CASE STUDY

A Comparison of Nursing Maths Support Approaches

John Little, Department for the Enhancement of Learning, Teaching and Access, Robert Gordon University, Aberdeen, United Kingdom. Email: j.little@rgu.ac.uk

Abstract

This article compares interventions for improvement of nursing and midwifery students' numeracy and drug calculations skills respectively; numeracy test / re-test with maths support versus a numeracy workshop and a calculations drop-in. In the first intervention supported students (n=11) increased their mean score from 19 out of 40 to 30 out of 40 and their peers (n=165) increased from 27 out of 40 to 30 out of 40, both statistically significant increases. In the second intervention the highest scoring group of students made use of both the workshop and the drop-in. Whilst there was no statistically significant difference in mean scores across the four identified groups of students (those with no maths support (n=205), those with foundation numeracy support only (n=18), those with calculations drop-in support only (n=11) and those (n=6) with both foundation numeracy and calculations drop-in support) there were no low scoring outliers amongst the students who made use of any of the support offered. Although there is stronger evidence for the first intervention having impact, the second intervention is more aligned to drug calculations skills development. However, the latter would need a larger scale sample to confirm efficacy.

Keywords: dosage, calculations, numeracy, nursing.

1. Introduction

The Nursing and Midwifery Council (NMC), the responsible professional body in the UK, seeks to focus on numeracy skills development, implied through its code (NMC, 2015a) and, more explicitly, through the standards that underpin the code (NMC, 2015b), albeit there is no national standard drug calculations test.

At the author's institution one to one, small group and large group maths support outside of main teaching programmes has been available to nursing students since 2003 and over the last ten years a commercial e-assessment package has been adopted as the means to deliver drug calculations proficiency. Without a doubt this has brought great benefits, not least by indirect promotion of a self-test strategy via online practice and formative assessments that aligns well to the known successful study and learning strategy of frequent testing and spread effort (Hartwig and Dunlosky, 2012). This author has also previously noted the benefits of multiple online formative test attempts (Little, 2006).

Note, in this article p values are mostly reported to the three decimal places presented in SPSS outputs with an assumed acceptance level of 0.05, the exception being p values obtained for transformed data.

2. Session 2012-2013 Intervention 1: Re-test of Foundation Numeracy Skills with Interim Maths Support

In session 2012-2013 students who didn't achieve full marks on their formative foundation numeracy (FN) assessment were asked to make a second attempt at the assessment, with the option of one to one or small group maths support (MS) appointments as an intervention in the interim. Note, the following figures only include students who completed both first and second attempts at the intended times.

Of 205 students, 29 are excluded due to the absence of one or more tests scores. Second re-test data for 6 of those 29 excluded students is available (all of these having appeared to miss their original second attempt) but only one of those six scored full marks and only one took advantage of maths support. Of the remaining 176 cases a group of 11 made use of maths support.

This test / re-test approach arguably aimed to exploit the learning effect of test repetition which Hartwig and Dunlosky (2012) allude to. Shapiro Wilk normality tests indicate that only the maths support students' data is suitable for assessing the change in test score using a paired samples t test (n=11, p=0.401), the other group (n=165, p<0.001) ostensibly necessitating use of the non-parametric alternative, the Wilcoxon Signed Rank comparison of medians test or, preferably, data transformation. In the latter case, for which Shapiro-Wilk is less sensitive due to the large group size (n>30), the relevant histogram indicates an obvious positive skew for the non-maths support data, an indicator of data that probably doesn't belong to a normal distribution (see Appendix for histogram). That said, it is known that the paired samples t test can be robust to departures from normality (Zumbo and Jennings, 2002) for larger samples (n>30) and effect size > 0.36 or, more commonly, effect size > 0.5 and so results from both tests and for the transformed non-maths support paired data are reported here. Assuming that mixed between-within ANOVA is similarly robust, results for the latter are also reported.

According to the SPSS help file (IBM Support, 2018) the most appropriate transformation for positively skewed data that includes values close to but not predominantly zero is a logarithmic transformation. However, this presents a difficulty in that it cannot be directly used with negative values i.e. a number of non-maths support students actually saw their scores fall after re-test. One option then is to add a constant to the paired difference data and then apply a logarithmic transformation. This in turn presents the problem of different results for different constants added. Notwithstanding the latter issue the t statistic can then be obtained from the mean of the transformed differences divided by the standard error of the transformed differences (sample standard deviation divided by the square root of the sample size n=165) with n-1 degrees of freedom (Shier, 2004) and a t table consulted to obtain the relevant critical value.

A paired samples t test of supported students' scores indicates a statistically significant increase in mean score from 19.09 to 30.36 (t = -5.214, n = 11, p < 0.001) and the corresponding Wilcoxon Signed Rank Test result is also significant (p = 0.003). A paired samples t test of the peer group also indicates a statistically significant increase in mean score from 26.65 to 30.35 (t = -10.140, n = 165, p < 0.001) and the corresponding Wilcoxon Signed Rank Test result is also significant (p < 0.001). The effect size, being calculated from $t^2/(t^2+(n-1))$ (Pallant, 2010), yields a large effect size of 0.731 and a medium effect size of 0.385 in each respective case, the latter exceeding the 0.36 robustness threshold suggested by Zumbo and Jennings (2002).

For completeness, however, a paired samples t test of non-maths support students' log transformed paired difference data was also performed (adding on a constant of 10 to eliminate negative and zero differences) and this yields t = 93.12 with t_{crit} for 164 degrees of freedom = 1.97, this indicating a statistically significant increase in mean score. Changing that constant to the smallest workable whole number yields t = 68.44 which, whilst a noticeably smaller value, still far exceeds the critical value. In either case p is < 0.00001 so the choice of constant doesn't affect the conclusion or, rather, since there is agreement between the paired samples t test, the Wilcoxon test and the test of the transformed data it is reasonable to conclude that there was a real increase in mean score for the non-maths support students.

Note, the change in sign from the negative t values noted earlier simply arises from the order that the paired differences are taken in. One would assume that entering the relevant assessment attempt score variables, mark 1 and mark 2, in that same order in the paired t test SPSS dialogue box that mark1 might be subtracted from mark2 but clearly SPSS interprets the left-hand cell as being the 'more recent' data. The transformed data t values were obtained using MS Excel with the paired differences obtained as mark2-mark1.

A mixed between-within anova yields a significant interaction between time and MS (Wilks Lambda $p < 0.001$), implying that caution must be exercised in interpretation of main effects. Time (re-test opportunity) is a significant main effect (Wilks Lambda $p < 0.001$) with a medium effect size (partial eta squared = 0.36) but the between subjects' effect for MS is not statistically significant ($p = 0.074$), indicating that the maths support intervention was not so important in increasing test scores as the re-test opportunity.

Based on a paired difference standard deviation of 4.68 derived from the present data set, Minitab power and sample size calculator indicates that a change in mean score of greater than 5.0 can be detected with a statistical power of 0.85 with 10 subjects. Much as Wright (2008) observes, it might be reasonably argued that an increase in mean score of 1.0 is clinically important as this may influence a future clinical decision on a calculation but then a much larger sample (n=199 for both test and re-test group) would be needed.

The results that have been obtained ostensibly represent the ideal outcome from a maths support perspective; the weaker group improved to a level comparable to their peers. The scatter graph in Figure 1 provides a visual illustration and further insight into the performance of the maths support group. Most students lie in the upper triangular part of the graph, which means the second attempts are better than the first attempts for most students. However, the students with maths supports got more improvement in their scores i.e. they are generally more to the top and left of the scatter graph than their peers. In fact, all 11 of the maths support students increased their scores whereas 26 out of 165 non-maths support students either scored more poorly or the same as in their first attempt.

However, there are arguably two groups of supported students – those who responded very well to the support (scoring >= 30) and those who responded less well (scoring < 30). This may be because the weaker students were harder to reach in the relatively short two-week period between assessment attempts and hence the latent need for more one to one or small group maths support.

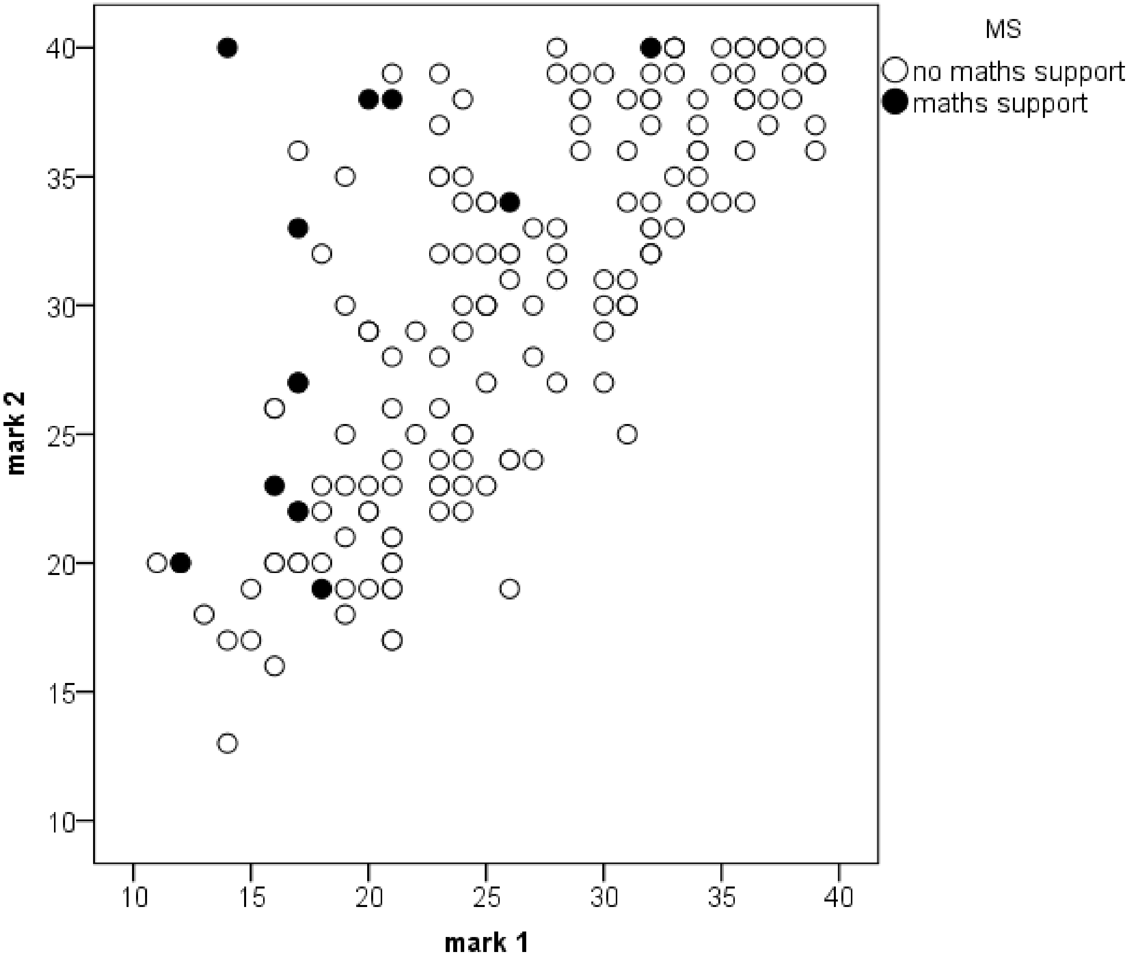

Figure 1. Scattergraph of 2nd Attempt (mark 2) vs 1st Attempt (mark 1) Clustered by MS

3. Session 2015-2016 Intervention 2: Foundation Numeracy Workshop and Drug Calculations Drop-in

In session 2015-2016 a foundation numeracy (FN) workshop intervention was employed to help students improve their numeracy skills. Additionally, a student-led large group drop-in session of solids, liquids and injections (SLI) calculations was arranged prior to the corresponding formative assessment.

Whilst attendance at workshops and uptake of the maths support appointment service was limited, the student led drug calculations drop-in was well attended and there is some weak evidence that both of these interventions helped students but especially so where both help opportunities were used i.e. there may have been a cumulative benefit.

Table 1. Range and Mean SLI Drug Calculation Scores for Four Different MS Groups: 1 (those who didn’t use maths support), 2 (those who attended the FN workshop), 3 (those who attended the SLI drop-in) and 4 (those who attended both the FN workshop and the SLI drop-in)

Maths Support Group	N	Min score	Max score	Mean score	Std. Deviation
1 No maths support	205	17	90	81.76	12.349
2 FN maths support only	18	37	90	80.06	12.379
3 SLI maths support only	11	65	90	80.82	7.859
4 SLI and FN maths support	6	83	90	87.17	2.714

While the widely differing sample sizes mean that these groups’ results are not directly comparable, Table 1 indicates that the mean SLI mark was highest for the small group of students who attended both the FN workshop and the SLI drop in. The only check against this group being biased towards self-motivated high achievers is anecdotal. That is, through discussion of question solutions with the students it became clear that all of the drop-in students exhibited some amount of difficulty with drug calculations during the session; the hope being, of course, that those difficulties were resolved by the session.

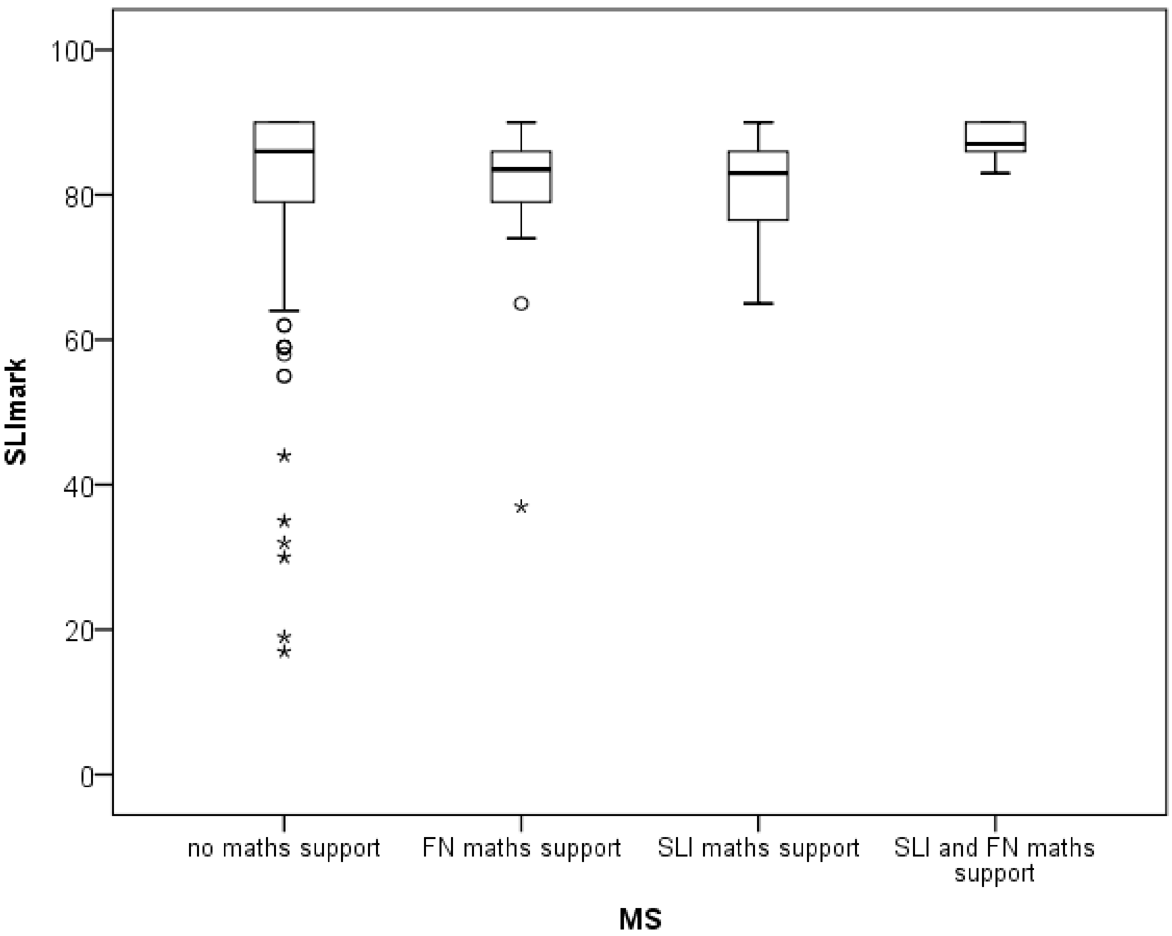

Figure 2. Boxplots of SLI Drug Calculation Marks for Different MS Groups: those who didn't use maths support (no maths support), those who attended the FN workshop, those who attended the SLI drop in and those who attended both the FN workshop and the SLI drop-in

In Figure 2 the horizontal lines within each box represent medians, the ends of each box are the upper and lower quartiles and the whiskers are drawn at the maximum and minimum value (excluding outliers defined as lying beyond 1.5 box lengths above or below the box with ordinary outliers denoted by circles and extreme outliers, more than 3 box lengths away from the box, denoted by asterisks) for each group (Pallant, 2010). There is a notable absence of low scoring outliers amongst students who attended the SLI drop-in. This supports Wright's (2008) finding that specialist skills intervention raised the minimum score on a numeracy test compared to a control or standard group.

Referring again to Table 1 it is difficult to compare group scores because the widely varied group sizes and potential associated heteroscedasticity (unequal variance across groups) presents a problem for comparing means using one-way ANOVA. Surprisingly, however, there is no violation of the homogeneity of variance assumption as Levene's test yields $p = 0.342$. However, normality tests indicate that the two largest groups do not belong to a normal distribution.

Assuming ANOVA is robust to departures from normality, the one-way ANOVA yields $p = 0.653$. For exponential transformed data the result is $p = 0.686$ so there is some agreement that there is no difference in group means.

For two of the comparison groups though the kurtosis to kurtosis standard error ratio (Gilchrist and Samuels, 2014) is greater than 2 so the robust test of equality of means must be consulted instead, of which the Welch test is deemed the more appropriate in this case.

The Welch test yields p = 0.006 and this contradicts the Kruskall-Wallis test result (p=0.187) and the earlier noted ANOVA results. An innocuous looking note in the SPSS output observes that group sizes are unequal and type 1 error levels are not guaranteed and, indeed, there is also disagreement between the robust tests of equality of means with the Brown-Forsythe test yielding p=0.391. It would seem reasonable to say then that there is some doubt over what conclusion can be drawn from the data and test results or it can be said that the Welch result either indicates a difference in means or represents a type 1 error.

Minitab power and sample size calculator indicates that larger numbers (n=143) are needed in each comparison group in order, for example, to be able to detect a difference in means of 5.0, assuming a test score standard deviation of 12 based on the present data set, and the figures for the non-parametric alternative are inevitably going to be higher. This would also cast doubt on any firm conclusion being drawn with smaller and unequal groups.

In one sense, this data is suggestive of ineffective interventions but if the aim is to ensure parity of performance across groups then there is perhaps scope to say that the second intervention was useful to students.

4. Conclusions

Students re-sitting a basic numeracy test improved their performance but more so, and to the level of their peers, if they also engaged in one to one or small group maths support.

Foundation numeracy workshops and drug calculations drop-in sessions seem to have a cumulative beneficial effect.

A drug calculations drop-in prior to formative assessment largely eliminated low scores for drop-in students in comparison to their peers.

E-assessment has broadly been a leap forward for students and promises more in the future (Sabin et al, 2013) but isn’t a panacea; there is still scope for a maths support tutor to add value.

Whilst there is stronger evidence for the first intervention having impact the second intervention is possibly more worthy of pursuit, since it is more aligned to calculations skills development. However, the latter would need a larger scale sample to confirm efficacy.

Logarithmic data transformation, indeed any of the recommended transformations, is problematic for paired differences that include zero but would still seem to yield a result that can lend support to conclusions based on non-parametric test results.

5. Acknowledgements

The author acknowledges the excellent, friendly and pragmatic assistance of Kate Goodhand, Marina Ritchie and colleagues in the School of Nursing and Midwifery.

6. Appendix

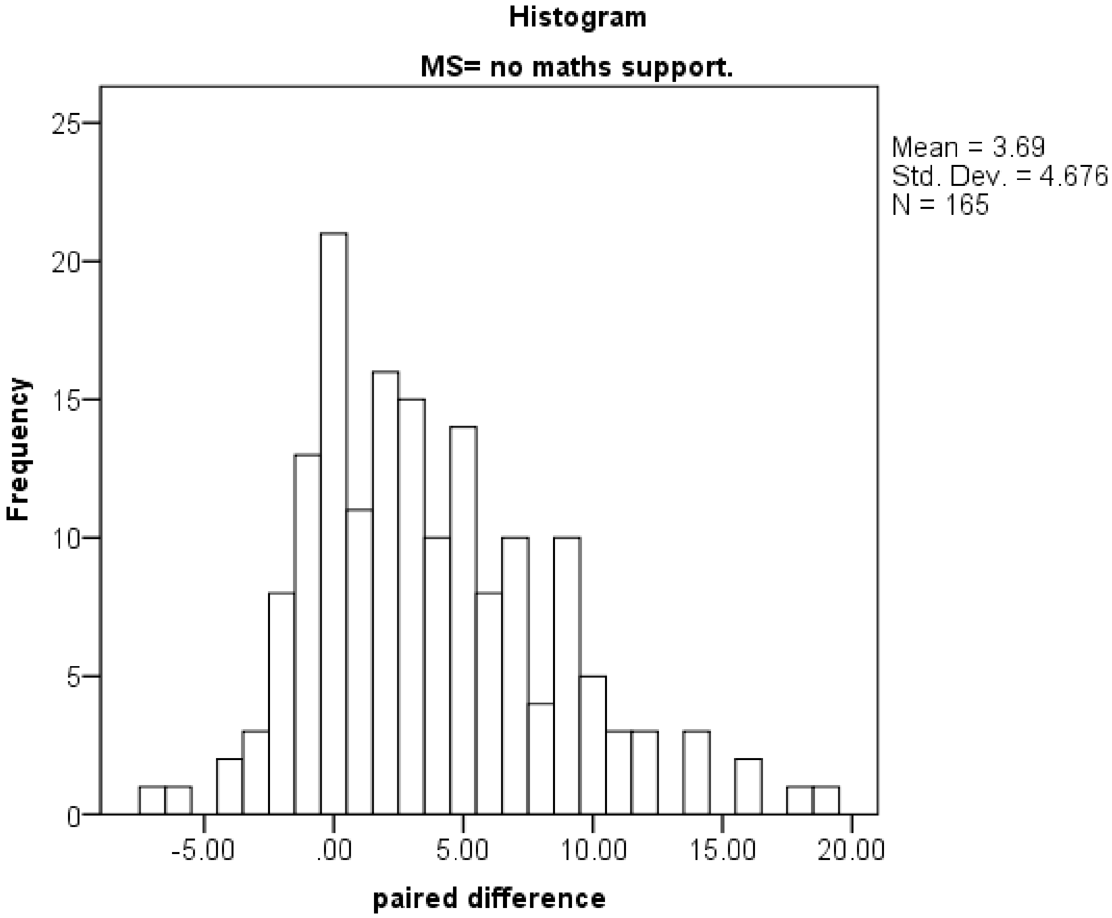

Figure 3. Histogram of Assessment Marks Paired Differences

7. References

Dunlosky, J. and Hartwig, M., 2012. Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychon Bull Rev* 19:126–134. Available at: http://dx.doi.org/10.3758/s13423-011-0181-y [Accessed 5 January 2017]

Gilchrist, M. and Samuels, P., 2014. *One Way ANOVA – Additional Material*. [pdf] Loughborough and Coventry: statstutor. Available at: http://www.statstutor.ac.uk/resources/uploaded/anovaadditional2.pdf [Accessed 8 January 2018]

IBM Support**,** 2018. *Transforming Variable to Normality for Parametric Statistics*. IBM: New York. Available at: http://www-01.ibm.com/support/docview.wss?uid=swg21479677 [Accessed 8 January 2018]

Little, J., 2006. Online support for medicine dosage calculations. *British Journal of Nursing* 5(21), pp. 1192-95. Available at: http://dx.doi.org/10.12968/bjon.2006.15.21.22380 [Accessed 27 May 2016]

Nursing and Midwifery Council, 2015a. *The NMC Code of Professional Conduct: Standards for Conduct, Performance and Ethics*. NMC: London. Available at: https://www.nmc.org.uk/standards/code/ [Accessed 27 May 2016]

Nursing and Midwifery Council, 2015b. *Standards for medicines management*. [pdf] NMC: London. Available at: https://www.nmc.org.uk/globalassets/sitedocuments/standards/nmc-standards-for-medicines-management.pdf [Accessed 1 July 2016]

Pallant, J., 2010. *SPSS* Survival *Manual 3rd Ed*. McGraw Hill / Open University Press: Maidenhead.

Sabin, M. et al., 2013. Safety in numbers 5: Evaluation of computer-based authentic assessment and high fidelity simulated OSCE environments as a framework for articulating a point of registration medication dosage calculation benchmark. *Nurse Education Today* 13(2): 55-65. Available at: http://dx.doi.org/10.1016/j.nepr.2012.10.009 [Accessed 5 January 2017]

Shier, R., 2004. *Statistics: Paired t tests*. [pdf] Loughborough and Coventry: statstutor. Available at: http://www.statstutor.ac.uk/resources/uploaded/paired-t-test.pdf [Accessed 8 January 2018]

Wright, K., 2008. Can effective teaching and learning strategies help student nurses to retain drug calculation skills? *Nurse Education Today*. 28(7): 856–864. Available at: http://doi:10.1016/j.nedt.2008.01.002 [Accessed 2 June 2016]

Zumbo, B. D. and Jennings, M. J., 2002. The Robustness of validity and efficiency of the related samples t-Test in the presence of outliers, *Psicológica*, 23(2), pp. 415-450.