

---

2019

## The Impact of Calculators on a Test of Clinician Numeracy: A Randomized Controlled Trial

Anne A. Taylor

*School of Medicine, Keele University, a.taylor1@keele.ac.uk*

Douglas R. Corfield

*University of Manchester, Cheshire, England, doug.corfield@manchester.ac.uk*

Lucie M. Byrne-Davis

*University of Manchester, Cheshire, England, lucie.byrne-davis@manchester.ac.uk*

Follow this and additional works at: <https://scholarcommons.usf.edu/numeracy>



Part of the [Medical Education Commons](#), [Nursing Commons](#), and the [Social and Behavioral Sciences Commons](#)

---

### Recommended Citation

Taylor, Anne A., Douglas R. Corfield, and Lucie M. Byrne-Davis. "The Impact of Calculators on a Test of Clinician Numeracy: A Randomized Controlled Trial." *Numeracy* 12, Iss. 2 (2019): Article 6. DOI: <https://doi.org/10.5038/1936-4660.12.2.6>

Authors retain copyright of their material under a [Creative Commons Non-Commercial Attribution 4.0 License](#).

---

# The Impact of Calculators on a Test of Clinician Numeracy: A Randomized Controlled Trial

## Abstract

Clinician numeracy (CN), the ability to use and understand quantitative data in patient care, is an important skill for healthcare professionals. Nonetheless, it is recognized that many healthcare professionals, including doctors, have deficiencies in CN, and that this may affect patient safety. In our previous research using the Medical Interpretation and Numeracy Test (MINT), we found that many doctors in training in the UK had low CN. However, participants were not permitted to use calculators when taking the MINT, even though staff has access to calculators in clinical practice. Therefore, our original study may have underestimated doctors' CN, compared to their ability in clinical practice.

We designed a randomized controlled trial to assess the impact of calculators on MINT score. We recruited 110 third-year medical students to participate in the study. Our results show that having access to a calculator had no impact on test scores. We consider that this is due to two factors: (1) CN is a complex construct that involves problem-solving and analysis, skills that are not improved by using calculators; (2) errors made by participants in our study are predominantly errors of understanding rather than calculation errors. We suggest that participants taking CN tests should have access to calculators as they would do in the workplace. We recognize that further research is needed, but suggest that educational interventions to improve CN should primarily be directed at improving understanding of mathematical concepts rather than focusing on calculation skills.

## Keywords

clinician numeracy, physician numeracy, health numeracy, numeracy assessment, Medical Interpretation and Numeracy Test

## Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial 4.0 License](https://creativecommons.org/licenses/by-nc/4.0/)

## Cover Page Footnote

Anne Taylor is a hospital dean at the School of Medicine, Keele University. She is involved in undergraduate and postgraduate medical education, and researches numeracy in medical practice.

Doug Corfield is Professor of Medical Sciences and Head of the Division of Medical Education, University of Manchester where he is Programme Director for the MBChB. He researches the neurophysiological control of breathing in humans.

Lucie Byrne-Davis is a senior lecturer in assessment and psychometrics at Division of Medical Education, University of Manchester. She researches assessment of constructs related to health professional education, training and practice.

## Introduction

Clinician numeracy (CN) is the ability of healthcare professionals “to use numbers and numeric concepts in the context of taking care of patients” (Caverly et al. 2012). CN is important across the spectrum of clinical work for doctors, from routine tasks such as calculating drug doses to medical decision making. It is thus essential to patient safety (Lesar et al. 1997; Hughes and Edgerton 2005; Gigerenzer et al. 2007; Coben and Weeks 2014; Williams and Walker 2014).

There is evidence, however, that many medical students and doctors have difficulty in calculating drug doses (Rowe et al. 1998; Selbst et al. 1999; Wheeler et al. 2007; Simpson et al. 2009; Harries and Botha 2013). There is also evidence that they may struggle to understand medical data underpinning clinical treatment options (Gigerenzer et al. 2007; Windish et al. 2007; Rao and Kanter 2010; Gigerenzer and Gray 2011; Moyer 2012; Johnson et al. 2014; Malhotra et al. 2015).

That medical students and doctors may struggle with drug dose calculation is clinically important, because medication errors are common and a significant cause of morbidity and mortality worldwide. It is estimated that there are approximately 240 million medication errors annually in the NHS in England (Elliott et al. 2018), while adverse drug events are estimated to cost almost \$20 billion annually in the US (da Silva and Krishnamurthy 2016). The World Health Organisation (WHO) has launched a global challenge to reduce the incidence of medication-related harm by 50% over five years (WHO 2017).

Drug dose calculation errors are a cause of medication error and have been researched extensively in the nursing literature (Johnson and Johnson 2002; Hutton et al. 2010; McMullan et al. 2010; Wright 2010; Sabin et al. 2013; McDonald et al. 2013; Weeks et al. 2013a, b, c; Young et al. 2013; Coben and Weeks 2014; Fleming et al. 2014; Bagnasco et al. 2016). However, there has been little research on drug dose calculation skills of medical students and doctors, perhaps due to the assumption that entry to medical school assures good numeracy (Rowe et al. 1998; Simpson et al. 2009; Harries and Botha 2013).

### ***Assessing Clinician Numeracy***

Assessment of CN in medical students would have many potential uses, including selection, formative assessment to identify areas of learning difficulty, and summative assessment for progression decisions. Assessing CN would be particularly salient if low CN was associated with difficulties in clinical practice. In order to measure CN in medical students and doctors, we previously developed an assessment of CN, the Medical Interpretation and Numeracy Test (MINT). The MINT is a 43-item assessment with questions testing computational, analytical, and statistical numeracy (Taylor and Byrne-Davis 2016). Our research using

MINT adds to the evidence demonstrating that medical students and doctors may have deficiencies in CN (Taylor and Byrne-Davis 2017).

The participants in our earlier study, however, did not have access to calculators, so it is possible that our finding of low CN on the MINT might not translate to difficulties in clinical practice, where calculators are readily available.

Not all numeracy questions require calculators. Close et al. (2008) classified numeracy questions as calculator-appropriate (complex calculations), calculator-optional (where it is unnecessary but not unreasonable to use a calculator), and calculator-inappropriate (simple calculations that can be answered readily either mentally or with pen and paper). Questions that are important in determining overall CN, such as data interpretation questions, would also be classified as calculator-inappropriate. In our previous research with the MINT, we considered that calculators would be unnecessary, as its content was largely calculator-inappropriate. Calculators would not help with analytical questions, involving the interpretation of data presented in charts and graphs, for example, or with statistical questions, testing clinical mathematical reasoning. Furthermore, most computational questions in the MINT were straightforward, and based on numbers that would be easy to manipulate either mentally or using pen and paper. On this basis, we originally classified only one of our 43 questions, a complex calculation, as calculator-appropriate.

On the other hand, we recognise that there may be a significant overlap between the various numeracy constructs, and the idea that questions classified as primarily “analytical” or “statistical” may also have significant computational elements (Golbeck et al. 2005). Therefore, our original classification of questions as calculator-appropriate or not may have been inaccurate: many MINT questions, whether computational, analytical or statistical, involve multiple steps and calculations, and so they could be considered to be either calculator-optional or calculator-appropriate.

We reviewed our test material, classifying all 18 computational questions, along with three analytical and two statistical questions as either calculator-appropriate or calculator-optional. We considered the remaining 20/43 questions to be calculator-inappropriate. Therefore, the lack of access to calculators in our initial research with the MINT may have resulted in an underestimate of CN in doctors compared to the real-life clinical situation where calculators are readily available. Thus, in that case, our previous finding of low CN in doctors would be less relevant to clinical practice, and could also mean that the MINT had lower construct validity.

To test the hypothesis that using calculators would improve MINT scores, we conducted a randomized controlled trial of the effect of calculators on clinician numeracy, comparing MINT scores in medical students randomly allocated to having or not having a calculator.

## **Methods**

### ***Study design***

The study was a randomized controlled trial. Participants were randomly allocated into one of two groups: group C, who received calculators, and group N, who did not. Ethical approval for the study was obtained from the University of Manchester (UoM) Research and Ethics Committee (UREC).

### ***Participants***

Participants were third-year medical students studying at a single institution in England. The MINT was incorporated into the formative mid-year assessments for these students. All students in the year group were eligible for entry to the study. One month prior to the formative assessment, these students attended a teaching session on clinician numeracy and its importance for healthcare professionals and were given preliminary information about the study. Further information and an invitation to participate in the research were sent by email.

### ***Interventions***

The Medical Interpretation and Numeracy Test (MINT) is an assessment of clinician numeracy, consisting of 43 questions, testing computational, analytical and statistical constructs; it has high internal consistency reliability as measured by a Cronbach's alpha score of 0.868 (Taylor and Byrne-Davis 2016). The MINT is available as a multiple-choice test, and in a short answer (constructed response) format. For this study, we used the constructed response format.

### ***Outcomes***

The outcome measure was the mean score of participants in groups N and C. We also measured the facility of each test item for participants in groups N and C.

### ***Sample size***

In order to calculate the sample size required for the trial, we considered the previous mean and standard deviation of the MINT in similar participant groups. The mean MINT score achieved by participants in a previous study was 32.76/43 with a standard deviation of 6.64 (Taylor and Byrne-Davis 2017). We considered that a change of up to 2 marks (less than 5 percentage points) might represent normal variation (a "good" vs. "bad" day for an individual), but that a change in score of 4 marks (almost 10 ppt) would demonstrate that an intervention had had a positive effect. With a minimum difference to be detected of 4 marks, and a standard deviation of 6.64, a type 1 error rate of 0.05, and a type 2 error of 0.2, we calculated that 88 participants (44 in each group) would be required (online tool for sample size calculation: Brant n.d.).

## **Randomization**

We invited 116 students to participate in the study; therefore, a list of potential participants was made, with study identification (ID) numbers from 1 to 116. A table of random numbers was used to allocate the ID numbers into study groups C (calculator) and N (no calculator). Test answer sheets were prepared, and recorded the study ID number and group allocation code “C” or “N”.

**Allocation concealment.** The test answer sheets were placed in a brown A4 envelope, alongside the MINT paper, a pencil and an eraser. Basic pocket calculators were added to test envelopes for test papers coded “C”. All envelopes were sealed. Because the calculators were small and flat, envelopes containing calculators appeared similar to those containing only a pencil and eraser. The study envelopes were randomly distributed on desks in the examination room, and participants were allowed to select their own seats. Therefore, neither the researcher nor the participants were aware of group allocation until the test commenced, and participants opened their envelopes.

**Implementation.** The test was carried out under examination conditions, with 90 minutes to complete the test. Once the test was completed, participants returned all test materials to the study envelopes. Participants were aware of the hypothesis that using a calculator would improve test score; those allocated to group N were given the option to request a calculator. When students opted to change their allocation, the coding on their answer sheets was changed accordingly, and this change was recorded.

## **Statistical Methods**

Data were analysed in Microsoft EXCEL, and an online statistical tool (MedCalc Software bvba (BE) a, b). We described the distribution of scores for each group, and then used Student’s *t*-test to compare the means of participants in the two main study groups (N and C); the primary analysis relates to the intent-to-treat group allocation. We also analysed data relating to the final (per-protocol) group allocations.

We assessed the magnitude of the difference associated with use of a calculator by calculating the effect size. Because the comparison is of mean test scores, it is more appropriate to calculate the absolute effect size rather than using an effect size index (Sullivan and Feinn 2012).

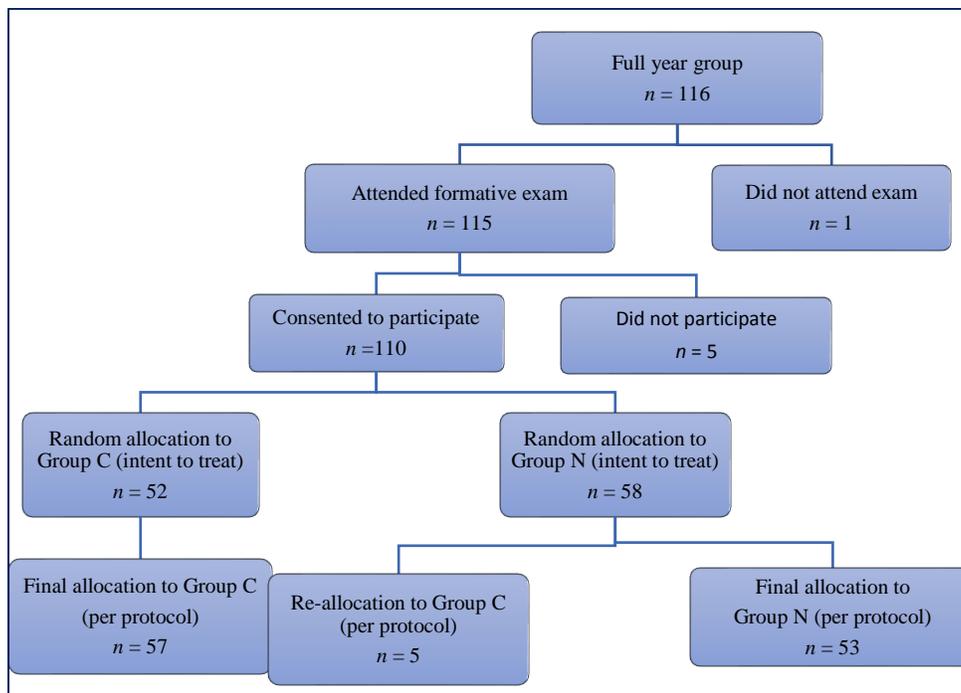
**Subgroup analyses.** Participants were asked to indicate their gender, as there is evidence from a study investigating quantitative literacy in US university students that female gender may be associated with lower numeracy (Sikorskii et al. 2011). Furthermore, we asked participants whether they had dyslexia because there is some overlap between dyslexia and dyscalculia (Gibson and Leinster 2011;

British Dyslexia Association 2017). We recorded these data to ascertain whether these attributes were evenly distributed across groups, and if not, to ensure that any effects did not confound observed differences between calculator and non-calculator groups. We performed a logistic regression analysis to assess any apparent effect relating to these characteristics.

## Results

### *Participants*

Of 116 third-year students, 110 (95%) consented to participate in the study. 52/110 (47%) of students were allocated to Group C (calculators), while 58/110 (53%) participants were allocated to Group N (no calculators). Five students who had been allocated to Group N requested calculators, and so were reassigned to Group C. Recruitment of participants and allocation to study groups is shown in Figure 1.



**Figure 1.** Recruitment and allocation of study participants.

## Demographic data

Of the 110 participants, 59 (54%) were female and 36 (33%) were male, and 15 (13%) did not declare their gender. Twelve students (11%) declared a diagnosis of dyslexia, 75 (68%) stated that they were not dyslexic, and 23 (21%) did not report their dyslexia status (Table 1). (The 23 students who did not comment on their dyslexia status include all 15 who did not indicate their gender).

**Table 1**  
**Demographic Data of Study Groups**

	<i>N</i>	Male <i>n</i> (%)	Female <i>n</i> (%)	Unknown Gender <i>n</i> (%)	Dyslexia <i>n</i> (%)	No dyslexia <i>n</i> (%)	Unknown dyslexia <i>n</i> (%)
<b>Total</b>	110	36 (33%)	59 (53%)	15 (14%)	12 (11%)	75 (68%)	23 (21%)
<b>Group C*</b>	52	14 (27%)	30 (58%)	8 (15%)	4 (8%)	36 (69%)	12 (23%)
<b>Group N*</b>	58	22 (38%)	29 (50%)	7 (12%)	8 (14%)	39 (67%)	11 (19%)

\**intent to treat*

## Mean scores

Test scores for all study groups are shown in Table 2, which includes scores for the full cohort of 110 participants as well as scores of participants in different groups. Although the performance of all groups was similar, the mean scores of those who had calculators were higher than mean scores of those without calculators. However, statistical analysis using Student's *t*-test to compare the mean scores of participants in different groups indicated that the apparent difference in scores was not significant. The primary analysis is based on intention to treat, and thus represents participants whose original allocations were to groups C ( $n=52$ ) and to group N ( $n=58$ ). There was no statistical difference in performance of participants in these groups ( $difference = 1.9$ ;  $SE = 0.98$ ;  $CI_{95\%} = -0.05 - 3.8$ ;  $t = 01.9$ ;  $DF = 108$ ;  $Sig = ns$ ).

**Table 2**  
**Test Score: All Groups**

	<i>N</i>	Mean ( <i>SD</i> )	Median	Range	IQR
<b>All</b>	110	31.8 (5.2)	33	19-43	29-35
<b>Group C</b> (intent to treat)	52	32.8 (5.1)	33	19-43	30-36
<b>Group N</b> (intent to treat)	58	30.9 (5.2)	33	19-41	28-34
<b>Group C</b> (per-protocol)	57	32.5 (5.2)	33	19-43	29-36
<b>Group N</b> (per-protocol)	53	31.1 (5.3)	33	19-41	28-35

\*interquartile range

Table 1 also provides data relating to the per-protocol group allocations: five students allocated to group N requested calculators and so were re-allocated to group C. Again, there was no statistical difference between the C and N groups ( $difference = 1.4$ ;  $SE = 1.0$ ;  $CI_{95\%} = -0.58 - 3.38$ ;  $t = 1.39$ ;  $DF = 108$ ;  $Sig = ns$ ).

There was also no difference in performance of those whose original and final allocations were to group C (*difference* = 0.3; *SE* = 0.98; *CI*<sub>95%</sub> = -1.66 – 2.25; *t* = 0.3; *DF* = 107; ns); or to group N (*difference* = 0.2; *SE* = 0.99; *CI*<sub>95%</sub> = -1.77 – 2.2; *t* = 0.201; *DF* = 109; *Sig* = ns).

### **Effect Size**

The absolute effect size is the difference in the mean scores of Groups N and C. Data were analysed using the intent-to-treat groups; thus the absolute effect size was 1.9.

### **Facility of test items**

In addition to analysing the mean scores of participants in groups C and N, we compared performance on individual MINT items to assess whether use of a calculator conferred an advantage for individual questions. Raw data show that facility was the same for 4/43 questions, was higher in Group C for 29/43 questions and higher in Group N for 10/43 questions (Table 3). We used the N-1 Chi-squared test to assess whether these differences were significant. Because this involved conducting 43 individual tests, it was necessary to apply the Bonferroni correction (Perneger 1998); therefore, significance  $p < 0.05/43$ , i.e. a difference was significant at the 5% level only if  $p < 0.001$ . We found a statistically significant difference in performance in 2/43 questions: in both cases, participants in group C performed better than those in group N. Both questions were computational.

### **Subgroup Analyses**

The mean score of male participants was 34.4/43, while that of females was 30.5/43; therefore, the effect size is 3.9/43. The mean score of participants with dyslexia was 29.3/43, and that for those who were not dyslexic was 32.4/43; the effect size is 3.1/43. Logistic regression analysis indicated that the difference related to gender was significant (Table 4).

## **Discussion**

We found that having a calculator did not affect overall scores on our test of clinician numeracy, the MINT. The mean score of Group C was slightly higher than that of Group N. The difference was not statistically significant. The absolute effect size was 1.9, i.e. participants in Group C had a mean score of 1.9/43 (4.4 ppt) higher than participants in Group N. We do not think that this is clinically important. The study was designed to detect an effect size of 4 (9.3 ppt) (see Methods).

**Table 3**  
**Facility of Test Items: Group C v Group N (*Intent to Treat*)**

Q nbr	Primary construct	Facility (proportion correct)			Diff <sup>ψ</sup> (ppt)	95% CI	χ <sup>2</sup>	Sig
		All n = 110	Group C n = 52	Group N n = 58				
25	Computational	.96	.96	.97	-1		ns	
20	Computational	.95	.96	.93	3		ns	
1	Computational	.92	.96	.88	8		ns	
16	Computational	.88	1.0	.78	22	11 - 34	12.6	0.0004*
4	Computational	.85	.90	.81	9			ns
22	Computational	.85	.88	.83	5			ns
12	Computational	.84	.88	.81	7			ns
35	Computational	.81	.87	.74	13			ns
21	Computational	.79	.79	.79	0			ns
43	Computational	.78	.83	.74	9			ns
5	Computational	.70	.65	.74	-9			ns
29	Computational	.67	.65	.69	-4			ns
30	Computational	.64	.77	.53	24	6 - 40	6.8	0.009**
19	Computational	.62	.62	.62	0			ns
40	Computational	.57	.56	.55	1			ns
33	Computational	.54	.73	.36	37	18 - 52	15	0.0001*
2	Analytical	1.0	1.0	1.0	0			ns
28	Analytical	.91	.92	.90	2			ns
31	Analytical	.89	.90	.88	2			ns
34	Analytical	.88	.90	.86	4			ns
42	Analytical	.84	.85	.81	4			ns
18	Analytical	.83	.85	.81	4			ns
23	Analytical	.83	.81	.84	-3			ns
39	Analytical	.83	.88	.78	10			ns
8	Analytical	.69	.69	.69	0			ns
32	Analytical	.64	.73	.55	18			ns
13	Analytical	.60	.63	.57	6			ns
37	Analytical	.54	.56	.50	6			ns
6	Analytical	.52	.48	.55	-7			ns
24	Analytical	.39	.42	.36	6			ns
38	Statistical	.98	.96	1.0	-4			ns
41	Statistical	.96	.96	.97	-1			ns
7	Statistical	.93	.94	.91	3			ns
36	Statistical	.93	.92	.91	1			ns
17	Statistical	.88	.90	.86	4			ns
10	Statistical	.87	.87	.88	-1			ns
26	Statistical	.84	.85	.83	2			ns
9	Statistical	.77	.83	.71	12			ns
3	Statistical	.58	.52	.64	-12			ns
11	Statistical	.44	.44	.41	3			ns
14	Statistical	.44	.42	.45	-3			ns
15	Statistical	.27	.31	.24	7			ns
27	Statistical	.25	.27	.24	3			ns

\*  $p < 0.001$ , therefore, significant at 5% level when the Bonferroni correction is applied.

\*\*  $p > 0.001$ , therefore, not significant at 5% level when the Bonferroni correction is applied.

<sup>ψ</sup> positive value indicates C > N; negative value indicates N > C.

**Table 4**  
**Logistic Regression Analysis of Subgroups**

Independent variable	<i>b</i>	<i>SE</i>	<i>T</i>	<i>Prob</i>
Calculator	1.7	1.01	1.7	.091
Dyslexia	-3.2	1.638	-1.94	0.56
Gender	-3.9	1.045	-3.75	0.000

We found little difference in terms of performance on individual questions. We had considered that using a calculator might improve performance on the 16 computational questions in the test; in addition, we classified three analytical and three statistical questions as either calculator-appropriate or calculator-optional, and so there were 22/43 questions where having a calculator might prove beneficial. However, participants who had calculators performed better on only two questions: one was a complex calculation; the other simply required calculating the mean of four values.

Because research on the use of calculators in tests of CN is limited to tests of drug dose calculation in nursing, with small study samples, it is difficult to compare our results to the existing literature. However, the evidence from nursing studies is unsettled: some researchers found that using calculators improved performance (Shockley et al. 1989; Bliss-Holtz 1994), while others observed little or no impact (Murphy and Graveley 1990; Tarnow and Werst 2000). Interestingly, there is some debate in the nursing literature about whether to permit the use of calculators in drug dose calculation tests; for example, McMullan et al. (2010) argue that calculators should not be allowed as they would constitute “a substitute for arithmetical knowledge and skills.” However, we consider that medical students, doctors, and other healthcare staff taking drug dose calculation tests and other tests of CN should be allowed to use calculators, because they are readily available in clinical practice. Furthermore, our results suggest that using calculators will not conceal evidence of low CN.

Our finding that using calculators did not have a positive impact on test scores supports the observation that CN is a complex construct that entails more than the ability to perform simple mathematical operations. The complexity of CN is highlighted by Coben and Weeks (2014), who note that numeracy in nursing practice requires being “competent, confident, and comfortable with one’s judgments on whether to use mathematics in a particular situation and if so, what mathematics to use, how to do it, what degree of accuracy is appropriate, and what the answer means in relation to the context.” Another nursing study describes four distinct areas of competence necessary for accurate drug dose calculation (“the 4 Cs”): computation, conceptualisation, conversion, and critical analysis (Johnson and Johnson 2002). Therefore, multiple skills are necessary for competence in CN and safe clinical practice: these skills are needed not only for drug dose calculation, but also for clinical tasks involving data interpretation,

including basic statistical analysis. Clearly, these skills are required by medical students and doctors as well as nursing students and nurses.

Our results may provide some insight into the type of errors being made by doctors and medical students in the MINT. We consider that these errors may relate to one or more of the “4 Cs”. Research in nursing practice has shown that using calculators reduces the incidence of computational errors, but has no impact on conceptual errors (Murphy and Graveley 1990; Bliss-Holtz 1994). Our finding that the two questions for which calculators improved performance were computational accords with this literature. Similarly, our finding that calculators did not influence mean test scores may indicate that participants are primarily making conceptual rather than computational errors. Furthermore, errors may occur when converting between different units of measurement: Wheeler et al. (2004, 2007) note that doctors commonly make such errors in drug dose calculation. Finally, participants in our study may not have critically analysed their answers to assess whether they were likely to be correct. There is evidence that errors made by bioscience students (Tariq 2008) and nursing students (Galligan and Hobohm 2015) in numeracy tests are often due to failure to cross-check their answers; therefore, it is likely that medical students also make this type of error. Determining the type of error being made is an important step in developing appropriate educational intervention, because successful remediation requires that the intervention is targeted at the area of weakness (Wallace 2019). Further research is needed in this area.

The lack of impact of calculators on MINT scores in this study reinforces our original observation that some doctors have low CN. This finding is important in relation to patient safety, as errors in drug dose calculation and in data interpretation may lead to serious patient harm (Lesar et al. 1997; Hughes and Edgerton 2005; Gleason et al. 2010; Gigerenzer and Gray 2011; Abramson et al. 2012; Seden et al. 2013; Vincent et al. 2014; Williams and Walker 2014; Malhotra et al. 2015). Moreover, the finding that calculators do not overcome apparent deficiencies in CN is supported by the observation that the introduction of electronic prescribing has had less impact on the prevalence of medication errors than was initially anticipated (Tully 2012; Ahmed et al. 2016). Further work is required to elucidate how and why doctors and medical students make errors in tests of CN, as this may have implications for their clinical practice and their education.

We asked participants to report on gender because there is evidence that female gender may be associated with lower numeracy (Sikorskii et al. 2011; Stoet and Geary 2013; Bagnasco et al. 2016). However, a study by Bridgeman et al. (1995) and a large meta-analysis by Lindberg et al. (2010) found no difference in mathematical ability related to gender; nonetheless, Lindberg et al. (2010) found strong evidence of stereotyping girls and women as being inferior at

mathematics. In our study, participants identifying as male performed better than those identifying as female. We consider that further research into the association of gender and CN in medical students and doctors could help tease apart different CN constructs to see if some of the different findings are related to gender effects on different aspects of numeracy.

We recorded dyslexia because of the overlap between dyslexia and dyscalculia (Gibson and Leinster 2011; British Dyslexia Association 2017). We found no statistically significant difference in performance of participants with dyslexia, compared to non-dyslexic participants.

### **Limitations**

All participants in this study were from a single medical school, and so effects might be related to the context of the course itself, although this is unlikely due to the random allocation between groups. Furthermore, drug dose calculation is a complex task, for which several distinct competencies are required, and we have explored only one area of competence. Nonetheless, we consider that our findings provide insight into the type of numeracy errors made by doctors and medical students, and may be valuable in terms of determining the direction of educational intervention to remediate drug calculation error.

### **Conclusion**

Using calculators did not affect overall MINT score. We consider that this outcome may be related to two key factors: first, because a large proportion of the test material can be classified as either calculator-inappropriate or calculator-optional, a calculator would not be expected to confer any benefit; and second, our findings suggest that the errors being made in the MINT are not remediable by using calculators, i.e., the errors are conceptual rather than arithmetical. This finding has implications for educational intervention to reduce drug calculation errors in doctors and medical students.

### **Acknowledgments**

We would like to thank the reviewers of this manuscript for their constructive feedback. We are very grateful to Professor Len Vacher, and to colleagues at the School of Medicine, Keele University for their support and encouragement of our research.

### **References**

Abramson, E. L., D. W. Bates, C. Jenter, L. A. Volk, Y. Barron, J. Quaresimo, A. C. Seger, E. Burdick, S. Simon, and R. Kaushal. 2012. "Ambulatory

- Prescribing Errors among Community-based Providers in Two States.” *Journal of the American Medical Informatics Association* 19:644–8. <https://doi.org/10.1136/amiajnl-2011-000345>.
- Ahmed, Z., S. Garfield, Y. Jani, S. Jheeta, and B. D. Franklin. 2016. “Impact of Electronic Prescribing on Patient Safety in Hospitals: Implications for the UK.” *Clinical Pharmacist* 8(5): 1–11. <https://doi.org/10.1211/CP.2016.20201013>.
- Bagnasco, A., L. Galaverna, G. Aleo, A. M. Grugnetti, F. Rosa, and L. Sasso. 2016. “Mathematical Calculation Skills Required for Drug Administration in Undergraduate Nursing Students to Ensure Patient Safety: A Descriptive Study.” *Nurse Education in Practice* 16: 33–39. <https://doi.org/10.1016/j.nepr.2015.06.006>.
- Bliss-Holtz, J. 1994. “Discriminating Types of Medication Calculation Errors in Nursing Practice.” *Nursing Research* 43(6): 373–375. <https://doi.org/10.1097/00006199-199411000-00010>.
- Brant, R. n.d. *Inference for Means: Comparing Means for Two Independent Samples*. <http://www.stat.ubc.ca/~rollin/stats/ssize/n2.html>. [accessed 15 April 2018]
- Bridgeman, B., A. Harvey, and J. Braswell. 1995. “Effects of Calculator Use on Scores on a Test of Mathematical Reasoning.” *Journal of Educational Measurement* 32(4): 323–340. <https://doi.org/10.1111/j.1745-3984.1995.tb00470.x>.
- British Dyslexia Association. 2017. “Dyslexic Definitions.” <http://www.bdadyslexia.org.uk/dyslexic/definitions>.
- Caverly, T. J., A. Prochazka, I. Binswanger, J. S. Kutner, and D. Matlock. 2012. “Getting the Gist of Health Risks.” Section on Statistical Education—JSM. <http://www.statlit.org/pdf/2012-Caverly-ASA.pdf>.
- Close, S., E. Oldham, P. Surgenor, G. Shiel, T. Dooley, and M. O’Leary. 2008. “The Effects of Calculator Use on Mathematics in Schools and in Certificate Examinations,” Final report on phase 2. Dublin: St. Patrick’s College, Trinity College and Educational Research Centre. [http://www.erc.ie/documents/calculator\\_final\\_report\\_phase2.pdf](http://www.erc.ie/documents/calculator_final_report_phase2.pdf).
- Coben, D., and K. Weeks. 2014. “Meeting the Mathematical Demands of the Safety-Critical Workplace: Medication Dosage Calculation Problem-solving for Nursing.” *Educational Studies in Mathematics* 86: 253–270. <https://doi.org/10.1007/s10649-014-9537-3>.
- da Silva, B. A., and M. Krishnamurthy. 2016. “The Alarming Reality of Medication Error: A Patient Case and Review of Pennsylvania and National Data.” *Journal of Community Hospital Internal Medicine Perspectives* 6: 31758. <https://doi.org/10.3402/jchimp.v6.31758>.

- Elliott, R.A., Camacho, E., Campbell, F., Jankovic, D., Martyn St James, M., Kaltenthaler, E., Wong, R., Sculpher, M.J., and Faria, R. 2018. "Rapid Evidence Synthesis and Economic Analysis of the Prevalence and Burden of Medication Error in the UK." Available from: <http://www.eepru.org/wp-content/uploads/2018/02/eepru-report-medication-error-feb-2018.pdf>.
- Fleming, S., A. M. Brady, and A. M. Malone. 2014. "An Evaluation of the Drug Calculation Skills of Registered Nurses." *Nurse Education in Practice* 14(1): 55–61. <https://doi.org/10.1016/j.nepr.2013.06.002>.
- Galligan, L., and C. Hobohm. 2015. "Investigating Students' Academic Numeracy in 1st Level University Courses." *Mathematics Education Research Journal* 27:129–145. <https://doi.org/10.1007/s13394-014-0132-9>.
- Gibson, S., and S. Leinster. 2011. "How Do Students with Dyslexia Perform in Extended Matching Questions, Short Answer Questions and Observed Structured Clinical Examinations?" *Advances in Health Sciences Education*, 16: 395–404. <https://doi.org/10.1007/s10459-011-9273-8>.
- Gigerenzer, G., W. Gaissmaier, E. Kurz-Milcke, L. M. Schwartz, and S. Woloshin. 2007. "Helping Doctors and Patients Make Sense of Health Statistics." *Psychological Science Public Interest* 8(2): 53–96. <https://doi.org/10.1111/j.1539-6053.2008.00033.x>.
- Gigerenzer, G., and M. Gray, eds. 2011. *Better Doctors, Better Patients, Better Decisions: Envisioning Health Care 2020*. Cambridge Mass: MIT. <https://doi.org/10.7551/mitpress/9780262016032.001.0001>.
- Gleason, K. M., M. R. McDaniel, J. Feinglass, D. W. Baker, L. Lindquist, D. Liss, and G. A. Noskin. 2010. "Results of the Medications at Transitions and Clinical Handoffs (MATCH) Study: An Analysis of Medication Reconciliation Errors and Risk Factors at Hospital Admission." *Journal of General Internal Medicine* 25:441–7. <https://doi.org/10.1007/s11606-010-1256-6>.
- Golbeck, A. L., C. R. Ahlers-Schmidt, A. M. Paschal, and S. E. Dismuke. 2005. "A Definition and Operational Framework for Health Numeracy." *American Journal of Preventive Medicine* 29(4): 375–376. <https://doi.org/10.1016/j.amepre.2005.06.012>.
- Harries, C. S. and J. H. Botha. 2013. "Can Medical Students Calculate Drug Doses?" *Southern African Journal of Anaesthesia and Analgesia* 19(5): 248–251. <https://doi.org/10.1080/22201173.2013.10872934>.
- Hughes, R. G., and E. A. Edgerton. 2005. "Reducing Pediatric Medication Errors: Children Are Especially at Risk for Medication Errors." *American Journal of Nursing* 105(5):79–84. <https://doi.org/10.1097/00000446-200505000-00035>.
- Hutton, M., D. Coben, C. Hall, D. Rowe, M. Sabin, K. W. Weeks, and N. Woolley. 2010. "Numeracy for Nursing, Report of a Pilot Study to Compare Outcomes of Two Practical Simulation Tools: An Online Medication Dosage

- Assessment and Practical Assessment in the Style of Objective Structured Clinical Examination.” *Nurse Education Today* 30: 608–614.  
<https://doi.org/10.1016/j.nedt.2009.12.009>.
- Johnson, S. A., and L. J. Johnson. 2002. “The 4 Cs: A Model for Teaching Dosage Calculations.” *Nurse Educator* 27(2):79–83.  
<https://doi.org/10.1097/00006223-200203000-00011>.
- Johnson, T. V., A. Abbasi, E. D. Schoenberg, R. Kellum, L. D. Speake, C. Spiker, A. Foust, A. Kreps, C. W. M. Ritenour, O. W. Brawley, and V. A. Master. 2014. “Numeracy Among Trainees: Are We Preparing Physicians for Evidence-based Medicine?” *Journal of Surgical Education* 71(2):211–215.  
<https://doi.org/10.1016/j.jsurg.2013.07.013>.
- Lesar, T., L. Briceland, and D. Stein. 1997. “Factors Related to Errors in Medication Prescribing.” *Journal of the American Medical Association* 277: 312–7. <https://doi.org/10.1001/jama.1997.03540280050033>.
- Lindberg, S. M., J. S. Shibley Hyde, and J. L. Petersen. 2010. “New Trends in Gender and Mathematics Performance: A Meta-Analysis.” *Psychological Bulletin* 136(6): 1123–1135. <https://doi.org/10.1037/a0021276>.
- Malhotra A., D. Maughan, J. Ansell, R. Lehman, A. Henderson, M. Gray, T. Stephenson, and S. Bailey. 2015. “Choosing Wisely in the UK: The Academy of Medical Royal Colleges’ Initiative to Reduce the Harms of Too Much Medicine.” *British Medical Journal* 350: 2308.  
<https://doi.org/10.1136/bmj.h2308>.
- McDonald, K., K. W. Weeks, and L. Moseley. 2013. “Safety in Numbers 6: Tracking Pre-registration Nursing Students’ Cognitive and Functional Competence Development in Medication Dosage Calculation Problem Solving: The Role of Authentic Learning and Diagnostic Assessment Environments.” *Nurse Education in Practice* 13: e66–77.  
<https://doi.org/10.1016/j.nepr.2012.10.015>.
- McMullan, M., R. Jones, and S. Lea. 2010. “Patient Safety: Numerical Skills and Drug Calculation Abilities of Nursing Students and Registered Nurses.” *Journal of Advanced Nursing* 66(4): 891–899. <https://doi.org/10.1111/j.1365-2648.2010.05258.x>.
- MedCalc Software bvba (BE) a. “Comparison of Means Calculator.”  
[http://www.medcalc.org/calc/comparison\\_of\\_means.php](http://www.medcalc.org/calc/comparison_of_means.php).
- MedCalc Software bvba (BE) b. “Comparison of Proportions Calculator.”  
[http://www.medcalc.org/calc/comparison\\_of\\_proportions.php](http://www.medcalc.org/calc/comparison_of_proportions.php).
- Moyer, V. A. 2012. “What We Don’t Know Can Hurt Our Patients: Physician Innumeracy and Overuse of Screening Tests.” *Annals of Internal Medicine* 156(5): 392–393. <https://doi.org/10.7326/0003-4819-156-5-201203060-00015>.

- Murphy, M. A., and E. A. Graveley. 1990. "Drug Calculation Examinations: Do Calculators Make a Difference?" *Nurse Educator* 15(1):35–43. <https://doi.org/10.1097/00006223-199001000-00008>.
- Perneger, T.V. 1998. "What's Wrong with Bonferroni Adjustments." *British Medical Journal* 316:1236–8. <https://doi.org/10.1136/bmj.316.7139.1236>.
- Rao, G., and S. L. Kanter. 2010. "Physician Numeracy as the Basis for an Evidence-based Medicine Curriculum." *Academic Medicine* 85(11): 1794–1799. <https://doi.org/10.1097/ACM.0b013e3181e7218c>.
- Rowe, C., T. Koren, and G. Koren. 1998. "Errors by Pediatric Residents in Calculating Drug Doses." *Archives Disease Childhood* 79(1): 56–58. <https://doi.org/10.1136/adc.79.1.56>.
- Sabin M., K. W. Weeks, D. A. Rowe, B. M. Hutton, D. Coben, C. Hall, and N. Woolley. 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 in Practice* 13: e55–65. <https://doi.org/10.1016/j.nepr.2012.10.009>.
- Seden, K., J. J. Kirkham, T. Kennedy, M. Lloyd, S. James, A. McManus, A. Ritchings, J. Simpson, D. Thornton, and A. Gill. 2013. "Cross-sectional Study of Prescribing Errors in Patients Admitted to Nine Hospitals across North West England." *BMJ Open* 3(1): e002036. <https://doi.org/10.1136/bmjopen-2012-002036>.
- Selbst, S. M., J. A. Fein, K. Osterhoudt, and H. Wayne. 1999. "Medication Errors in a Pediatric Emergency Department." *Pediatric Emergency Care* 15(1): 1–4. <https://doi.org/10.1097/00006565-199902000-00001>.
- Shockley, J. S., W. C. McGurn, C. Gunning, E. Graveley, and D. Tillotson. 1989. "Effects of Calculator Use on Arithmetic and Conceptual Skills of Nursing Students." *Journal of Nursing Education* 28(9): 402–405. <https://doi.org/10.3928/0148-4834-19891101-06>.
- Sikorskii, A., V. Melfi, D. Gilliland, J. R. Kaplan, and S. Ahn. 2011. "Quantitative Literacy at Michigan State University, 1: Development and Initial Evaluation of the Assessment." *Numeracy* 4(2): Article 5. <https://doi.org/10.5038/1936-4660.4.2.5>.
- Simpson, C. M., G. B. Keijzers, and J. F. Lind. 2009. "A Survey of Drug-dose Calculation Skills of Australian Tertiary Hospital Doctors." *Medical Journal Australia* 190(3):117.
- Stoet, G., and D. C. Geary. 2013. "Sex Differences in Mathematics and Reading Achievement Are Inversely Related: Within- and across-nation Assessment of 10 years of PISA data." *PLoS ONE* 8(3): e57988. <https://doi.org/10.1371/journal.pone.0057988>.

- Sullivan, G. M., and R. Feinn. 2012. "Using Effect Size—or Why the P Value Is Not Enough." *Journal of Graduate Medical Education* 4(3): 279–282. <https://doi.org/10.4300/JGME-D-12-00156.1>.
- Tariq, V.N. 2008. "Defining the Problem: Mathematical Errors and Misconceptions Exhibited by First-year Bioscience Undergraduates." *International Journal of Mathematical Education in Science and Technology*. 39:7, 889–904. <https://doi.org/10.1080/00207390802136511>.
- Tarnow, K.G., and C. L. Werst. 2000. "Drug Calculation Examinations: Do Calculators Make a Difference?" *Nurse Educator*, 25(5): 213–215. <https://doi.org/10.1097/00006223-200009000-00010>.
- Taylor A. A., and L. M. Byrne-Davis. 2016. "Clinician Numeracy: The Development of an Assessment Measure for Doctors." *Numeracy* 9(1): Article 5. <https://doi.org/10.5038/1936-4660.9.1.5>.
- Taylor A. A., and L. M. Byrne-Davis. 2017. "Clinician Numeracy: Use of the Medical Interpretation and Numeracy Test in Foundation Trainee Doctors." *Numeracy* 10(2): Article 5. <https://doi.org/10.5038/1936-4660.10.2.5>.
- Tully, M. 2012. "Prescribing Errors in Hospital Practice." *British Journal of Clinical Pharmacology* 74(4): 668–675. <https://doi.org/10.1111/j.1365-2125.2012.04313.x>.
- Vincent, C., S. Burnett, and J. Carthey. 2014. "Safety Measurement and Monitoring in Healthcare: A Framework to Guide Clinical Teams and Healthcare Organisations in Maintaining Safety." *BMJ Quality & Safety* 23(8): 670–677. <https://doi.org/10.1136/bmjqs-2013-002757>.
- Wallace, D. 2019. "Parts of the Whole: Theories of Pedagogy and Kolb's Learning Cycle." *Numeracy*, 12(1): Article 17. <https://doi.org/10.5038/1936-4660.12.1.17>.
- Weeks, K. W., J.M. Clochesy, B.M. Hutton, and L. Moseley. 2013a. "Safety in Numbers 3: Authenticity, Building Knowledge & Skills and Competency Development & Assessment: The ABC of Safe Medication Dosage Calculation Problem-solving Pedagogy." *Nurse Education in Practice* 13: e33–42. <https://doi.org/10.1016/j.nepr.2012.10.011>
- Weeks, K.W., B.M. Hutton, D. Coben, J.M. Clochesy, and D. Pontin. 2013b. "Safety in Numbers 4: The Relationship between Exposure to Authentic and Didactic Environments and Nursing Students' Learning of Medication Dosage Calculation Problem Solving Knowledge and Skills." *Nurse Education in Practice* 13: e43–54. <https://doi.org/10.1016/j.nepr.2012.10.010>.
- Weeks, K. W., B.M. Hutton, S. Young, D. Coben, J. M. Clochesy, and D. Pontin. 2013c. "Safety in Numbers 2: Competency Modelling and Diagnostic Error Assessment in Medication Dosage Calculation Problem-solving." *Nurse*

- Education in Practice* 13: e23–32.  
<https://doi.org/10.1016/j.nepr.2012.10.013>.
- Wheeler, D. W., D. D. Remoundos, K. D. Whittlestone, T. P. House, and D. K. Menon. 2004. “Calculation of Doses of Drugs in Solution.” *Drug safety* 27(10):729–734. <https://doi.org/10.2165/00002018-200427100-00003>.
- Wheeler, D. W., S.J. Wheeler, and T.R. Ringrose. 2007. “Factors Influencing Doctors’ Ability to Calculate Drug Doses Correctly.” *International Journal of Clinical Practice* 61(2):189–194. <https://doi.org/10.1111/j.1742-1241.2006.01273.x>.
- WHO. See World Health Organisation.
- Windish, D. M., S. J. Huot, and M. L. Green. 2007. “Medicine Residents’ Understanding of the Biostatistics and Results in the Medical Literature.” *Journal American Medical Association* 298(9):1010–1022.  
<https://doi.org/10.1001/jama.298.9.1010>.
- Williams, D. J., and J.D. Walker. 2014. “A Nomogram for Calculating the Maximum Dose of Local Anaesthetic.” *Anaesthesia* 69(8): 847–853.  
<https://doi.org/10.1111/anae.12679>.
- World Health Organisation. 2017. “WHO Global Patient Safety Challenge: Medication Without Harm.” Available from:  
<https://www.who.int/patientsafety/medication-safety/medication-without-harm-brochure/en/>.
- Wright, K. 2010. “Do Calculation Errors by Nurses Cause Medication Errors in Clinical Practice? A Literature Review.” *Nurse Education Today* 30(1), 85–97. <https://doi.org/10.1016/j.nedt.2009.06.009>.
- Young, S., K. W. Weeks, and B.M. Hutton. 2013. “Safety in Numbers 1: Essential Numerical and Scientific Principles Underpinning Medication Dose Calculation.” *Nurse Education in Practice* 2013(13):e11–22.  
<https://doi.org/10.1016/j.nepr.2012.10.012>.