; Stark, M. S. Measuring and Reducing Chemical Spills by Students: A Randomized Controlled Trial of Providing Feedback. *J. Chem. Educ.* **2019**, *96* (10), 2180−2187.

(8) *United Nations Economic Commission for Europe Globally Harmonized System of Classification and Labelling of Chemicals (GHS)*. https://unece.org/ghs-1st-edition-2003 (accessed 2nd June 2023).

(9) Karapantsios, T. D.; Boutskou, E. I.; Touliopoulou, E.; Mavros, P. Evaluation of chemical laboratory safety based on student comprehension of chemicals labelling. *Education for Chemical Engineers* **2008**, *3* (1), e66−e73.

(10) Magallón, S.; Narbona, J.; Crespo-Eguílaz, N. Acquisition of Motor and Cognitive Skills through Repetition in Typically Developing Children. *PLoS One* **2016**, *11* (7), No. e0158684.
(11) Moir, T. Why Is Implementation Science Important for Intervention Design and Evaluation Within Educational Settings? *Frontiers in Education* **2018**, *3*, DOI: 10.3389/feduc.2018.00061
(12) Kreider, M. L.; Spencer Williams, E. Interpreting REACH guidance in the determination of the derived no effect level (DNEL). *Regul. Toxicol. Pharmacol.* **2010**, *58* (2), 323−329.
(13) Serlachius, A.; Hamer, M.; Wardle, J. Stress and weight change in university students in the United Kingdom. *Physiology & Behavior* **2007**, *92* (4), 548−553.
(14) European Chemicals Agency. *Potassium cyanide: Toxicological Summary* (accessed 30th June 2023).
(15) Stark, M. S.; Robertson, S. A. *Pretrial protocol for RCT to Assess the Effect of Online Information on Spillage on Safe Chemical Handling Skills*; Center for Open Science, Open Science Framework Registries: 2019.
(16) Ronaldson, S.; Adamson, J.; Dyson, L.; Torgerson, D. Waiting list randomized controlled trial within a case-finding design: methodological considerations. *Journal of Evaluation in Clinical Practice* **2014**, *20* (5), 601−605.
(17) Zwarenstein, M.; Treweek, S.; Gagnier, J. J.; Altman, D. G.; Tunis, S.; Haynes, B.; Oxman, A. D.; Moher, D. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ.* **2008**, *337*, a2390.
(18) Rosenkranz, G. K. A note on the Hodges-Lehmann estimator. *Pharmaceutical Statistics* **2010**, *9* (2), 162−167.
(19) Hodges, J. L.; Lehmann, E. L. Estimates of Location Based on Rank Tests. *Annals of Mathematical Statistics* **1963**, *34*, 598−611.
(20) Bonett, D. G.; Price, R. M. Statistical inference for a linear function of medians: confidence intervals, hypothesis testing, and sample size requirements. *Psychol Methods* **2002**, *7* (3), 370−83.
(21) Mann, H. B.; Whitney, D. R. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Annals of Mathematical Statistics* **1947**, *18* (1), 50−60.
(22) Stevens, S. S. On the Theory of Scales of Measurement. *Science* **1946**, *103* (2684), 677−680.
(23) Smith, G. D.; Ebrahim, S. Data dredging, bias, or confounding. *BMJ* **2002**, *325* (7378), 1437−1438.
(24) *IBM SPSS Software.* https://www.ibm.com/uk-en/analytics/spss-statistics-software (accessed Jul 12, 2019).
(25) Porta, N.; Bonet, C.; Cobo, E. Discordance between reported intention-to-treat and per protocol analyses. *J. Clin Epidemiol* **2007**, *60* (7), 663−9.
(26) Gupta, S. K. Intention-to-treat concept: A review. *Perspect Clin Res.* **2011**, *2* (3), 109−12.
(27) Schulz, K. F.; Altman, D. G.; Moher, D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *Trials* **2010**, *11*, 32.
(28) Wolfe, H. *Nonparametric Statistical Methods*, 1st ed.; Wiley: New York, 1973.
(29) Rosenkranz, G. K. A note on the Hodges-Lehmann estimator. *Pharm. Stat* **2010**, *9* (2), 162−7.
(30) Nakagawa, S.; Cuthill, I. C. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews* **2007**, *82* (4), 591−605.
(31) Altman, D. G.; Bland, J. M. Uncertainty beyond sampling error. *BMJ: British Medical Journal* **2014**, *349*, g7065.
(32) Wilcoxon, F. Probability Tables for Individual Comparisons by Ranking Methods. *Biometrics* **1947**, *3* (3), 119−122.
(33) Cohen, J. Statistical Power Analysis. *Current Directions in Psychological Science* **1992**, *1* (3), 98−101.
(34) Stark, M. S.; Tsokou, A. M.; Howells, A. Chemical Spillage as a Model for Accident Causation. *J. Chem. Educ.* **2021**, *98* (1), 50−56.
(35) Faigman, D. L.; Monahan, J.; Slobogin, C. Group to Individual (G2i) Inference in Scientific Expert Testimony. *University of Chicago Law Review* **2014**, *81* (2), 417−480.
(36) Sterne, J. A. C.; Cox, D. R.; Smith, G. D. Sifting the evidence—what's wrong with significance tests? *BMJ.* **2001**, *322* (7280), 226−231.
(37) Lewis, S.; Clarke, M. Forest plots: trying to see the wood and the trees. *BMJ.* **2001**, *322* (7300), 1479−1480.
(38) Riley, R. D.; Lambert, P. C.; Abo-Zaid, G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ.* **2010**, *340*, c221.
(39) Altman, D. G.; Bland, J. M. Statistics notes: Absence of evidence is not evidence of absence. *BMJ.* **1995**, *311* (7003), 485.
(40) Madon, T.; Hofman, K. J.; Kupfer, L.; Glass, R. I. Implementation Science. *Science* **2007**, *318* (5857), 1728−1729.
(41) Bauer, M. S.; Damschroder, L.; Hagedorn, H.; Smith, J.; Kilbourne, A. M. An introduction to implementation science for the non-specialist. *BMC Psychology* **2015**, *3* (1), 32.
(42) Ioannidis, J. P. Why most published research findings are false. *PLoS Med.* **2005**, *2* (8), No. e124.
(43) Open Science Collaboration. Estimating the reproducibility of psychological science. *Science* **2015**, *349* (6251), aac4716.
(44) Maxwell, S. E.; Lau, M. Y.; Howard, G. S. Is psychology suffering from a replication crisis? What does "failure to replicate" really mean? *American Psychologist* **2015**, *70*, 487−498.
(45) Holme, T. A. Reproducibility, Replication, and Generalization in Research about Teaching Innovation. *J. Chem. Educ.* **2019**, *96* (11), 2359−2360.
(46) Cooper, M. M. The Replication Crisis and Chemistry Education Research. *J. Chem. Educ.* **2018**, *95* (1), 1−2.
(47) *Reproducibility and Replicability in Science*; The National Academies Press: Washington, DC, 2019; p 256.