Research Report

Benchmark assessment of numeracy for nursing:
Medication dosage calculation at point of registration

Authors:

- Professor Diana Coben, BA, DipACE, PhD, FRSA, FHEA, Professor of Adult Numeracy, King’s College London
- Dr Carol Hall, RGN, RSCN, PhD, BSc (Hons) PG Dip Ad/Ed PG Dip ResM, FHEA, Associate Professor, School of Nursing, University of Nottingham
- Dr Meriel Hutton, RGN, RNT, PhD, BA, CertEd, DN, Senior Visiting Research Fellow, King’s College London
- Dr David Rowe, PhD, MSc, BEd (Hons), FACSM, FRC, Reader in the Department of Sport, Culture and the Arts, University of Strathclyde
- Dr Keith Weeks, RN, PhD, BSc (Hons), BEd, DipN, RNT, FHEA, Reader of Health Professional Education, Faculty of Health, Sport and Science, University of Glamorgan
- Norman Woolley, MSc, MN, RGN, DipN, PGCE, RNT, Head of Learning and Teaching; Associate Head Dept of Professional Education and Service Delivery, Faculty of Health, Sport and Science, University of Glamorgan
## Contents

Report to NHS Education for Scotland on work to develop a benchmark assessment in numeracy for nursing ................................................................. 4

  Background ........................................................................................................ 4

Benchmark assessment of numeracy for nursing: Medication dosage calculation at point of registration ................................................................. 10

  Aim ....................................................................................................................... 10
  Research questions .............................................................................................. 10
  Experimental protocol and methods .................................................................. 10
    Instrument Design ............................................................................................... 11

Design and development of the medication dosage calculation problem-solving competency model ................................................................. 12

Design and development of the computer-based authentic assessment data collection tool ................................................................. 15

Design and development of the Practical Simulation Assessment data collection tool ...................................................................................... 19

Methods and Methodology .................................................................................. 20

Stages of the research study .................................................................................. 21
  Pilot study (March - April, 2008) ........................................................................... 22
  Main Study – Stage 1 (April – September, 2008) ................................................. 22
  Main Study – Stage 2 (September 2008 - June 2009) ........................................... 23
  Main Study – Stage 3 (July – December 2009) .................................................... 23

Sample Frame for the Main Study ........................................................................ 23

  Sample for the Main Study .................................................................................. 23

Ethics and Access .................................................................................................. 24
  Ethics ................................................................................................................... 24
  Access .................................................................................................................. 26

Research protocol .................................................................................................. 27
  Data Collection ..................................................................................................... 27
  Practical Issues .................................................................................................... 27

The Assessment Process ......................................................................................... 28
  The Computer-based Assessment Process .......................................................... 29
  The Practice Based Assessment Process ............................................................. 30
  Calculators ........................................................................................................... 33
The Evaluation Instrument .................................................................................. 33

Data processing and analysis ........................................................................... 34
- Data processing .......................................................................................... 34
- Summary data analyses ............................................................................. 35

Results .............................................................................................................. 37
- Demographic information on participants .................................................. 37
- Summary analyses of subscale and total scale scores .................................. 37
- Item-level data analyses ............................................................................. 47
- Tabular, Histogram and Summary Results .................................................. 48
- Tablet and Capsule Section Summary ....................................................... 48
- Oral Liquid Medicine Section Summary ................................................... 55
- Injections Section Summary ..................................................................... 68
- Intravenous Infusion Section Summary ..................................................... 82

Discussion ........................................................................................................ 93
- Tablets and capsules ................................................................................... 94
- Oral liquid medicine .................................................................................. 95
- Injections .................................................................................................... 96
- Intravenous fluid delivery rates ................................................................... 97

Conclusions .................................................................................................... 100

Summary .......................................................................................................... 102

References ...................................................................................................... 103

Annexe 1: Report of the pilot study ................................................................. 106
Report to NHS Education for Scotland on work to develop a benchmark assessment in numeracy for nursing

Sponsor:
Michael Sabin, RGN BSc (Hons) MSc (Education) PG Cert. Learning Teaching and Assessment, Programme Director, NHS Education for Scotland, currently on secondment as Nursing Officer Workforce Development, Chief Nursing Officer’s Directorate, Scottish Government

Research team:
- Professor Diana Coben, BA, DipACE, PhD, FRSA, FHEA, Professor of Adult Numeracy, King’s College London
- Dr Carol Hall, RGN, RSCN, PhD, BSc (Hons) PG Dip Ad/Ed PG Dip ResM, FHEA, Associate Professor, School of Nursing, University of Nottingham
- Dr Meriel Hutton, RGN, RNT, PhD, BA, CertEd, DN, Senior Visiting Research Fellow, King’s College London
- Dr David Rowe, PhD, MSc, BEd (Hons), FACSM, FRC, Reader in the Department of Sport, Culture and the Arts, University of Strathclyde
- Dr Keith Weeks, RN, PhD, BSc (Hons), BEd, DipN, RNT, FHEA, Reader of Health Professional Education, Faculty of Health, Sport and Science, University of Glamorgan
- Norman Woolley, MSc, MN, RGN, DipN, PGCE, RNT, Head of Learning and Teaching; Associate Head Dept of Professional Education and Service Delivery, Faculty of Health, Sport and Science, University of Glamorgan

Background
The development of competence in numeracy by healthcare staff and students of healthcare programmes within higher education is a key area for concern, as shown by successive studies (Jukes, 2006). The NPSA, in partnership with the British Medical Journal (BMJ) Publishing Group, has provided a standardised online educational package for junior doctors (NPSA, 2006) but as yet nothing has been instigated for nurses and other healthcare professionals.

Medication calculations are an important part of the numeracy required for healthcare. Medication errors are an aspect of clinical governance highlighted recently by the National Patient Safety Agency (NPSA) in England and Wales and have been targeted for remedial action over recent years (NPSA, 2006a; NPSA,
The number of injuries and deaths attributable to medication error in the NHS is unknown but 9% of incidents reported to the NPSA in its annual audit involved medicines (NPSA, 2003); and this is consistent with historical data. While miscalculation is not separately identified as an aspect of medication error, the NPSA do identify that

\textit{Miscalculation, failure to titrate the dose to the patient's needs, miscommunication and failure on the part of all team members to check the dose before dispensing, preparing or administering a dose are the most common factors contributing to dosing errors.}

(NPSA, 2009, p.19)

Education and training is identified as contributory both positively and negatively. The Department of Health report on \textit{Improving Medication Safety} (Smith, 2004) put some of the blame for medication errors on inadequacies in the education and training of both doctors and nurses.

In Scotland, \textit{Safe Today, Safer Tomorrow} (NHS QIS, 2006) focused on patient safety and risk management, including medication error and recommended that NHS Quality Improvement Scotland (NHS QIS) should take responsibility for co-ordinating the development of a structured action plan, including costs and timescales, in order to take this work forward (NHS QIS, 2006, p.16). NHS QIS has a lead role in the promotion of patient safety and works in partnership with the National Patient Safety Agency (NPSA) and other patient safety organisations internationally. NHS QIS is leading and co-ordinating the Scottish Patient Safety Programme (SPSP) within which a key initial goal was to drive improvements in Medicines Management.

In the context of this concern about calculation competence in the healthcare professions, there have been calls both for benchmarks on requirements for entry to programmes of preparation and for some absolute measure of achievement and performance at the point where the student or trainee formally becomes a practising professional. Indeed, from September 2008 the body regulating the Nursing profession in the UK, the Nursing and Midwifery Council (NMC), has required students to achieve 100% in a test of numeracy in practice
(NMC, 2007a) before they are allowed to register as nurses. In addition to testing at the point of registration, the NMC requires universities to judge nursing students’ mathematics ability at Entry to Programme and, more recently, also at Entry to Branch (Adult, Child, Mental Health, Learning Disability) (NMC, 2007b).

Unfortunately, there are currently no national standards for teaching or assessment of numeracy during pre-registration nurse education and, in the absence of a robust evidence-based standard (a benchmark), a relativistic position has emerged with a multiplicity of tests, processes and criteria being developed and deployed locally, all ostensibly offering fixed points against which competence could be determined or predicted.

We argue that without such a benchmark, any measure of numerical competence is:

> ... in the eye of the recipient of evidence of that competence, be it higher education institutions, regulators, employers or service users.

(Hutton, 2004)

There are several linked issues here:

1. The a priori or foundation mathematical skills which a prospective nurse should have before commencing professional training;
2. Progressive competence development towards a fixed point at which competence must be shown;
3. The requirement for a benchmark;
4. The placement of the benchmark;
5. The nature of such a benchmark.

The first point proposes that without a particular set of (usually foundational) mathematical skills the prospective student will not reach the level of numerical competence which will be required to meet the collective (public and professional) expectations and requirements of the registered professional. Such markers of future capability are common within academic study and constitute the minimum grades sought for programme entry. Many Higher Education
Institutions (HEIs) set specific grades in mathematics within their requirements for entry to their pre-registration Nursing programmes on the assumption that a particular level of school mathematics qualification can act as a reliable proxy for future academic capability. However, the results of a study by Hutton suggest that qualifications in mathematics, other than those of above GCSE grade C or equivalent, should not be relied upon to predict performance in a test of calculations required in nursing (Hutton, 1998a). In any case, without a reliable benchmark for competence it is not possible to link entry criteria to subsequent competence development and this may result in either unnecessarily restrictive entry criteria or a failure to identify additional learning needs in our student body.

The second point recognises that, as a core outcome of the educational process, both general conceptual and operationally specific learning will occur and be measurable if appropriate learning, teaching and assessment processes are in place. The duration of educational programmes are generally set not just to allow us to fill students with all the knowledge and skills we believe they will need, but also to allow a step-by-step approach to the development of understanding and competence over time. Thus, an understanding of the key skills required and the students’ existing capabilities allows appropriate support and intervention across the duration of their programme of professional preparation.

The third point focuses on why such a benchmark might be needed. This gets to the heart of the rationale for professional, as opposed to purely academic, preparation. Professional competence is articulated not just in the quality of the individual's thinking, but in the way that such thinking is put to the service of those who are in receipt of the professional’s services. Notwithstanding an individual professional’s decision to undertake or withhold any action, if the professional act includes technical skills, then it is this technical competence which becomes the embodiment of the required and expected competence. In this way, the recent focus upon safety in healthcare has been the stimulus for both education to enhance competence and for measures which might demonstrate and govern high quality clinical performance. A benchmark
therefore seeks to provide both a target towards which competence development can be directed and a robust measure against which its achievement can be governed and assured.

If such a benchmark can be set, at what point should it be deployed? If professional competence is to be characterised in public protection and safety terms (and the NMC clearly sets its stall in this domain), then the logical point is that at which the student transforms into the professional. The point of registration is, by its nature, the point at which the individual professional accepts their accountability for their acts and omissions as measured against a set of (sometimes specific, sometimes rather general) standards. Up until that point, the individual's learning is transitional from the novice state to that which meets the agreed expectations.

Given the link between the required competence and its public consumption, the nature of any benchmark necessarily requires not just an assessment method which is reliable and valid in educational terms, but one which directly and authentically represents its purpose and the context in which it will be performed. Such a process does not undermine the transferability of learning but firmly cements the relationship between the desired expectation of competence and the governance of its development and subsequent performance.

Once established, a robust assessment benchmark provides not just some assurance of baseline professional standards but for the first time an opportunity to move away from relativistic interpretations of numerical competence and explore the relationship between entry qualifications, in-programme preparation, placement experience, remedial support and subsequent achievement. Thus, it would be possible to work backwards from this fixed criterion to look at key stage indicators and the probability of achieving the required standard. Such data would be extremely powerful in supporting subsequent educational intervention. Further, it would also allow us to look forward to subsequent stages in the individual’s career where either higher level skills require to be built on, or a refreshing of core skills can be facilitated by
recourse to the core standard.

Against this background, following on from a review of relevant literature (Sabin, 2001) and recommendations from a consultation on healthcare numeracy (NES Numeracy Working Group, 2006; Sabin, 2006a; Sabin, 2006b) and in line with a strategy developed by NES (Sabin, 2006a; Sabin, 2006b), NHS Education for Scotland (NES) brought together an interdisciplinary group of subject experts from across the UK to explore the key issues associated with determining the achievement of competence in numeracy for professional practice in nursing.

The overall aim of this work is to develop a proposed benchmark assessment for numeracy for nursing in Scotland. In the first instance we propose the establishment of a benchmark at entry to the NMC Register because this is the point at which students become registered nurses, accountable for their professional practice within the NMC’s regulatory framework.

Our work towards this aim is outlined in the remaining sections of this report.
Research Report

Benchmark assessment of numeracy for nursing: Medication dosage calculation at point of registration

Aim

The aim of the study was to evaluate empirical evidence of the reliability and convergent validity of a computer-based learning and assessment tool of medicine dosage calculations by comparing its outcomes with the outcomes of a practical activity requiring the same calculations and to determine learner-perceived acceptability of the assessment tools in relation to authenticity, relevance, fidelity and value. This was intended as a first step towards creating a benchmark for numeracy for nursing.

Research questions

1. What is the internal consistency reliability of the computer-based assessment and the practice based assessment?

2. What is the criterion-related validity of the computer-based assessment and the practice-based assessment?

3. How acceptable to nursing students are the assessments in terms of authenticity, relevance, fidelity and value?

Experimental protocol and methods

The research began with preliminary work undertaken between November 2006 and April 2007 (reported in Coben, 2008). This preliminary work comprised:

- adoption of a definition of numeracy applicable to nursing (Coben, 2000, p.35);
- development of evidence-based principles for numeracy for nursing;
- development of criteria for a benchmark assessment of numeracy for nursing;
- development of an evidence-based prototype benchmark assessment tool
covering practical knowledge in medicine dosage calculations

**Instrument Design**

Having reviewed the literature, we recognised that an authentic numeracy assessment tool should be:

- **Realistic**: Evidence-based literature in the field of nursing numeracy (Hutton, 1997; Weeks, 2001) strongly supports a realistic approach to the teaching and learning of calculation skills, which in turn deserve to be tested in an authentic environment. Questions should be derived from authentic settings.

- **Appropriate**: The assessment tool should determine competence in the key elements of the required competence (OECD, 2005; Sabin, 2001).

- **Differentiated**: There should be an element of differentiation between the requirements for each of the branches of nursing (Hutton, 1997).

- **Consistent with adult numeracy principles**: The assessment should be consistent with the principles of adult numeracy learning teaching and assessment, having an enablement focus (Coben, 2000).

- **Diagnostic**: The assessment tool should provide a diagnostic element, identifying which area of competence has been achieved, and which requires further intervention (Black & Wiliam, 1998).

- **Transparent**: The assessment should be able to demonstrate a clear relationship between ‘test’ achievement and performance in the practice context (Weeks, Lyne, Mosely, & Torrance, 2001).

- **Well-structured**: The assessment tool should provide a unique set of questions with a consistent level of difficulty and a structured range of complexity. (Hodgen & William, 2006)

- **Easy to administer**: the assessment should provide the opportunity for rapid collation of results, error determination, diagnosis and feedback (Black & William, 1998).

(Coben et al., 2008, pp 96-97)

In response to these requirements, the evidence-based benchmark computer assessment tool was designed by Weeks and Woolley to develop nurses'
medication dosage calculation skills using a framework derived from the Authentic World® computer programme. The assessment tool uses graphics and participant interaction to create close proximity to real world practice. It covers the full range of complexity within the hierarchy of dosage calculation problems which Adult Branch nurses are likely to meet at the point of registration. This includes unit dose, sub- and multiple-unit dose, complex problems and conversion of Systeme Internationale (SI) units (Weeks, Lyne, Mosely, & Torrance, 2001). The 28 practical simulation items were a subset of 50 computer simulation items selected on the basis of their high internal consistency reliability in an earlier study with students from Case Western Reserve University, USA, and the University of Glamorgan (Clochesy, 2008; Weeks & Woolley, 2007).

The practical simulation assessment was designed and developed to mirror the computer-based assessment.

Design and development of the medication dosage calculation problem-solving competency model

The generic numeracy definition adopted by the NES expert numeracy reference group was that defined by Coben as follows:

To be numerate means to be competent, confident, and comfortable with one’s judgements 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.

(Coben, 2000, p. 35, emphasis in the original)

For the purposes of extrapolating and operationalising a definition of dosage calculation competence from this definition of numeracy, the reference group adapted and crystallized a competency model derived from an initial premise described by Weeks, Lyne & Torrance (2000) and elaborated by Authentic World Ltd. (Figure 2, below).

Healthcare numeracy is a subset of a wider numeracy and medication dosage calculation is both a subset of numeracy and of a wider competency in medicines management (Figure 1).
Figure 1: Medication dosage calculation in the context of Numeracy and Medicines Management (Sabin, et al., 2008)

Medication dosage calculation was chosen because this is the most commonly cited exemplar for nursing numeracy. The public and professional focus on medication-related calculation is linked to patient safety since incorrect calculation of medication dosage may harm patients and, by default, the reputation of the profession and professional education.

While acknowledging the existence of wider medicines management factors that permeate the prescription, dispensing, preparation and administration of medicines, this competency model developed in the project focused specifically on medication dosage calculation problem-solving (Figure 2). However, once such a benchmark is established, it should be possible to create further benchmarks in numeracy for nursing at other key points in nurse training and to extend these to include other areas of numeracy for nursing, such as fluid balance, nutrition, management of nursing provision, etc.
Figure 2: Medication dosage calculation problem-solving competency model

This model articulates the inter-relationship between the three sub-elements that combine to form competence in medication dosage calculation problem-solving:

1. **Conceptual competence:** The need to identify, interpret, extract and understand the relationship between the medication nomenclature, dosage and numerical information embedded in medication prescriptions and medication container labeling/monographs; and to set up an accurate and appropriate dosage or rate equation that illustrates the relationship between the essential elements of the dosage problem.

2. **Calculation competence:** The need to undertake appropriate arithmetical operations and computations to calculate a numerical value that falls within an appropriate degree of accuracy for the required dose.
or rate.

3. **Technical measurement competence:** The need to select an appropriate measurement and delivery vehicle or vehicles and to transfer and measure the calculated numerical value for the correct dose or rate to be administered.

The confluence (central white interface) of the model highlights how all three sub-elements must be present in order to achieve a correct dosage or rate solution. Conversely, if an error is made within any one or more of the sub-elements then a medication dosage calculation error will result.

**Design and development of the computer-based authentic assessment data collection tool**

The computer based authentic assessment data collection tool designed for use in the NES research project was derived from an original constructivist-based program (Weeks 2001; Weeks *et al.*, 2001) and *Authentic World Ltd*. Figures 3 – 7 illustrate the original *Authentic World Ltd* design features of the constructivist based program and how this reflects iconic representations of the three sub-competency elements of the competency model:

![Figure 3: Tablet and Capsule Dosage Calculation Competence](image-url)

*Figure 3: Tablet and Capsule Dosage Calculation Competence*
Figure 4: Oral Liquid Medicine Dosage Calculation Competence

Figure 5: Injection Dosage Calculation Competence
The original constructivist-based Authentic World Ltd program reflected a diagnostic assessment of cognitive competence (European Commission for Education & Culture 2008) within the medication dosage calculation problem-solving competency domain. This was subsequently adapted for the purposes of designing an assessment and data collection tool for use in this project. The assessment and data collection tool focused specifically upon the presentation of medication prescriptions, medication containers and labels and the technical
measurement element of the competency model. The rationale for this adaptation was to:

1. Provide a framework within which a prospective competency benchmark in medication dosage calculation at the point of registration could be articulated and evaluated;
2. Provide an appropriate assessment platform that could be utilised by students who had been exposed to education programmes not based on constructivist pedagogies.

Figures 8 and 9 illustrate an injection problem from the adapted model developed as an exemplar assessment tool and used for data collection purposes in the project.

*Figure 8: Stage 1 of the assessment exemplar tool*
**Design and development of the Practical Simulation Assessment data collection tool**

Alongside the computer-based authentic assessment tool a practical simulation model was developed to mirror the key assessment elements. The equipment purchased to allow testing of these items in simulated practice in a maximum of two centres at any one time consisted of:

- 10 medicine cabinets containing:
  - appropriately labelled bottles for placebo tablets (6 per cabinet);
  - bottles for placebo liquid medicines (6 in each);
  - 6 x 10 mL ampoules of water for injection (appropriately labelled as placebo drugs);
- syringes in a range of sizes;
- needles for drawing up injections;
- plastic quills for drawing up liquid medicines;
- plastic medicine tots (10 ml and 30 ml capacity);
- placebo tablets;
- food colouring to mix with water for placebo liquid medicine;
- 10 mL plastic ampoules of water for injection for placebo drugs.
In both types of assessment, the medication dosage problem-solving questions consisted of:

- tablet and capsule, liquid medicine and injection problems presented in a hierarchy of unit dose, sub- and multiple unit dose, complex problems and conversion of SI unit problems;
- intravenous (IV) infusion problems focused on millilitre per hour calculations for administration of IV fluid solutions via volumetric pump delivery devices; and drop per minute calculations for administration of IV fluid solutions via a variety of IV administration set devices.

Questions were generated from a bank of 600 authentic medication dosage and IV infusion problems identified and extracted from clinical practice settings. The sample of questions used in the programme of research was selected on the basis of pre-assessed reliability and internal consistency as calculated by the statistical advisor to the project from data reported by Clochesy (2008): r = .98 and Weeks & Woolley (2007): r = .92.

**Methods and Methodology**

The methods and methodology utilized in the project may be summarized as follows in relation to the research questions:

1. What is the internal consistency reliability of the computer-based assessment and the practice based assessment?

   This was evaluated with Adult Branch nursing students in their 3rd year by administering a set of 40 computer-based assessment items and 28 matched practice-based assessment items. Cronbach’s alpha was then calculated for the total set of dosage calculation items, and on subsets of items (e.g., tablets and capsules, etc.), for both forms of test (computer-based and practice-based)
2. What is the criterion-related validity of the computer-based assessment and the practice-based assessment?

This was evaluated with Adult Branch nursing students in their 3\textsuperscript{rd} year by comparing their performance on the computer-based assessment tool with a practice-based assessment using a modified Objective Structured Clinical Examination (OSCE) format with the students undertaking the same calculations.

3. How acceptable to nursing students are the assessments in terms of authenticity, relevance, fidelity and value?

This was investigated by reviewing students’ online evaluation of the assessment tools they have experienced.

The project commenced investigation of these questions with a Pilot Study in a university in England; this is reported on in Annexe 1. In the main study, all universities in Scotland with pre-registration nurse training programmes were invited to participate. Assessment tools were refined in the light of experience on the pilot study and the practical equipment was moved to Scotland for the main study. Practical testing equipment was set up similarly in each site and students were assessed according to a generic practice schedule provided for the data collectors. Students had access to the practice schedule before participating in the assessment as part of the information given at prior presentations.

\textit{Stages of the research study}

The research process was structured as follows:

- Pilot study (March-April, 2008)
  - Stage 1: Materials testing
  - Stage 2: Piloting of the assessment of 3\textsuperscript{rd} year Nursing students in a university in England

- Main Study
  - Stage 1 (April – September 2008)
Stage 2 (September 2008 – June 2009)
Stage 3 (July – December 2009)

Pilot study (March - April, 2008)

The Pilot Study is summarised below (full details are in the Pilot Study report, attached below as Annexe 1 – see also Hutton et al., pub. pending).

The Pilot Study enabled a practical evaluation of both methods and materials with a typical sample group of final year student nurses within a large HEI similar to those planned for the main study. The university's research ethics committee gave ethical approval for the study. By undertaking the Pilot in England we ensured that Pilot data would not contaminate the main study.

Student performance in 28 medication calculation tasks was assessed using computer-based and practice-based formats. Both formats endeavoured to simulate clinical reality. The computer assessment was based on the Authentic World® computer programme which had already been extensively tested for face and content validity (Weeks, Lyne, Mosely & Torrance, 2001; Weeks, 2001) and research item internal reliability and the practice simulation exercise involved materials similar to those used in the local NHS Trust.

A computerised 50-item assessment (10 tablet & capsule; 10 liquid medicine; 10 injection; 10 two-part IVI = 50 items) was undertaken by all volunteering Adult Branch nursing students at commencement of their third year of study. Twenty eight of these were selected (a random stratified sample) to complete the final stage: a further computer assessment of the twenty eight most reliable set of items (drawn from the results of the baseline computer assessment) matched to the practice-based assessment.

Main Study – Stage 1 (April – September, 2008)

The main study took place in Scotland where six HEIs expressed interest in participating. These universities all included departments with students of
nursing in their third year of studies, and all agreed that students could be approached to volunteer to participate in the study, subject to ethical clearance. In the event, student volunteers from only four HEIs participated, due to logistical reasons. Stage 1 entailed adoption of the agreed ethical approval by interested Scottish universities, with extension where required, and recruitment and initial baseline assessment of students, from which a sample undertook Stage 2.

**Main Study – Stage 2 (September 2008 - June 2009)**

Stage 2 of the Main Study comprised cross-over assessment of the sample established in Stage 1 \((n = 63)\) in which half the students undertook the final computer assessment followed by the practical test and the other half undertook the assessments in reverse order.

**Main Study – Stage 3 (July – December 2009)**

Stage 3 comprised analysis and interpretation of results.

**Sample Frame for the Main Study**

The total population of new third year Adult Branch students at the volunteering HEIs in Scotland during the data collection period was estimated at 1000 across both Autumn and Spring entry cohorts.

**Sample for the Main Study**

- The numbers of students involved in the Main Study were as follows:
  - estimated number of eligible participants: 500;
  - number who undertook the initial online assessment and were invited to participate: 80;
  - number who participated: 64;
  - number who completed the assessment process (i.e., both online and practice): 63.
There may be several reasons for the small number of students who participated; this requires further investigation. The difference between those participating and those completing is explained by one student who felt unable to continue following the online assessment.

**Ethics and Access**

**Ethics**

The plan for attaining ethical approval was discussed extensively within the research team and also directly with research leads from the university nursing departments who were considering participation. A major difficulty to be resolved was that while the NHS has cross-boundary approval arrangements, this is not the situation between nor even within universities. Because the study was a multi-site project, ethical approval would theoretically be needed individually from every site used. We decided to base the project plan for ethical approval on work undertaken by Lauder et al (2007) as that required a similar multi-site arrangement. Initial contact was made to HEIs through introductory workshops which showcased the intended project within Scotland, and through telephone correspondence with directors of HEI pre-registration Nursing programmes. It was agreed that full ethical approval could be applied for through the English University. Once this had been granted, each individual school was contacted with details of the submission and the initial acceptance letter; the Scottish universities then determined what type of submission for approval was needed for their institution. This took time but worked well, with shortened submission arrangements successfully negotiated.

In designing the project, ethical concerns were specifically identified within three areas. The first related to the outcome of the assessments undertaken, if the tool subsequently demonstrated validity. There was a risk that student nurses who thought that they had performed badly might experience feelings of inadequacy or low self esteem. Secondly, there might be individuals who performed sufficiently badly for their potential professional practice to be
brought into question. A final issue related to the imposition of an assessment which might be perceived as increasing the learners’ existing workload or compromising their summative assessment schedule.

In order to resolve these issues the following mechanisms were incorporated into the project.

1. **The outcome of the assessment**: Students were advised that the assessment was deliberately timed to allow them to address any shortfalls in their knowledge identified by the assessment. Students were offered feedback and support after the assessment through access to the HEIs’ own facilities and, in negotiation with the Schools, through additional access to the computer learning package Authentic World® from which the online assessment was originally derived. Information given to participating HEIs included the possibility that some students would need support in relation to drug dosage calculation. This was reiterated at presentations to the lecturers in each School. They were also advised that all data would be confidential, i.e., known only to the researchers. Students were offered opportunities to discuss the project at any stage with the project team.

2. **Poor achievement in the assessment**: Assessment data from this project would not contribute to any summative assessment within each School. The students’ existing curricula had already been validated and reviewed conjointly by the NMC and the HEIs as an effective means for the preparation of student nurses. The project sought to illuminate methods and potential for future development which may, if deemed significantly valid, be adopted by each HEI as part of its curriculum. While trends in the data would be fed back to schools with the option of inclusion in the development of their curriculum, this decision would remain that of the HEI within the overall context of its governing drivers. In this context it would not be possible to identify final outcomes in relation to professional practice for individual students who have not completed
their programme of study, merely to identify the needs of the cohort in relation to further development prior to completion. This anonymised information was thus offered to the participating HEIs.

3. **Issues around workload:** Participation in the project would necessarily take up some student time in order to be completed effectively. However, we recognised that students would have to learn about medicine dosage calculation and practise this as part of their existing curriculum. It was expected that learning outcomes from participation in the project would match existing provision designed to meet NMC (2007) requirements. It was anticipated that in some circumstances students would have the opportunity to accredit learning from the study in place of existing preparation for the calculation and administration of medicines. Further, as the project ran at different times from students’ existing study sessions, they would not be disadvantaged.

**Access**

Once the work had received full ethical approval, we needed to access students for the following purposes:

- A 1 hour introductory session to outline the project and recruit volunteers. In this meeting students were given a presentation about the project and also invited to take a volunteers’ information form advising them about the project and their potential role within it. They were also invited to take a consent form to sign and return.
- A period of time for students to reflect on the original presentation before committing themselves to participation.
- A 2 hour online baseline assessment of all volunteers’ performance on the online assessment developed for the project.
- A 2 hour further computer assessment and a 2 hour practical assessment for a sample of students selected on the basis of their scores.
- A short online questionnaire to evaluate students’ perceptions of the usefulness of the two activities.
**Research protocol**

In each of the main study sites we assessed:

- All student volunteers using the computerised assessment programme.
- Up to 10 students (depending upon number volunteering) via practical assessment (Day 1) followed by a computerised assessment (Day 2).
- Up to 10 students (depending upon number volunteering) via the computerised assessment (Day 1) followed by the practical assessment (Day 2).

In some cases the model was slightly adapted, within the limits of ethical approval, in order to suit the needs of individual HEIs.

**Data Collection**

Data collection took place at each of the Main Study HEIs (14 days data collection in total). Students received feedback on their work after all data were collected and were offered access to the full Authentic World® computer programme for the remainder of their course of study.

**Practical Issues**

In line with ethical clearance, the member of the team working for NES was included in the data collection only where the results of the assessment could not be seen. This meant he could participate only in the introductory presentations and in the computer assessments (where data were submitted centrally online).

We negotiated access to the computer facilities and skills laboratories which would be needed at each site. We compiled a set of equipment necessary for the practical assessment and this was sent to each site by courier in time for the data collection activity. This meant that all materials and documentation were the same at each site. The one exception was the use of local infusion pumps (a) to
avoid transporting these and (b) to provide students with familiar equipment.

**The Assessment Process**

The assessment process comprised computerised assessment of numeracy items associated with administering medication and an assessment of identical items in a simulated practice setting. Students completed the assessments in a counterbalanced order (i.e., some completed the computer simulation first, then completed the practical simulation test, and other participants completed the tests in the opposite order). A break was given between tests.

The assessment items encompassed:

- Tablets and Capsules,
- Liquid Medicines
- Injections
- IV Infusions

Within each section, items ranged in difficulty, namely

- unit dose,
- sub and multiple unit dose,
- complex calculations
- SI unit conversion

IV infusions were subdivided into responses given in mL/hr and drops/minute (i.e., the same problem was answered in different units).

Participants responded to 50 items in the computer-based assessment and 28 items in the practical assessment; the latter were a subset of the 50 items tested via computer simulation. The computer-based assessment problem rubric illustrating medication domains, levels of complexity and numbers of questions within each is illustrated in Table 1.
<table>
<thead>
<tr>
<th>Part 1</th>
<th>Tablet &amp; Capsule</th>
<th>Liquid Medicine</th>
<th>Injection</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit Dose</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>Sub &amp; Multiple Unit Dose</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>Complex Arithmetic</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Conversion SI Units</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>10</strong></td>
<td><strong>10</strong></td>
<td><strong>10</strong></td>
<td><strong>30</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Part 2</th>
<th>IV Infusions</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>MI per Hour</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Drops per Minute</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>20</strong></td>
<td><strong>20</strong></td>
</tr>
</tbody>
</table>

Table 1: Problem rubric

The Computer-based Assessment Process

In our introductory presentation we demonstrated the computer tasks (please see the screen slide prints (figs. 8 and 9 as included within the previous section) of the computerised assessment for examples of questions). Students were then given the opportunity to volunteer and contact details of the research teams were made available to students in case they had any questions about the proposed assessment.

Prior to the assessment participating students were given access to personal computers in computer laboratories and password access to demonstration materials, together with support in learning how to use the computer package.

Students were asked to undertake the computer-based assessment alone. One hour was allowed for the initial computer assessment although some students finished more quickly.

During the online assessment the students were observed by the researchers who were also available to clarify the process. Responses were submitted by the
students to a central database where they were collated and marked automatically by the computer program.

The Practice Based Assessment Process

The practical assessment activities were also outlined in the introductory presentation. The medication dosage problem-solving tasks consisted of a total of 23 prescriptions requiring 28 calculations, which comprised:

- 6 each of: tablet and capsule; liquid medicine; and injection problems. These included unit dose, sub- and multiple-unit dose, complex calculations and conversion of SI units (18 calculations in total).
- 5 IV infusion prescriptions each requiring millilitre per hour calculations for administration via volumetric pump and drop per minute calculations for administration via a variety of IV administration drip sets (10 calculations in total).

Over two hours, students were assessed at two workstations, one for IV infusion calculations and one for liquid medicines, tablets and capsules, and injections. Each workstation had all the equipment required for completion of the assessment.

We designed the layout for the practical assessment of tablets and capsules, liquid medicines and injections so that we could see all the students but each student was unable to see what other students were doing (see Figure 10). In this position students could act as they would in practice and help from the researchers would be available quickly should they require it. This arrangement also allowed researchers to remove completed medication trays for marking without disturbing other students.
Medicine dosage calculations} Identical workstations with a medicine cabinet and a selection of medicine pots, syringes and needles at each were arranged on tables. The tables were arranged so that the open medicine cabinet acted as a screen between adjacent students. Paper and pencil were provided for any working out needed. Also on the table were three colour-coded dinner trays, each holding six small papier-maché trays and a prescription sheet containing six prescriptions (formatted similarly to those used in the NHS). One sheet was of prescriptions for tablets and capsules, one liquid medicines and one injections. For each, students had to select the correct preparation from the medicine cabinet and prepare it as for presentation to a patient.
When the student had prepared the prescriptions from one sheet, the completed tray was removed for checking, leaving the student with space to work through the next prescription sheet. We recorded students’ answers on a results sheet. One researcher acted as invigilator for the whole room while up to two others were required to collect and record students’ results.

**Intravenous fluid rates** For calculations involving IV fluid rates, a prescription sheet with five fluid prescriptions was provided to each student. Students were asked to calculate two rates per prescription – one in millilitres per hour for entry into the volumetric pump and the other for drops per minute. They then
set the rate by adjusting the manual roller on a giving set and inputting the correct numbers on a volumetric pump. Following pre-pilot materials testing, it was considered impractical for the researcher to count the drops each time and so students were asked to tell the researcher how many drops they were aiming for and this was recorded. This was done by the student noting their intended answer on a piece of paper and showing the examiner, thus avoiding the risk that other students might overhear the answers given. Many students lacked experience in setting rates on volumetric pumps and so again the researcher keyed in the numbers as dictated by the student, using the same approach as above. Students’ answers were recorded on their results sheet as they finished each task.

**Calculators**

Students were not permitted to use calculators. We do not object in principle to nurses using calculators (see Hutton, 1998b), but disallowing calculators enabled us to standardize the testing procedures across all students.

**The Evaluation Instrument**

After the completion of the Main Study students were emailed a questionnaire regarding their perception of their experience of the assessment process. The questionnaire design was informed by Gulikers’ 5 dimensional framework for authentic assessment (Gulikers, 2004). This framework was developed to inform and evaluate the design of assessment authenticity, with professional practice as the starting point. The five dimensions: task; physical context; social context; result/form; and criteria; subsequently provided a practice-focused framework for evaluating students’ perceptions of the authenticity and representativeness of the assessment environments. However, in the event, a disappointingly small response rate precluded this element from our final analysis.
**Data processing and analysis**

**Data processing**

Following completion of the two forms of assessment, the response of each participant on each item was marked as correct (entered as a 1 in the data) or incorrect (entered as 0 in the data) for the purposes of summary data analyses (see explanation of data analyses provided below). Where incorrect responses were given, field notes and observations were consulted to diagnose sources of error for item-level diagnostic analyses (see explanation of data analyses provided below).

Two sets of computer data were considered. The ‘full set’ included all items (50 items total; 10 items per subscale). The ‘reduced set’ included only those items that were common to the computer and practical simulation tests (28 items total; 6 items per subscale for tablets and capsules, liquid medicines, and Injections, 5 items per subscale for IV infusions mL/hr and IV infusions drops/min).

Two sets of practical simulation data were considered. In the first set, responses were marked based on the volume of liquid present, regardless of the position of the syringe plunger. In this scenario, technical error was penalised. For example, if 5mL of liquid was needed in the syringe, a response was marked incorrect if the participant aligned the head of the syringe plunger with the correct volume measurement bar on the syringe barrel (5 mL), but considerable air was introduced into the syringe barrel (resulting in an incorrect volume of liquid in the syringe). This data set is denoted by the abbreviation MA. In the second set, responses were marked based on the numerical answer provided on the instrument, regardless of whether the correct volume of liquid was present. In this scenario, technical error was not penalised. For example, if 5mL of liquid was needed in the syringe, a response was marked correct if the respondent aligned the head of the syringe plunger with the correct volume measurement bar on the syringe barrel (5 mL), even if considerable air was introduced into the syringe barrel. This data set is denoted by the abbreviation NUM.
This differentiated marking system affected responses to the Liquid Medicines and Injections items, in which syringes were used to provide the answer; it did not affect the tablet and capsule or IV infusion items. These two methods of marking the responses were followed to recognise the difference between the arithmetical component of the problem and the technical component of the problem (please see the Discussion section for a fuller explanation of the tie-in with the paradigm of numeracy problem-solving developed during the course of the NES study).

The responses to the two forms of the test were analysed in two fundamentally different ways, namely the ‘traditional’ or ‘summary’ data analyses, and the ‘item-level’ or ‘error-diagnosis level’. The traditional/summary data analyses answered the global questions regarding internal consistency reliability of the tests and subtests, and regarding the extent to which total score on the computer assessment matched total score on the equivalent items administered using the practical simulation (i.e., criterion-related validity). The item-level/error diagnosis analysis involved detailed inspection of the number and nature of errors at the item level, in order to inform how use of the two different test forms might identify problem-solving errors arising from different parts of the competency model as illustrated in Figure 2.

**Summary data analyses**

Summary data analyses were conducted on combined item scores, either on the complete version of the test (all items), or at the sub-test level (items on the tablets and capsules, liquid medicines, injections and IV infusions subtests).

Analysis Set 1 consisted of various descriptive statistics (minimum score, maximum score, mean, standard deviation, skewness and kurtosis). This gives a general idea of the range of responses, whether the data were relatively normally distributed (indicating whether the use of parametric data analysis is appropriate), whether some sub-tests were more difficult than others, and enables future comparison to performance by other samples.
Analysis Set 2 was the internal consistency reliability analyses. Internal consistency is the degree to which items within a test differentiate consistently between respondents with high or low levels of numerical competency. Cronbach’s alpha (intraclass correlation coefficient from a 2-way ANOVA model) was used, which is equivalent to the Kuder-Richardson (KR-20) coefficient traditionally used with dichotomous (correct/incorrect) knowledge test items.

Analysis Set 3 consisted of Pearson correlations between the various forms of the tests. The most important correlations were those between the computer simulation and practical simulation tests. This analysis is central to the main research question of whether computer simulation items are an accurate substitute for practical simulation questions (the issue of criterion-related validity evidence). Correlations tell us about relative validity (do the two tests put people in a similar order?)

Analysis set 4 included mean comparison of the two forms of assessment (computer simulation and practicals), using repeated measures t-tests and Cohen’s D. Similar to the above, this pertains to the criterion-related validity of the computer simulation form of the test, but from the perspective of absolute validity, namely whether the computer simulation format gives similar scores to the practical simulation format, or systematically higher/lower scores. The t-test is an inferential significance test that tells us whether any mean difference is due to chance or is more likely a systematic bias that exists in the population, when responding to the two forms of the test. Because repeated measures t-tests can be overpowered in larger samples, especially if there is a high correlation between the two sets of scores (as was expected in the current study), this can lead to what is called a Type I error (a conclusion that a mean difference is statistically significant, even though it may be of trivial clinical or practical importance). The Cohen’s D estimates the clinical meaningfulness or importance of a mean difference by standardising it against the within-group variability. Cohen provided the following guidelines for interpreting values of D: 0.2 = small effect; 0.5 = medium effect; > 0.8 = large effect (Cohen, 1968).
Results

Demographic information on participants

In the main phase of the study, 63 third year Adult Branch nursing students from 4 different institutions completed both forms of the dosage calculation test.

Summary analyses of subscale and total scale scores

Summary analysis Set 1: Descriptive statistics – total test scores and sub-tests. Descriptive statistics on the total scale and subscales are presented in Table 2. Mean scores were generally high (i.e., close to the maximum possible score), especially for the total test score and the Tablets and Capsules and Liquid Medicines subscales. It is also notable that the standard deviations were lower for these subscales. Together, these two statistical results (means and standard deviations) indicate that these sub-tests (Tablets and Capsules and Liquid Medicines) were easier for the participants, and that their scores were therefore less variable (i.e., most participants answered correctly on most items in the Tablets and Capsules and Liquid Medicines subscales).

The data were normally distributed (skewness and kurtosis < |2.0|), meaning that they met the distributional assumptions necessary for conducting the later parametric analyses (correlations and t-tests). This also has implications for the use of these items and similar items in future applied research studies that involve the use of parametric data analysis (e.g., t-tests, ANOVAs, Pearson correlation, etc.). Although slightly more non-normally distributed, most of the sub-tests also could be analyzed using parametric data analyses. In comparison to data previously collected on similar students in Wales (used to select items for the practical simulation in the current study), scores in this sample were generally higher (except for the IV subscales) and more homogeneous. This has relevance to later results reported in this section.
Table 2: Descriptive statistics for test performance on main sample (N = 63) Test

<table>
<thead>
<tr>
<th></th>
<th># items</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Computer Simulation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Full:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td>19</td>
<td>50</td>
<td>40.4</td>
<td>8.9</td>
<td>-0.8</td>
<td>-0.6</td>
</tr>
<tr>
<td>T/C</td>
<td>10</td>
<td>7</td>
<td>10</td>
<td>9.6</td>
<td>0.7</td>
<td>-1.7</td>
<td>2.1</td>
</tr>
<tr>
<td>L/M</td>
<td>10</td>
<td>5</td>
<td>10</td>
<td>9.1</td>
<td>1.2</td>
<td>-1.2</td>
<td>1.0</td>
</tr>
<tr>
<td>INJ</td>
<td>10</td>
<td>6</td>
<td>10</td>
<td>8.8</td>
<td>1.3</td>
<td>-0.8</td>
<td>-0.6</td>
</tr>
<tr>
<td>IV-ML</td>
<td>10</td>
<td>0</td>
<td>10</td>
<td>7.3</td>
<td>3.6</td>
<td>-1.2</td>
<td>-0.1</td>
</tr>
<tr>
<td>IV-DP</td>
<td>10</td>
<td>0</td>
<td>10</td>
<td>5.6</td>
<td>4.2</td>
<td>-0.4</td>
<td>-1.6</td>
</tr>
<tr>
<td><strong>Computer Simulation Reduced:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>28</td>
<td>10</td>
<td>28</td>
<td>22.7</td>
<td>4.8</td>
<td>-0.8</td>
<td>-0.3</td>
</tr>
<tr>
<td>T/C</td>
<td>6</td>
<td>4</td>
<td>6</td>
<td>5.7</td>
<td>0.6</td>
<td>-1.9</td>
<td>2.4</td>
</tr>
<tr>
<td>L/M</td>
<td>6</td>
<td>3</td>
<td>6</td>
<td>5.6</td>
<td>0.7</td>
<td>-1.9</td>
<td>3.5</td>
</tr>
<tr>
<td>INJ</td>
<td>6</td>
<td>3</td>
<td>6</td>
<td>5.1</td>
<td>1.0</td>
<td>-0.7</td>
<td>-0.6</td>
</tr>
<tr>
<td>IV-ML</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>3.6</td>
<td>1.9</td>
<td>-1.1</td>
<td>-0.5</td>
</tr>
<tr>
<td>IV-DP</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>2.7</td>
<td>2.1</td>
<td>-0.2</td>
<td>-1.7</td>
</tr>
<tr>
<td><strong>Practical Simulation MA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>28</td>
<td>9</td>
<td>28</td>
<td>23.1</td>
<td>4.3</td>
<td>-1.0</td>
<td>0.9</td>
</tr>
<tr>
<td>T/C</td>
<td>6</td>
<td>4</td>
<td>6</td>
<td>5.8</td>
<td>0.5</td>
<td>-3.0</td>
<td>7.6</td>
</tr>
<tr>
<td>L/M</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td>5.1</td>
<td>1.2</td>
<td>-1.4</td>
<td>1.4</td>
</tr>
<tr>
<td>INJ</td>
<td>6</td>
<td>0</td>
<td>6</td>
<td>4.6</td>
<td>1.6</td>
<td>-1.0</td>
<td>0.1</td>
</tr>
<tr>
<td>IV-ML</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>4.2</td>
<td>1.4</td>
<td>-2.0</td>
<td>3.2</td>
</tr>
<tr>
<td>IV-DP</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>2.9</td>
<td>2.0</td>
<td>-0.4</td>
<td>-1.4</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th></th>
<th>NUM</th>
<th>Total</th>
<th>28</th>
<th>9</th>
<th>28</th>
<th>22.6</th>
<th>4.4</th>
<th>-1.2</th>
<th>1.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>T/C</td>
<td></td>
<td></td>
<td>6</td>
<td>4</td>
<td>8</td>
<td>5.9</td>
<td>0.6</td>
<td>-1.4</td>
<td>7.0</td>
</tr>
<tr>
<td>L/M</td>
<td></td>
<td></td>
<td>6</td>
<td>1</td>
<td>6</td>
<td>5.3</td>
<td>1.1</td>
<td>-1.8</td>
<td>3.3</td>
</tr>
<tr>
<td>INJ</td>
<td></td>
<td></td>
<td>6</td>
<td>0</td>
<td>6</td>
<td>4.9</td>
<td>1.4</td>
<td>-1.5</td>
<td>2.3</td>
</tr>
<tr>
<td>IV-ML</td>
<td></td>
<td></td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>4.2</td>
<td>1.4</td>
<td>-2.0</td>
<td>3.2</td>
</tr>
<tr>
<td>IV-DP</td>
<td></td>
<td></td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>2.9</td>
<td>2.0</td>
<td>-0.4</td>
<td>-1.4</td>
</tr>
</tbody>
</table>

Notes: T/C = Tablets and capsules, L/M = Liquid medicines; INJ = injections; IV = Intravenous infusions; ML = mL/hr; DP = drops/min

Summary analysis set 2: Internal consistency reliability. Results for the reliability analysis are presented in Table 3. Although reliability was generally high for the total scale scores (for all forms of the test, i.e., Computer Simulation Full, Computer Simulation Reduced, Practical Simulation MA and Practical Simulation NUM), it was low for several of the subscale scores. This was due to two factors. First, several items were automatically excluded from the analyses due to zero variance (i.e., every student answered the item correctly). Calculation of correlation statistics in these circumstances is impossible because of the incorporation of item variance in the mathematical computations (and therefore the impossibility of dividing and multiplying by zero). Second, the variability (standard deviations) of these subscales was consequently reduced (and reliability tends to be lower in tests with a lower number of items). For subscales that had a complete item set in the analysis (and for the total scale scores), reliability was moderate to high, and similar to the reliabilities calculated on the data previously collected on similar students in Wales. Generally, the IV sub-tests were more reliable than the other sub-tests, and the sub-tests taken by the practical simulation method were slightly more reliable than the same sub-tests taken via computer simulation.

There are two reasons why the internal consistency was better in the Welsh data than in this study. First, there were 10 items per sub-test in the Welsh data, whereas this study had shortened sub-tests of 5-6 items. Second, and more
influential, was the effect of the “zero-variance” items in the main study. This did not happen in the Welsh data – every item had several people who got a wrong answer (so there were zeroes in every item, and there were more zeroes overall in the previous data set, creating greater variability and therefore higher correlations (internal consistency reliability estimates).

Table 3: Internal consistency reliability for test performance on main sample (N = 63)

<table>
<thead>
<tr>
<th>Test</th>
<th>Original # items</th>
<th>Items excluded*</th>
<th>Final # items</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer Simulation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full:</td>
<td>Total</td>
<td>50</td>
<td>1, 2, 3, 4, 11, 13, 15, 21, 22, 23</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>T/C</td>
<td>10</td>
<td>1, 2, 3, 4</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>L/M</td>
<td>10</td>
<td>11, 13, 15</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>INJ</td>
<td>10</td>
<td>21, 22, 23</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>IV-ML</td>
<td>10</td>
<td>21, 22, 23</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>IV-DP</td>
<td>10</td>
<td>21, 22, 23</td>
<td>10</td>
</tr>
<tr>
<td>Computer Simulation</td>
<td>Reduced:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>28</td>
<td>1, 3, 11, 22, 23</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>T/C</td>
<td>6</td>
<td>1, 3</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>L/M</td>
<td>6</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>INJ</td>
<td>6</td>
<td>22, 23</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>IV-ML</td>
<td>5</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>IV-DP</td>
<td>5</td>
<td>21, 22, 23</td>
<td>5</td>
</tr>
<tr>
<td>Practical Simulation</td>
<td>MA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Summary analysis set 3: Intercorrelation between tests. These results are presented in Table 4. Although these results are briefer than the others, they are among the most important results. The significant, and meaningful, correlations between the computer simulation format (full and reduced) and the practical simulation format (MA and NUM-scored versions) indicate that the computer simulation format tends to put testees in a similar order to the practical assessment format (i.e., higher scorers on the computer simulation also tend to be higher scorers on the practical simulation, lower scorers on the computer simulation also tend to be lower scorers on the practical simulation, middle scorers on the computer simulation also tend to be middle scorers on the practical simulation). This supports the proposition of using a computer simulation for testing dosage calculation skills in trainee nurses. Importantly, these intercorrelations (all between $r = .73$ and $.77$) were estimated from a reduced item set of 28 items (because of the limitation on the number of items that could be tested in the practical simulation test), whereas the proposed
computer simulation test would have a greater number of items (closer to 50), in which case the “true” correlation between the computer simulation and the practical simulation (if an equivalent item set of 50 items were used for both) would be higher. Using the Spearman-Brown prophecy formula, this would approximate to correlations of between $r = .83$ and $.86$.

Although not reported in this section, the intercorrelations between the practical simulation and computer simulation for separate subscales were lower (these are contained in the $t$-test results, but range from $r = .29$ to $.76$). This is to be expected because these consist of far fewer items.

*Table 4: Pearson correlations between different forms of dosage calculation tests on main sample (N = 63)*

<table>
<thead>
<tr>
<th></th>
<th>Computer Simulation (All)</th>
<th>Computer Simulation (Reduced)</th>
<th>Practical Simulation MA</th>
<th>Practical Simulation NUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer Simulation (All)</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer Simulation (Reduced)</td>
<td>.98</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practical Simulation MA</td>
<td>.73</td>
<td>.74</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Practical Simulation NUM</td>
<td>.76</td>
<td>.77</td>
<td>.97</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*Note: All correlations were significant (p < .01)*

Summary analysis set 4: t-tests comparing mean performance. Due to the fact that the full computer simulation test contained 50 items, it would have been inappropriate to compare mean scores of the full computer simulation test to the practical simulation test, which had a reduced set of items, therefore only computer simulation scores from the reduced version were compared to the practical simulation. This was done for the total test score and each subscale score. These results are presented in Tables 5a through Table 5f.
Table 5a: Mean comparison of Computer Simulation (MA and NUM scoring) vs. Practical Simulation, total score (N = 63)

<table>
<thead>
<tr>
<th>Computer Simulation vs. Practical Simulation MA</th>
<th>Practical Simulation MA Total</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer Simulation Total</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>r</td>
<td>t</td>
</tr>
<tr>
<td>Computer Simulation vs. Practical Simulation MA Total</td>
<td>22.7</td>
<td>4.8</td>
<td>22.6</td>
<td>4.4</td>
<td>.74</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Computer Simulation vs. Practical Simulation NUM</th>
<th>Practical Simulation NUM Total</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer Simulation Total</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>r</td>
<td>t</td>
</tr>
<tr>
<td>Computer Simulation vs. Practical Simulation NUM Total</td>
<td>22.7</td>
<td>4.8</td>
<td>23.1</td>
<td>4.3</td>
<td>.75</td>
<td>-1.2</td>
</tr>
</tbody>
</table>

Results statement: Mean total scores from the Computer Simulation test were not significantly different (p > .05) from mean total scores on the Practical Simulation test. Differences were clinically trivial (Cohen's D ≈ 0.02-0.09)

Table 5b: Mean comparison of Computer Simulation (MA and NUM scoring) vs. Practical Simulation tablets and Capsules sub-test score (N = 63)

<table>
<thead>
<tr>
<th>Computer Simulation vs. Practical Simulation MA</th>
<th>Practical Simulation MA T/C</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer Simulation T/C</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>r</td>
<td>t</td>
</tr>
<tr>
<td>Computer Simulation vs. Practical Simulation MA T/C</td>
<td>5.7</td>
<td>0.6</td>
<td>5.8</td>
<td>0.5</td>
<td>.76</td>
<td>-2.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Computer Simulation vs. Practical Simulation NUM</th>
<th>Practical Simulation NUM T/C</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer Simulation T/C</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>r</td>
<td>t</td>
</tr>
<tr>
<td>Computer Simulation vs. Practical Simulation NUM T/C</td>
<td>5.7</td>
<td>0.6</td>
<td>5.9</td>
<td>0.6</td>
<td>.70</td>
<td>-2.6</td>
</tr>
</tbody>
</table>

Results statement: Mean scores on the Tablets and Capsules questions were significantly (p < .05) higher on the Practical Simulation than on the Computer Simulation. However, this corresponded to a clinically trivial to moderately meaningful difference (Cohen’s D ≈ 0.2-0.4)
Table 5c: Mean comparison of Computer Simulation (MA and NUM scoring) vs. Practical Simulation, Liquid Medicines sub-test score (N = 63)

<table>
<thead>
<tr>
<th>Computer Simulation vs. Practical Simulation MA</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>L/M</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td><strong>SD</strong></td>
</tr>
<tr>
<td>5.6</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Computer Simulation vs. Practical Simulation NUM

<table>
<thead>
<tr>
<th><strong>L/M</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td><strong>SD</strong></td>
</tr>
<tr>
<td>5.6</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Results statement: Mean scores on the Liquid Medicines questions were significantly (p < .05) lower on the Practical Simulation than on the Computer Simulation. This corresponded to a clinically moderately meaningful difference (Cohen’s D ≈ 0.3-0.5).

Table 5d: Mean comparison of Computer Simulation (MA and NUM scoring) vs. Practical Simulation, Injections sub-test score (N = 63)

<table>
<thead>
<tr>
<th>Computer Simulation vs. Practical Simulation MA</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inj</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td><strong>SD</strong></td>
</tr>
<tr>
<td>5.1</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Computer Simulation vs. Practical Simulation NUM

<table>
<thead>
<tr>
<th><strong>Inj</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td><strong>SD</strong></td>
</tr>
<tr>
<td>5.1</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Results statement: Mean scores on the Injections questions were significantly (p < .05) lower on the Practical Simulation MA than on the Computer Simulation. This corresponded to a clinically moderately meaningful difference (Cohen’s D ≈ 0.4). Mean scores on the Injections questions were not significantly (p > .05) different between the Practical Simulation NUM and the Computer Simulation.
Table 5e: Mean comparison of Computer Simulation (MA and NUM scoring) vs. Practical Simulation, IV sub-test score, mL/hr (N = 63)

<table>
<thead>
<tr>
<th>Computer Simulation vs. Practical Simulation MA</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer Simulation</td>
<td>Practical Simulation</td>
<td>MA IV-ML</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV-ML</td>
<td>IV-ML</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>3.6</td>
<td>1.9</td>
<td>4.2</td>
<td>1.4</td>
<td>.57</td>
</tr>
</tbody>
</table>

Results statement: Mean scores on the IV (mL/hr) questions were significantly (p < .05) higher on the Practical Simulation than on the Computer Simulation. This corresponded to a clinically moderately meaningful difference (Cohen’s D ≈ 0.4)

Table 5f: Mean comparison of Computer Simulation (MA and NUM scoring) vs. Practical Simulation, IV sub-test score, drops/min (N = 63)

<table>
<thead>
<tr>
<th>Computer Simulation vs. Practical Simulation MA</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer Simulation</td>
<td>Practical Simulation</td>
<td>MA IV-DP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV-DP</td>
<td>IV-DP</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>2.7</td>
<td>2.1</td>
<td>2.9</td>
<td>2.0</td>
<td>.68</td>
</tr>
</tbody>
</table>

Results statement: Mean scores on the IV (drops/min) questions were not significantly different (p > .05) between the Computer Simulation and Practical Simulation versions. The mean difference was clinically trivial (Cohen’s D ≈ 0.1)
Similar to the previous results, these also are very important results, as they assess the *absolute* validity between the two forms of the test – namely, on average, do participants get similar scores from the computer simulation and practical simulation (or alternatively, does one test result in higher scores, indicating a systematic bias)?

The results here are again very promising for the total test score. Average score over all 28 items is essentially the same regardless of whether the questions are answered using the practical simulation or computer simulation format. They are statistically not significantly different and clinically/practically not meaningfully different (as indicated by trivial effect size). The same is not always true for the separate sub-tests. In some cases (e.g., Tablets and Capsules), the practical simulation format resulted in a higher mean score (though the difference was relatively minor). In other cases (e.g., Liquid Medicines), the computer simulation format resulted in a higher mean score than the practical simulation format. In other cases (e.g., IV drops/min), there was no difference between mean scores on the two test formats. The irregular pattern of results, in addition to the often relatively small differences (even though sometimes statistically significant), combined with the often homogeneous data (small standard deviations increase statistical power) lead to the conclusion that any differences between formats (computer simulation vs. practical simulation) on the sub-tests is coincidental, or a function of sampling variability (random differences).
Item-level data analyses

The following section provides:

a) Summary tables which compare congruence (criterion-related validity) between participant performance on assessment items within the computer simulation and the practical assessment. Operational definitions for congruence were manifestation of a CORRECT performance on an item in the computer assessment and on its counterpart item in the practical assessment AND manifestation of an INCORRECT performance on an item in the computer assessment and on its counterpart item in the practical assessment (denoted in green in the tables). Operational definitions for incongruence were manifestation of a CORRECT performance on an item in the computer assessment and an INCORRECT performance on its counterpart item in the practical assessment, and vice-versa (denoted in red in the tables). As described previously, within the liquid medicines and injections sections a further level of evaluation of performance was undertaken and variation denoted using the abbreviations MA and NUM. This is reflected in the inclusion of two tables for both these sections.

b) Item by item histograms, analysis and commentary on individual assessment items. Individual item analyses includes the complexity level of the dosage/rate problem, the correct solution to the problem, the congruence level between the computer and practice based assessments expressed as a histogram and a percentage value, and the range of incorrect solution values manifested by participants.

This data and item-by-item representation provides a detailed discrete and comparative analysis of within section participant performance and the congruence (criterion-related validity) between the two assessment environments.
### Tabular, Histogram and Summary Results

#### Tablet and Capsule Section Summary

*Table 6: Tablet and Capsule Summary Results*

<table>
<thead>
<tr>
<th>Question</th>
<th>Congruent Outcome</th>
<th>Incongruent Outcome</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% Computer &amp; practice performance correct</td>
<td>% Computer &amp; practice performance incorrect</td>
<td>% Computer performance correct/ Practice performance incorrect</td>
</tr>
<tr>
<td>1 (1)</td>
<td>63 (100%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 (3)</td>
<td>63 (100%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 (5)</td>
<td>62 (98.4%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 (6)</td>
<td>61 (96.8%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 (9)</td>
<td>53 (84.1%)</td>
<td>4 (6.4%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 (10)</td>
<td>48 (88.9%)</td>
<td>4 (7.4%)</td>
<td>2 (3.7%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sub-Total</td>
<td>94.7%</td>
<td>2.3%</td>
<td>0.6%</td>
</tr>
<tr>
<td>Congruence Totals</td>
<td>97%</td>
<td></td>
<td>3%</td>
</tr>
</tbody>
</table>

**Participants:** Problems 1 – 5 (n = 63); Problem 6 (n = 54)
Analysis and commentary on result

- Problem – unit dose: Prescribed 20mg / Dispensed 20mg: 1 Tablet
- Correct solution = 1 Tablet
- Congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 100%
- Computer environment errors: 0
- Quasi-practice environment errors: 0
Analysis and commentary on result

- Problem – unit dose: Prescribed 7.5mg / Dispensed 7.5mg: 1 Tablet
- Correct solution = 1 Tablet
- Congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 100%
- Computer environment errors: 0
- Quasi-practice environment errors: 0
Analysis and commentary on result

- Problem – multiple unit dose: Prescribed 4mg / Dispensed 1mg: 1 Tablet
- Correct solution = 4 Tablets
- Congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 98%
- Computer environment errors: 1 (1 tab)
- Quasi-practice environment errors: 0
Analysis and commentary on result

- Problem – multiple unit dose: Prescribed 100mg / Dispensed 50mg: 1 Tablet
- Correct solution = 2 Tablets
- Congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 97%
- Computer environment errors: 2 (0.5 tab)
- Quasi-practice environment errors: 0
Analysis and commentary on result

- Problem – SI Unit Conversion: Prescribed 1.5mg / Dispensed 500mcg: 1 Tablet
- Correct solution = 3 Tablets
- Congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 90%
- Computer environment errors: 10 (Did Not Attempt (DNA), 0.5 tab, 1.5 tab, 2 tabs, 2.5 tabs)
- Quasi-practice environment errors: 4 (DNA, 1 tab, 1.5 tabs)
Analysis and commentary on result

- Problem – SI Unit Conversion: Prescribed 0.6gram / Dispensed 200mg: 1 Tablet
- Correct solution = 3 Tablets
- Congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 96%
- Computer environment errors: 4 (DNA)
- Quasi-practice environment errors: 6 (DNA, 1 tab, 0.5 tab)
### Oral Liquid Medicine Section Summary

#### Table 7: Oral Liquid Medicines Summary Results

<table>
<thead>
<tr>
<th>Question</th>
<th>Congruent Outcome</th>
<th>Incongruent Outcome</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% Computer &amp; practice correct</td>
<td>% Computer &amp; practice incorrect</td>
<td>% Computer performance correct/ Practice performance incorrect</td>
</tr>
<tr>
<td>7 (11)</td>
<td>60 (95.2%)</td>
<td>0 (0%)</td>
<td>3 (4.8%)</td>
</tr>
<tr>
<td>8 (12)</td>
<td>53 (84.2%)</td>
<td>0 (0%)</td>
<td>5 (7.9%)</td>
</tr>
<tr>
<td>9 (14)</td>
<td>56 (88.9%)</td>
<td>0 (0%)</td>
<td>5 (7.9%)</td>
</tr>
<tr>
<td>10 (16)</td>
<td>54 (85.7%)</td>
<td>0 (0%)</td>
<td>8 (12.7%)</td>
</tr>
<tr>
<td>11 (19)</td>
<td>57 (90.4%)</td>
<td>2 (3.2%)</td>
<td>3 (4.8%)</td>
</tr>
<tr>
<td>12 (20)</td>
<td>42 (66.7%)</td>
<td>10 (15.9%)</td>
<td>9 (14.2%)</td>
</tr>
<tr>
<td>Sub-Total</td>
<td>85.2%</td>
<td>3.2%</td>
<td>8.7%</td>
</tr>
<tr>
<td>Congruence Totals</td>
<td>88.4%</td>
<td>11.6%</td>
<td>100%</td>
</tr>
</tbody>
</table>
Table 8: Oral Liquid Medicines Summary Results (Results adjusted for observed minor technical measurement error variation from the correct numerical value; or correct numerical value transferred to the measurement vehicle with inclusion of an air bolus of >10% volume displacement.)

<table>
<thead>
<tr>
<th>Question</th>
<th>Congruent Outcome</th>
<th>Incongruent Outcome</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% Computer &amp; practice performance correct</td>
<td>% Computer &amp; practice performance incorrect</td>
<td>% Computer performance correct/Practice performance incorrect</td>
</tr>
<tr>
<td>7 (11)</td>
<td>58 (92.1%)</td>
<td>0 (0%)</td>
<td>5 (7.9%)</td>
</tr>
<tr>
<td>8 (12)</td>
<td>53 (84.2%)</td>
<td>0 (0%)</td>
<td>5 (7.9%)</td>
</tr>
<tr>
<td>9 (14)</td>
<td>53 (84.2%)</td>
<td>0 (0%)</td>
<td>8 (12.7%)</td>
</tr>
<tr>
<td>10 (16)</td>
<td>52 (82.6%)</td>
<td>0 (0%)</td>
<td>10 (15.8%)</td>
</tr>
<tr>
<td>11 (19)</td>
<td>56 (88.8%)</td>
<td>2 (3.2%)</td>
<td>4 (6.4%)</td>
</tr>
<tr>
<td>12 (20)</td>
<td>41 (65.1%)</td>
<td>10 (15.9%)</td>
<td>10 (15.8%)</td>
</tr>
<tr>
<td>Sub-Total</td>
<td>82.8%</td>
<td>3.2%</td>
<td>11.1%</td>
</tr>
<tr>
<td>Congruence Totals</td>
<td>86.0%</td>
<td>14.0%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Participants: n = 63
Review of results

Analysis and commentary on result

- Problem – unit dose: Prescribed 5mg / Dispensed 5mg: 5mL
- Correct solution = 5mL
- Congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 95%
- Computer environment errors: 0
- Quasi-practice environment errors: 3 (6mL, 2mL, 1mL); two participants failed to displace air bubbles from oral syringes

Technical measurement error adjusted results (observed minor error
variation from the correct numerical value; or correct numerical value transferred to the measurement vehicle with inclusion of an air bolus of >10% volume displacement):

• Two quasi-practice oral syringe measurement errors observed: 4.8mL and 5.2mL;
• Corrected congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 92%. 
Analysis and commentary on result

- Problem – unit dose: Prescribed 2.5mg / Dispensed 2.5mg: 5mL
- Correct solution = 5mL
- Congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 84%
- Computer environment errors: 5 (2.5mL, 4mL)
- Quasi-practice environment errors: 5 (1mL, 2.5mL, 4mL); two participants failed to displace air bubbles from oral syringes
Analysis and commentary on result

- Problem – multiple unit dose: Prescribed 160mg / Dispensed 40mg: 5mL
- Correct solution = 20mL
- Congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 89%
- Computer environment errors: 2 (15mL)
- Quasi-practice environment errors: 5 (6mL, 15mL, 25mL); one participant failed to displace air bubbles from an oral syringe

Technical measurement error adjusted results *(observed minor error variation from the correct numerical value; or correct numerical value transferred to the measurement vehicle with inclusion of an air bolus of >10% volume)*
displacement):

- Three quasi-practice measurement pot measurement errors observed: 22mL, 22mL and 22mL;
- Corrected congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 84%
Analysis and commentary on result

- Problem – multiple unit dose: Prescribed 300mg / Dispensed 50mg: 5mL
- Correct solution = 30mL
- Congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 86%
- Computer environment errors: 1 (22mL)
- Quasi-practice environment errors: 8 (15mL, 22mL, 27mL, 33mL, 34mL, 35mL)

Technical measurement error adjusted results (observed minor error variation from the correct numerical value; or correct numerical value transferred to the measurement vehicle with inclusion of an air bolus of >10% volume)
displacement): 

- Two quasi-practice measurement pot measurement errors observed: 31mL and 32mL;
- Corrected congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 83%.
Analysis and commentary on result

- Problem – SI Unit Conversion: Prescribed 0.5gram / Dispensed 5mg: 5mL
- Correct solution = 5mL
- Congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 94%
- Computer environment errors: 3 (DNA, 0.5mL, 1mL)
- Quasi-practice environment errors: 5 (DNA, 3mL, 10mL, 12mL, 18mL); four participants failed to displace air bubbles from oral syringes

Technical measurement error adjusted results (observed minor error variation from the correct numerical value; or correct numerical value transferred to the measurement vehicle with inclusion of an air bolus of >10% volume)
displacement):

- One quasi-practice oral syringe measurement error observed: syringe plunger drawn to 5mL volume graduation point with inclusion of 2.5mL air bolus;
- Corrected congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 92%.
Analysis and commentary on result

- Problem – SI Unit Conversion: Prescribed 1 gram / Dispensed 200mg: 5mL
- Correct solution = 25mL
- Congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 83%
- Computer environment errors: 12 (DNA, 2.5mL, 9.5mL, 20mL, 22mL)
- Quasi-practice environment errors: 19 (2.5mL, 3mL, 5mL, 15mL, 20mL, 28mL, 29mL, 30mL, 60mL)

Technical measurement error adjusted results (observed minor error variation from the correct numerical value; or correct numerical value transferred to the measurement vehicle with inclusion of an air bolus of >10% volume displacement):
• One quasi-practice measurement pot measurement error observed: 23mL;
• Corrected congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 81%.
## Injections Section Summary

### Table 9: Injections Summary Results

<table>
<thead>
<tr>
<th>Question</th>
<th>Congruent Outcome</th>
<th>Incongruent Outcome</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% Computer &amp; practice performance correct</td>
<td>% Computer &amp; practice performance incorrect</td>
<td>% Computer performance correct/Practice performance incorrect</td>
</tr>
<tr>
<td>13 (22)</td>
<td>58 (92.1%)</td>
<td>0 (0%)</td>
<td>5 (7.9%)</td>
</tr>
<tr>
<td>14 (23)</td>
<td>62 (98.4%)</td>
<td>0 (0%)</td>
<td>1 (1.6%)</td>
</tr>
<tr>
<td>15 (24)</td>
<td>53 (84.1%)</td>
<td>2 (3.2%)</td>
<td>6 (9.5%)</td>
</tr>
<tr>
<td>16 (27)</td>
<td>27 (42.9%)</td>
<td>18 (28.5%)</td>
<td>8 (12.7%)</td>
</tr>
<tr>
<td>17 (28)</td>
<td>38 (60.3%)</td>
<td>8 (12.7%)</td>
<td>10 (15.8%)</td>
</tr>
<tr>
<td>18 (29)</td>
<td>50 (79.3%)</td>
<td>4 (6.4%)</td>
<td>5 (7.9%)</td>
</tr>
<tr>
<td>Sub-Total</td>
<td>76.2%</td>
<td>8.5%</td>
<td>9.2%</td>
</tr>
<tr>
<td>Congruence Totals</td>
<td>84.7%</td>
<td></td>
<td>15.3%</td>
</tr>
</tbody>
</table>
Table 10: Injection Summary Results (Results adjusted for observed minor technical measurement error variation from the correct numerical value; or correct numerical value transferred to the measurement vehicle with inclusion of an air bolus of >10% volume displacement.)

<table>
<thead>
<tr>
<th>Question</th>
<th>Congruent Outcome</th>
<th>Incongruent Outcome</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% Computer &amp; practice performance correct</td>
<td>% Computer &amp; practice performance incorrect</td>
<td>% Computer performance correct/ Practice performance incorrect</td>
</tr>
<tr>
<td>13 (22)</td>
<td>54 (85.7%)</td>
<td>0 (0%)</td>
<td>9 (14.3%)</td>
</tr>
<tr>
<td>14 (23)</td>
<td>58 (92.1%)</td>
<td>0 (0%)</td>
<td>5 (7.9%)</td>
</tr>
<tr>
<td>15 (24)</td>
<td>49 (77.8%)</td>
<td>2 (3.2%)</td>
<td>10 (15.8%)</td>
</tr>
<tr>
<td>16 (27)</td>
<td>25 (39.7%)</td>
<td>18 (28.5%)</td>
<td>10 (15.8%)</td>
</tr>
<tr>
<td>17 (28)</td>
<td>35 (55.5%)</td>
<td>8 (12.7%)</td>
<td>13 (20.6%)</td>
</tr>
<tr>
<td>18 (29)</td>
<td>47 (74.6%)</td>
<td>4 (6.4%)</td>
<td>8 (12.6%)</td>
</tr>
<tr>
<td>Sub-Total</td>
<td>70.9%</td>
<td>8.5%</td>
<td>14.5%</td>
</tr>
<tr>
<td>Congruence Totals</td>
<td>79.4%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incongruent Totals</td>
<td></td>
<td>20.6%</td>
<td></td>
</tr>
</tbody>
</table>

Participants: n = 63
Review of results

Analysis and commentary on result

- Problem – unit dose: Prescribed 5mg / Dispensed 5mg: 1mL
- Correct solution = 1mL
- Congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 92%
- Computer environment errors: 0
- Quasi-practice environment errors: 5 (0.6mL, 0.8mL, 1.2mL, 5mL, 10mL); seven participants failed to displace air bubbles from injection syringes
**Technical measurement error adjusted results** *(observed minor error variation from the correct numerical value; or correct numerical value transferred to the measurement vehicle with inclusion of an air bolus of >10% volume displacement):*

- Four quasi-practice syringe measurement errors observed: 3 x 0.9mL (measured in a 5mL syringe); syringe plunger drawn to 1mL volume graduation point with inclusion of 0.98mL air bolus;

- Corrected congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 86%.
Analysis and commentary on result

- Problem – unit dose: Prescribed 60mg / Dispensed 60mg: 1mL
- Correct solution = 1mL
- Congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 98%
- Computer environment errors: 0
- Quasi-practice environment errors: 1 (10mL); seven participants failed to displace air bubbles from injection syringes

Technical measurement error adjusted results (observed minor error variation from the correct numerical value; or correct numerical value transferred to the measurement vehicle with inclusion of an air bolus of >10% volume displacement):
• Four quasi-practice syringe measurement errors observed: 0.92mL, 0.94mL; syringe plunger drawn to 1mL volume graduation point with inclusion of 0.1mL air bolus and 1mL air bolus;
• Corrected congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 92%.
Analysis and commentary on result

- Problem – multiple unit dose: Prescribed 75mg / Dispensed 50mg: 1mL
- Correct Solution = 1.5mL
- Congruence & equivalence of student medication dosage problem solving performance between computer & quasi-practice environments: 87%
- Computer environment errors: 4 (0.75mL, 1.25mL)
- Quasi-Practice environment errors: 8 (DNA, 0.75mL, 1mL, 1.3mL, 1.4mL, 1.7mL); eleven participants failed to displace air bubbles from injection syringes
Technical measurement error adjusted results (observed minor error variation from the correct numerical value; or correct numerical value transferred to the measurement vehicle with inclusion of an air bolus of >10% volume displacement):

- Four quasi-practice syringe measurement errors observed: 3 x 1.4mL; syringe plunger drawn to 1mL volume graduation point with inclusion of 1mL air bolus;
- Corrected congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 81%.
Analysis and commentary on result

- Problem – sub-unit dose: Prescribed 1mg / Dispensed 2.5mg: 1mL
- Correct solution = 0.4mL
- Congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 71%
- Computer environment errors: 28 (DNA, 0.1mL, 0.15mL, 0.2mL, 0.25mL, 0.3mL, 0.45mL, 2.5mL, 4mL, 5mL)
- Quasi-practice environment errors: 26 (DNA, 0.25mL, 0.3mL, 0.5mL, 0.6mL, 0.7mL, 0.8mL, 1mL, 1.1mL, 1.25mL, 1.5mL, 2.5mL, 3.8mL); seven participants failed to displace air bubbles from injection syringes

Technical measurement error adjusted results (observed minor error variation from the correct numerical value; or correct numerical value transferred to the measurement vehicle with inclusion of an air bolus of >10% volume displacement):
• Two quasi-practice syringe measurement errors observed: syringe plunger drawn to 0.4mL volume graduation point with inclusion of 0.08mL air bolus and 0.4mL air bolus;

• Corrected congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 68%.
Analysis and commentary on result

- Problem – sub-unit dose: Prescribed 2mg / Dispensed 5mg: 1mL
- Correct solution = 0.4mL
- Congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 73%
- Computer environment errors: 15 (DNA, 0.075mL, 0.2mL, 0.25mL, 0.3mL, 0.45mL, 1mL, 2mL, 2.5mL, 4mL)
- Quasi-practice environment errors: 18 (DNA, 0.1mL, 0.2mL, 0.25mL, 0.3mL, 0.33mL, 0.5mL, 0.6mL, 0.75mL, 1.9mL, 2mL, 2.4mL, 2.5mL); six participants failed to displace air bubbles from injection syringes

**Technical measurement error adjusted results** *(observed minor error variation from the correct numerical value; or correct numerical value transferred to the measurement vehicle with inclusion of an air bolus of >10% volume displacement):*
• Three quasi-practice syringe measurement errors observed: syringe plunger drawn to 0.4mL volume graduation point with inclusion of 0.07mL air bolus, 0.2mL air bolus and 0.4mL air bolus

• Corrected congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 68%
Analysis and commentary on results

- Problem– SI Units Conversion: Prescribed 0.25gram / Dispensed 500mg: 10mL
- Correct solution = 5mL
- Congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 86%
- Computer environment errors: 8 (DNA, 0.5mL, 2.5mL, 3mL, 10mL, 20mL)
- Quasi-practice environment errors: 9 (DNA, 0.15mL, 2.4mL, 2.5mL, 4.2mL, 4.5mL, 8); six participants failed to displace air bubbles from injection syringes

Technical measurement error adjusted results (observed minor error variation from the correct numerical value; or correct numerical value transferred to the measurement vehicle with inclusion of an air bolus of >10% volume)
displacement):

- Three quasi-practice syringe measurement errors observed: 4.6mL; syringe plunger drawn to 5mL volume graduation point with inclusion of 0.5mL air bolus and 5mL air bolus;
- Corrected congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 81%.
Intravenous Infusion Section Summary

Table 11: Intravenous Infusions Summary Results

<table>
<thead>
<tr>
<th>Question</th>
<th>Congruent Outcome</th>
<th>Incongruent Outcome</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% Computer &amp; practice performance correct</td>
<td>% Computer &amp; practice performance incorrect</td>
<td>% Computer performance correct/ Practice performance incorrect</td>
</tr>
<tr>
<td>19 (31) (mL/Hour)</td>
<td>40 (63.5%)</td>
<td>10 (15.8%)</td>
<td>2 (3.2%)</td>
</tr>
<tr>
<td>20 (31) (DPM)</td>
<td>27 (42.8%)</td>
<td>21 (33.3%)</td>
<td>8 (12.7%)</td>
</tr>
<tr>
<td>21 (32) (mL/Hour)</td>
<td>42 (66.6%)</td>
<td>7 (11.2%)</td>
<td>2 (3.2%)</td>
</tr>
<tr>
<td>22 (32) (DPM)</td>
<td>28 (44.5%)</td>
<td>21 (33.3%)</td>
<td>4 (6.4%)</td>
</tr>
<tr>
<td>23 (33) (mL/Hour)</td>
<td>49 (77.8%)</td>
<td>6 (9.5%)</td>
<td>3 (4.8%)</td>
</tr>
<tr>
<td>24 (33) (DPM)</td>
<td>25 (39.7%)</td>
<td>18 (28.5%)</td>
<td>11 (17.5%)</td>
</tr>
<tr>
<td>25 (34) (mL/Hour)</td>
<td>41 (65.1%)</td>
<td>8 (12.7%)</td>
<td>4 (6.4%)</td>
</tr>
<tr>
<td>26 (34) (DPM)</td>
<td>28 (44.5%)</td>
<td>18 (28.5%)</td>
<td>6 (9.5%)</td>
</tr>
<tr>
<td>27 (40) (mL/Hour)</td>
<td>40 (63.5%)</td>
<td>7 (11.2%)</td>
<td>2 (3.2%)</td>
</tr>
<tr>
<td>28 (40) (DPM)</td>
<td>25 (39.7%)</td>
<td>22 (34.9%)</td>
<td>5 (7.9%)</td>
</tr>
<tr>
<td>Sub-Total</td>
<td>54.8%</td>
<td>21.9%</td>
<td>7.5%</td>
</tr>
<tr>
<td>Congruence Totals</td>
<td>76.7%</td>
<td>23.3%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Participants: n = 63
Analysis and commentary on result

• Problem: Prescribed volume: 500mL / Infusion duration: 6 hours
• Correct solution = 83.33 (83) mL per hour
• Congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 79%
• Computer environment errors: 21 (DNA, 0mL, 12mL, 20mL, 82mL, 85mL, 87mL, 97.4mL, 123mL, 125mL, 160mL)
• Quasi-practice environment errors: 12 (DNA, 16mL, 17mL, 75mL, 76mL, 87mL, 97.4mL, 101.4mL, 123mL, 500mL)
Analysis and commentary on result

• Problem: IV administration set delivers 20 drops per millilitre
• Correct solution = 27.77 (28) drops per minute (DPM)
• Congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 76%
• Computer environment errors: 28 (DNA, 0 DPM, 3 DPM, 4 DPM, 5 DPM, 10 DPM, 14 DPM, 16 DPM, 20 DPM, 22 DPM, 24 DPM, 25 DPM, 29 DPM, 30 DPM, 32 DPM, 33 DPM, 36 DPM, 38 DPM, 60 DPM)
• Quasi-practice environment errors: 29 (DNA, 4 DPM, 5 DPM, 6 DPM, 9 DPM, 10 DPM, 12 DPM, 14 DPM, 15 DPM, 17 DPM, 20.18 DPM, 21 DPM, 22 DPM, 25 DPM, 29 DPM, 30 DPM, 31 DPM, 34 DPM, 38 DPM, 50 DPM, 51 DPM)
Analysis and commentary on result

• Problem: Prescribed volume: 1000mL / Infusion duration: 6 hours
• Correct solution = 166.66 (167) mL per hour
• Congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 78%
• Computer environment errors: 19 (DNA, 8mL, 12mL, 15.5mL, 60mL, 83.2mL, 140mL, 150mL, 168mL, 177.7mL, 183.3mL, 1000mL)
• Quasi-practice environment errors: 9 (DNA, 15mL, 165mL, 177.4mL, 267mL, 1000mL)
Analysis and commentary on result

- Problem: IV administration set delivers 12 drops per millilitre
- Correct solution = 33.33 (33) drops per minute (DPM)
- Congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 78%
- Computer environment errors: 31 (DNA, 0 DPM, 5 DPM, 8 DPM, 9 DPM, 10 DPM, 12 DPM, 16 DPM, 19 DPM, 28 DPM, 29 DPM, 31 DPM, 36 DPM, 48 DPM, 53 DPM, 60 DPM, 72 DPM, 83 DPM)
- Quasi-practice environment errors: 25 (DNA, 13 DPM, 15 DPM, 16 DPM, 20 DPM, 24 DPM, 26 DPM, 31 DPM, 32 DPM, 35.4 DPM, 36 DPM, 42 DPM, 43 DPM, 44 DPM, 55 DPM)
Analysis and commentary on result

- Problem: Prescribed volume: 500mL / Infusion duration: 4 hours
- Correct solution = 125 mL per hour
- Congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 87%
- Computer environment errors: 11 (DNA, 12mL, 48mL, 25mL, 120mL)
- Quasi-practice environment errors: 9 (DNA, 83mL, 120mL, 122mL, 185mL, 500mL)
Analysis and commentary on result

- Problem: IV administration set delivers 12 drops per millilitre
- Correct solution = 25 drops per minute (DPM)
- Congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 68%
- Computer environment errors: 27 (DNA, 0 DPM, 2 DPM, 5 DPM, 7 DPM, 9 DPM, 10 DPM, 12 DPM, 15 DPM, 21 DPM, 24 DPM, 26 DPM, 30 DPM, 55 DPM, 62 DPM)
- Quasi-practice environment errors: 29 (DNA, 6 DPM, 10 DPM, 12 DPM, 15 DPM, 17 DPM, 20 DPM, 20.12 DPM, 24 DPM, 29 DPM, 31 DPM, 36 DPM, 37 DPM, 38 DPM, 50 DPM)
Analysis and commentary on result

- Problem: Prescribed volume: 250mL / Infusion duration: 8 hours
- Correct solution = 31.25 (31) mL per hour
- Congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 78%
- Computer environment errors: 18 (DNA, 0mL, 30mL, 32.2mL, 36mL, 41mL, 60mL, 81.2mL, 125mL, 360mL)
- Quasi-practice environment errors: 12 (DNA, 33mL, 41mL, 41.66mL, 42mL, 62.5mL, 63mL, 250mL)
Analysis and commentary on result

• Problem: IV administration set delivers 60 drops per millilitre
• Correct solution = 31.25 (31) drops per minute (DPM)
• Congruence and equivalence of student medication dosage problem solving performance between computer and practice quasi-environments: 73%
• Computer environment errors: 29 (DNA, 0 DPM, 1 DPM, 12 DPM, 19 DPM, 25 DPM, 27 DPM, 30 DPM, 34 DPM, 36 DPM, 41 DPM, 48 DPM, 60 DPM, 69 DPM, 81 DPM)
• Quasi-practice environment errors: 24 (DNA, 1.7 DPM, 3 DPM, 4 DPM, 6 DPM, 17 DPM, 19 DPM, 20 DPM, 23 DPM, 25 DPM, 60 DPM, 63 DPM)
Analysis and commentary on result

• Problem: Prescribed volume: 1000mL / Infusion duration: 6 hours

• Correct solution = 166.66 (167) mL per hour

• Congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 75%

• Computer environment errors: 21 (DNA, 6mL, 20mL, 125mL, 140mL, 150mL, 155.5mL, 158mL, 168mL, 177.7mL, 183.3mL)

• Quasi-practice environment errors: 9 (DNA, 83mL, 165mL, 177.4mL, 1000mL)
Analysis and commentary on result

• Problem: IV administration set delivers 20 drops per millilitre
• Correct solution = 55.55 (56) drops per minute (DPM)
• Congruence and equivalence of student medication dosage problem solving performance between computer and quasi-practice environments: 75%
• Computer environment errors: 33 (DNA, 0 DPM, 2 DPM, 3 DPM, 8 DPM, 10 DPM, 18 DPM, 20 DPM, 27 DPM, 28 DPM, 31 DPM, 33 DPM, 36 DPM, 50 DPM, 52 DPM, 53 DPM, 54 DPM, 57 DPM, 65 DPM)
• Quasi-practice environment errors: 27 (DNA, 3 DPM, 8 DPM, 17 DPM, 18 DPM, 24 DPM, 33 DPM, 44 DPM, 50 DPM, 54 DPM, 59 DPM, 60 DPM, 66 DPM)
Discrete analysis of error classifications identified within the authentic computer environment and the practice environment

Discussion

In seeking to propose an evidence-based benchmark assessment for medication-related calculation in nursing, this project set out to evaluate the empirical evidence regarding the internal consistency reliability of both a computer-based assessment and a practice simulation assessment of medicine calculation competence. Further, it sought to determine the criterion-related validity of the computer-based assessment and the practice-based assessment to determine whether the computer-based assessment might act as a ‘proxy’ for the practice-based assessment. Finally, the study sought to determine whether, alone or in conjunction, these assessment processes could meet;

- the requirements of the professional regulator, employers and public in providing a reliable and valid measure of medicine-related calculation competence at point of registration
- the needs of HEIs in providing an educationally robust, fair and easily administered assessment
- the needs of students and practitioners in supporting accurate assessment of competence in the context within which it will be required.

A key focus of this study was to determine the validity of the computer simulation format of medication dosage calculations by testing it against the gold standard of a practical simulation format. The underlying rationale was that while a practical simulation format is a way of testing calculation skills across the full range of prescription types and complexity which would not be guaranteed in a practice setting, it is not feasible for mass testing. The computer simulation enables assessment of large numbers of people at the same time, in remote locations and with limited costs. If computer simulated assessment of medications calculations could be shown to operate similarly (provide similar results) to practical simulation testing, this would validate the use of computer
simulated testing as a relatively inexpensive, less time-consuming assessment method.

From the results presented above, the criterion-related validity of the computer simulation has been supported, both in terms of putting participants in a similar order of competence and in terms of participants obtaining similar absolute results (getting the same number of questions correct on the computer as they would in the practical situation). Different sections of the assessment are discussed below.

**Tablets and capsules**

The analysis of results in this section indicated that there were no significant differences between performance in the authentic computer environment and performance in the practice environment.

Very few errors were made overall and none at all in the unit dose calculations. The two errors made in calculations involving multiple unit doses were made in the computer model, but not in practice. One of these involved the choice of half a tablet which was possible in the computer model but not in the practical situation where the placebo tablets were unbreakable. The pilot study had identified this discrepancy between the two models of assessment but the reference group decided to leave the placebos as unbreakable since good practice guidelines are against breaking tablets.

Students found more difficulty in calculating between units where milligrams or grams were prescribed but the tablets were labeled in micrograms and milligrams respectively. There were a number of different wrong answers or failure to attempt these particular calculations in both the computer model and the practical situation, suggesting that students had difficulty with the calculation regardless of the method of presenting the problem.
Oral liquid medicine

The results from this section were analysed in two ways as reported under Results to make allowance for minor measurement errors in the practice environment. Nevertheless, we found a higher rate of discrepancy between computer results and those from the practical exercise in this section than in the tablets and capsules section. The unit dose calculations for liquid medicine were correctly done by most students on the computer but we observed a number of technical measurement errors in the practice environment. Several of these involved failure to dispel large air bubbles from the oral syringe used to measure the dose, resulting in a slightly smaller volume of medicine being available for administration than was suggested by the position of the plunger. In the vast majority of cases, this would not have been clinically significant. Other measured doses were incorrect by factors between 0.5mL and 22mL in the computer assessment and between 0.2 mL and 35mL in the practical assessment. Poor choice of delivery vehicle, for example a 30mL medicine pot used for a dose of 5mL or indeed for 30mL made some amounts difficult for the adjudicator to measure accurately, but this did not seem to occur to the students involved.

Multiple unit dose prescriptions were calculated correctly by most when using the computer, but again there were a number of technical measurement errors made in the practice environment, air in oral syringes being the most common cause. The students, even though they were either at the end of second year or in their third year of training, did not appear to have experience in drawing up liquids using syringes and some had very poor judgment of a suitable receptacle in which to measure the calculated amounts. Few were observed to be holding the medicine pot at eye-level to allow for a meniscus.

That fewer errors in the oral liquid medicine section were made on the computer model may be due to a number of factors. First, it was not possible to simulate the introduction of air into the syringe in the computer model. This has since been recognised as an item for design refinement for the future. Second, because the computer screen was at eye-level with clearly marked levels on the simulated medicine pot, filling it to the correct level was relatively easy. In the practice
environment, some of the markings on the pots were quite difficult to see and although the pot could be placed at eye-level by putting it on top of the medicine cabinet, few students did this. Consequently, they did not allow for the meniscus when filling the pot to the calculated mark and resulting amounts were slightly inaccurate because of this.

Conversion between SI units again caused difficulty for a number of students and more made errors on the computer in this section than in the unit and multiple unit dosage prescriptions. Similar errors were made in practice however, and congruence between the two assessment methods was relatively high suggesting that the calculations themselves were the main source of error.

**Injections**

Unit dose calculations were performed on the computer without error. The practical task resulted in a number of technical errors involving air in the syringe and also errors which suggested miscalculation of the amount required. One of these was an error by a factor of 5 and another by a factor of 10.

Multiple-unit dose errors were made in both assessment methods but again technical measurement errors in the practical use of syringes were common. The actual amount of fluid in a few syringes, particularly those of 1mL capacity, was negligible, suggesting that the student(s) in question concentrated on getting the plunger to the correct mark while failing to see that the needle tip was above the level of the fluid in the ampoule. Hence, a correctly calculated numerical value, although appropriately transferred to the correct measurement graduation point on the syringe barrel would in these cases manifest itself as a considerable sub-dose in reality.

The sub-unit dose prescriptions in this section produced the lowest congruence between the two assessment methods, with a high number of wide-ranging errors demonstrated in both environments. This would suggest that the calculation was sufficiently difficult or poorly understood that different
calculation errors were made each time. The commonly manifested technical measurement error of air not being displaced from the syringe accounted only for a small number of errors, leaving miscalculation to be the probable cause of the remaining errors.

The slightly lower level of congruence between the two methods of assessment for prescriptions relating to injections is difficult to explain. Lack of dexterity and practice using actual syringes and needles may account for many of the technical measurement errors observed but would not explain all the differences. Limited exposure to calculations for injections involving anything more complicated than a unit dose is a possible explanation which was borne out by students’ informal comments.

**Intravenous fluid delivery rates**

Participants were required to work out two rates of delivery for each of the fluid prescriptions. Calculating millilitres per hour tested their performance in setting the rate on a standard volumetric pump, while calculating drops per minute tested whether they were able to calculate the rate through a giving set without a pump attached.

The overall results showed fewer correctly calculated rates than the medicines sections, but there was a similar level of congruence between the two methods and a much higher congruency between incorrect answers suggesting that participants had more difficulty with the calculations themselves than the medium of the assessment.

Many students had not been exposed to calculating fluid delivery rates whether by volumetric pump or by IV fluid administration sets. They reported that they were not aware of any formula to help them to calculate the rate and subsequently performed more poorly in this section.
As there was evidence that many students had not been given the opportunity in the clinical area to apply the calculation of fluid administration rates to a practical situation, simulation, whether by computer or in the practice set-up, gave that opportunity and may be recommended as a practice tool.

We detected mathematical errors in drip rate calculations which involved decimal places. While some volumetric pumps permit the operator to enter decimal fractions, we asked the students to use a rounding technique to the nearest whole number. In spite of this instruction, drip rates were often calculated to 2 decimal places and left as this in the final answer. Participants did not recognise that 66.66 would round up to 67 nor that the answer needed to be in whole numbers when this related to drops per minute. The computer model marked such answers as incorrect. If assessed in practice, it could be argued that this type of error would have no clinical significance. It was apparent that students had had little exposure to this hands-on practice in setting IV fluid rates.

*General analysis of error classifications identified within the authentic computer environment and the practice environment*

Figure 13 illustrates a technical measurement error analysis model that defines the error classifications observed in the authentic computer environment and the practice environment:
Figure 13: Technical measurement error analysis model

This error analysis model illustrates that:

1. Of the four error classifications two were diagnosed in both the authentic computer program and the practice environment (inappropriate selection of a measurement vehicle; and incorrect measurement or setting of a medication dose or rate), and two were only measurable in the practice environment (inappropriate inclusion of minor air bubbles and air boluses in syringes).

2. If the student consistently provides an accurate technical measurement solution within a given domain of practice and at a given level of a hierarchical rubric then it can be assumed that they have also demonstrated conceptual and calculation competence within that specific domain (tablet and capsule, oral liquid medicine, injection or IV Infusion) and rubric level.
3. If the student provides an inaccurate technical measurement solution within a given domain of practice and at a given level of a hierarchical rubric then they have failed to demonstrate competence. However, importantly, the error may be conceptual, calculation, technical measurement or any combination of the latter in nature. Subsequently if the individual’s knowledge and skills within this domain are to be advanced then it is necessary at this point to undertake a diagnostic assessment of the underpinning and discrete inappropriate conceptual, calculation and technical measurement competency schemata that the student/practitioner manifests.

**Conclusions**

High congruence between results from the two methods of assessment in the tablets and capsules section suggests that for determination of calculation competence in management of this type of prescription, an authentic computer assessment is equivalent to an assessment through practice simulation.

In assessing calculation of liquid medicine doses, as illustrated in Figure 13, both the authentic computer environment and the practice environment assessments facilitated detection of technical measurement errors associated with selection of inappropriate measurement vehicles and measurement/setting of incorrect liquid doses or IV infusion rates. However, the definition of numeracy which we have used includes competence in knowing what the transfer of a calculated answer to a technical measurement vehicle means in practice and the computer model did not allow for technical measurement errors associated with failing to displace air bubbles and air boluses from syringes, error types which were manifested by several students in the practice environment. This was a measure of numeracy competency which has not been widely considered in the literature and was apparent in all the sections of the assessment which involved liquids. Some students made arithmetic or computational errors but the majority of errors were made in practical technical measurement. These errors were apparent in practice regardless of the accuracy of the original calculation of dose.
This element was appropriately identified in the practice simulation assessment and, if coupled with the authentic computer assessment, this competence could be assessed as a practical skill with any prescription requiring liquid medicine without recourse to repeated measures across the range of complexity. The same argument would apply to prescriptions for injections.

We conclude that for calculation of medicines dosage, the major advantage of the authentic computer environment was to provide prescriptions covering the full range of calculations likely to be met in practice. At the same time it allowed easy assessment of the mathematical element of these calculations with large numbers of students in a short time. In addition, given that marking and feedback generation was entirely automated, the process was quick, easy and totally objective. In assessing nurses’ calculation of medicine doses and IV fluid rates, an authentic computer model that presents dosage problems within an agreed rubric is invaluable in providing assessment of the full range of calculations likely to be met in practice as a newly qualified nurse.

In the practice context it would be impossible to ensure that all third year student nurses encountered and were reliably assessed upon the full range of problems supported by an authentic assessment process. However, the assessment of the numeracy element of a nurse’s competence in medicines management must include assessment of both the full range of calculations likely to be required (supplied through an authentic model – in this case computerised) and the measurement vehicle manipulation and measurement skills available in most clinical settings and/or able to be simulated in a practical environment.

We propose that the assessment tools and processes identified within this report provide a robust form of assessment that meets the needs of regulators, educators, employers, practitioners, students and public in reliably identifying conceptual, calculation and technical measurement competence in the context of medicines administration.

Further, we propose that this research provides a benchmark against which
other researchers and interested stakeholders can measure the impact of other innovations in learning, teaching and assessment strategies, and of recruitment, development and support/retraining strategies.

**Summary**

The main overall focus of this study was to determine the validity of the computer simulation format of delivering dosage calculation problems. The validity of the computer simulation format was tested against the gold standard practical simulation format. The underlying rationale was that the practical simulation format is not feasible for mass testing, particularly across the full range of question type (tablets, liquid medicines, injections, etc.) and complexity (single unit, multiple unit etc) whereas computer simulation enables testing across the range of question type and complexity of large numbers of people at the same time, in remote locations, with limited costs. If computer simulated testing could be shown to operate similarly (provide similar results) to practical simulation testing, this would validate the use of computer simulated testing for future research.

From the results presented above, the criterion-related validity of the computer simulation format has been supported, both in terms of putting participants in a similar order of competence and in terms of participants obtaining similar absolute results (getting the same number of questions correct on the computer simulation as they would on the practical simulation). These results supplement the more detailed item-by-item comparisons that Weeks has conducted (Weeks 2001), and which produced similar results in terms of confidence in the computer simulated testing format as a substitute method for practical assessment.

Some caveats remain, however:

- computer simulation does not test certain elements of the real-world dosage calculation problem (e.g., technical competency)
- these conclusions should only be applied to similar situations, populations, and constructs.
References


Annexe 1: Report of the pilot study

Numeracy for Nursing – creating a benchmark
Report of a Pilot study to compare outcomes of two medications dosage assessment methods using simulated reality

Diana Coben, King's College London
Carol Hall, University of Nottingham
B. Meriel Hutton, King's College London
David Rowe, University of Strathclyde
Mike Sabin, NHS Education Scotland/Scottish Government
Keith Weeks & Norman Woolley, Authentic World®, University of Glamorgan

Abstract
This is the report of a pilot study for the first stage of an inter-disciplinary project to establish a benchmark for numeracy in nursing by exploring the key issues in determining the achievement of competence in nursing numeracy. The project is funded by NHS Education for Scotland (NES).

Objective: To test the efficacy of a computer-based assessment of pre-registration nursing students' medication calculation skills by comparing outcomes from this with calculations presented in a practical setting. The pilot also acted as a test of the research design, and the participants’ evaluation of content will inform the suitability of items of assessment for inclusion within a benchmark standard.

Design: A quasi-experimental cross-over design in which half the participants were exposed to the computer-based assessment prior to undertaking the practical tasks, while the other half were exposed to the practical tasks before undertaking the computer-based assessment.
Setting: A large school of nursing in England with modern IT facilities to meet the specifications of the computer assessment and a large classroom for the practical activity.

Participants: Fifty volunteer early 3rd year (2008) students on the adult branch of a pre-registration nursing programme.

Main Outcome Measures:
Numerical scores from the computerised assessment compared with results from the practical exercise.

Results: Results showed a reasonable to very high level of congruence between the two methods (44%-100% congruence; mean congruence for n = 252 test item opportunities for error = 80% congruence).

Conclusions: Computerised assessment using this particular platform is likely to closely mirror medication calculations done in a practical setting.

Introduction

Medication errors are an aspect of clinical governance which has been highlighted recently by the National Patient Safety Agency (NPSA) and this issue is currently targeted for remedial action (NPSA, 2006a, 2006b). While medication errors are multi-factorial, lack of competence in numerical calculation is often cited as a key area of concern, especially with respect to medication dosage calculation (Weeks, Lyne, & Torrance, 2000). Numeracy is recognized as a key competence for professional practice in nursing, yet successive studies indicate a lack of proficiency within both the student population and amongst registered nurses (Hutton, 1997; Sabin, 2001). To inform the study, the following definition of numeracy was adopted:
To be numerate means to be competent, confident, and comfortable with one’s judgements 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.

(Coben, 2000, p. 35)

Numeracy in the healthcare context is much broader than medication calculation, but this is its most visible and commonly cited example, and greater consistency in assessment would strengthen support for the enhancement of learning and teaching approaches. From September 2008 the body regulating the profession in the UK, the Nursing and Midwifery Council (NMC) requires students to achieve 100% in a test of “numeracy in practice” (NMC, 2007) before being allowed to register as nurses, yet there are currently no national standards for what that test should include. Higher Education Institutions (HEIs) and employers are accordingly incorporating numeracy into their pre-registration nursing programmes and devising and setting their own tests of numeracy. Whilst these tests may be constructed on the basis of considerable professional expertise and with the best of intentions, as long as there is no recognised benchmark standard, such tests will remain variable both in level and quality; at worst, it is suggested, some may be professionally or mathematically inappropriate. Whilst it is accepted that in some instances, the establishment of benchmark standards may be controversial in terms of achieving minimum or best practice, the above considerations indicate that the development of a benchmark for nursing numeracy calculation is desirable in order to determine and debate which skills should be taught, and to ascertain when competence has been achieved.

The research team has made the case for such a benchmark fully elsewhere through international conference presentation and publication (Coben et al., 2008). The aim in this stage of the study is to address issues of parity and level in assessing competence in numeracy for successful calculation of medication dosages by nurses at the point of qualification. We focussed on medication
dosage calculation because it is probably the most widely-recognised and sensitive area of numeracy for nursing. Medication dosage calculation is one aspect (‘Calculation’) of ‘Medicines Management’ in the NMC’s ‘Essential Skills Clusters’ for pre-registration nurse education (NMC, 2007). There was no intention to assess ‘Medicines Management’ as a whole.

We believe that to assess such competence in practice with every student covering the full range of medication dosage calculations would be an extremely difficult task. Not only are students allocated to placements with very different dosage calculation opportunities, but it would be too complex a task to ensure that each student had been tested in a similar way. One solution would be to set up objective structured clinical examinations (OSCEs). However, these are extremely costly in terms of assessor/supervisor time and there is no guarantee that they would cover a full range of calculations.

From an extensive analysis of the literature, the following research-based criteria for an effective numeracy benchmark assessment tool was developed. Such a tool should be reliable, valid and capable of recreating the complexity of nursing numeracy in an authentic assessment environment. Specifically, it should be:

- **Realistic:** Evidence-based literature in the field of nursing numeracy (Hutton, 1997; Weeks, 2001) strongly supports a realistic approach to the teaching and learning of calculation skills, which in turn deserve to be tested in an authentic environment. Questions should be derived from authentic settings. A computer based programme of simulated practice in drug calculations, formative testing, with feedback on the nature of errors made, has been shown to develop competency in medication dosage calculation, which also can be demonstrated in the clinical areas (Weeks, Lyne, & Torrance, 2000). Exposure of students to real-world situations is recommended (Weeks, 2001).

- **Appropriate:** The assessment tool should determine competence in the key elements of the required competence (OECD, 2005; Sabin, 2001).
• **Differentiated:** There should be an element of differentiation between the requirements for each of the branches of nursing (Hutton, 1997).

• **Consistent with adult numeracy principles:** The assessment should be consistent with the principles of adult numeracy learning, teaching and assessment, having an enablement focus (Coben, 2000, Hodgen, Coben & Rhodes, 2008).

• **Diagnostic:** The assessment tool should provide a diagnostic element, identifying which area of competence has been achieved, and which requires further intervention (Black & Wiliam, 1998). Thus it should “provide information to be used by students and teachers that is used to modify the teaching and learning activities in which they are engaged in order better to meet student needs. In other words, assessment is used formatively to ‘keep learning on track’”. (Wiliam, 2006).

• **Transparent:** The assessment should be able to demonstrate a clear relationship between ‘test’ achievement and performance in the practice context (Weeks, Lyne, Mosely, & Torrance, 2001).

• **Well-structured:** The assessment tool should:
  
  • provide a unique set of questions with a consistent level of difficulty;
  
  • provide a structured range of complexity; and
  
  • take place within a defined framework, at points by which students can be effectively prepared, while allowing time for supportive remediation. (Hodgen & Wiliam, 2006)

• **Easy to administer:** The assessment should provide the opportunity for rapid collation of results, error determination, diagnosis and feedback (Black & Wiliam, 1998).

The pilot study described here focuses on establishing the validity and reliability
of the research instruments and the robustness of the research design. Medication dosage calculations covering a full range of complexity within the hierarchy, that is: unit dose, sub and multiple unit dose, complex problems and conversion of units, which had been previously tested for face and content validity by a team of pharmacists, nurse clinicians and nurse educators (Weeks, Lyne, Mosely and Torrance, 2001) were used as assessment items. The original assessment tool was designed by Keith Weeks and Norman Woolley as part of the Authentic World® program. This program is based on a constructivist-centred design drawn from the work of Piaget (1983), Bruner (1975), and von Glasersfeld (1987).

Aim of the study
The pilot study set out to test whether a computer-based assessment using the authentic diagnostic assessment outlined above mirrored the results of identical calculations presented in a (simulated) clinical environment. A pre-pilot was used to test the design of the practical tasks and the practicalities of the cross-over design of the study.

METHODS

Study design
A computer-based 40-item baseline assessment using web-based materials from Authentic World® was undertaken by all students who volunteered for the project. A representative sample of 20 students was then selected from these on the basis of results. Two weeks later, selected students’ performance in a further computer-based assessment of medication calculations was compared with their performance in a simulated practice situation involving the same calculations. A cross-over design took learning effects into account.

Participants
Following ethical approval by the pilot-site University and agreement from the course directors, 88 adult branch students from two cohorts commencing their 3rd year of Diploma/BSc Nursing Studies were approached. The cohorts
included traditional 3-year programme students and a small number of graduate entry students who had received a shortened first year programme. An initial hour-long introductory session by two members of the research team outlined the project and explained volunteers’ participation in it. All students were promised confidential feedback on their performance and a period of free access to the full web-based programme of instruction. Students were then invited to sign up as participants. These volunteers were given information sheets and asked to attend the preliminary computer baseline assessment on a given date three weeks later.

Fifty students attended for the baseline assessment. Because of the large number of volunteers, these were divided into two groups and allocated to either a morning or afternoon session. From the original 50, a representative sample of 20 was selected on the basis of their results in the baseline test to continue with the study. Of these, only nine students presented for the second stage.

**Tools**

**a) Computer-based assessment**

A computer-based assessment tool was devised from an existing programme of medication calculation instruction for nurses in which the necessary calculation skills are developed and assessed in an authentic context using high fidelity imagery of the elements of the dosage calculation problem (e.g. prescription charts, medication containers, syringes etc.). See Fig.1. This evidence-based programme (Authentic World®) has been refined over a number of years and is already used by several schools of nursing in the UK. It has been tested extensively for validity (Weeks, Lyne, Mosely, & Torrance, 2001). The 40-item baseline assessment drawn from the programme covered the full range of medication dosage calculations which the development team believed likely to be met by a student nurse in practice. (A further phase of the master study is planned to evaluate this assumption.)
Problem Presentation and Delivery Vehicle Selection

Answer Representation in Chosen Delivery Vehicle

Presentation of Feedback on Performance

Figure 1: Authentic World Screenshots
To accommodate the assessment of all individuals in the study, irrespective of the medication dosage calculation and problem-solving education process they had previously been exposed to, the assessment programme had been stripped of its constructivist-based modelling, coaching, scaffolding and detailed diagnostic feedback processes (although for participating in the research programme, all students were given access to the full programme, including detailed diagnostic feedback). On volunteering to participate in the study, students were issued with a log-in code which allowed them access to the Authentic World web-site and a specially prepared set of practice questions so that they could get used to the technology. The prepared programme covered competence in calculating doses of tablets/capsules, liquid medicines and injections, all at unit-dose, sub and multiple unit-dose, complex problem and conversion of SI unit levels. The final section included intravenous infusion rates for both volumetric pumps and manually controlled giving sets.

The assessment took place in a computer laboratory with one computer per student and two supervisors. After a short practice with the technology, students were asked to complete the 40-item baseline assessment. Two hours were allowed with an optional break between sections.

A representative sample of 20 participants were selected from the results of this baseline assessment and given a date to attend the next stage of the pilot. Of these, nine students presented for the second stage when they undertook a further computer assessment and a practical activity requiring the same calculations. To randomise sequence effects, half the students were allocated to the computer test in the morning and the practical in the afternoon, and vice versa. Each student was asked to use their ID from the original computer assessment on any paper used for working out and to identify their results.

The second computer assessment consisted of a 28-item test. Statistical work undertaken in the preparatory stage of the project confirmed the reliability of the selected Authentic World® test items; the internal consistency reliability of the selected items were $R = .92$. Six items were calculations involving tablets and capsules; six were liquid medicines and six injections. A further five calculations
were of intravenous (IV) fluid rates, each requiring a calculation for both volumetric pumps and manually controlled giving sets. Participants were supplied with paper and pencil for any working out they needed to do and they were given two hours in which to complete the assessment. An optional break was allowed during this time. Student participants were also required to evaluate the items of assessment for suitability as a benchmark standard, using an on-line evaluation tool.

b) The practical assessment activity

The practical activity had been designed for 10 students at a time. Five would be seated at tables arranged so that they could work independently but also be observed by a central invigilator in a similar manner to the computer assessment and the upright medicine cabinets would act as screens between the participants. The other five would be seated at tables where they would be presented with a labelled bag of fluid and giving set for each numbered ‘prescription’. As some students had notified withdrawal from the study prior to Stage 2, not all 20 students were expected and so the practical room was set up to accommodate eight students with a minimum of three supervisors. The four medicine cabinet ‘stations’ were arranged with the cabinets upright and open on the corner of each table thus acting as screens between students. (See Fig.2) Paper and pencils were provided at each ‘station’ and if used, collected afterwards.

The practical activity required students to do exactly the same medication calculations as in the computer assessment. These were presented in two sections: medicines calculations covering tablets and capsules, liquid medicines and injections; and calculations of IV rates. The medicine section consisted of three colour-coded prescription sheets (formatted similarly to those used in the NHS) to be worked through one at a time. One sheet was of prescriptions for tablets and capsules, one liquid medicines and one injections. Each sheet contained six numbered prescriptions and was presented on a colour-coded dinner tray (labelled with the student’s project ID number). Six small papier-
 maché trays (also colour-coded and numbered 1–6) were provided per dinner tray/prescription sheet (See Fig. 3).

Figure 2: Layout of room (accommodating 8 ‘stations’)

The dinner trays were stacked in random order and students were instructed to tackle them one at a time by selecting the correct medicine from the medicine cabinet, preparing it as for presentation to a patient and putting the medicine pot or syringe relating to each prescription on the appropriately numbered papier-maché tray and placing this onto the dinner tray (see Fig. 4). When they had completed the six items, or as many as they could, the students signalled for the tray to be collected for ‘marking’. They then dealt with the next tray.
Half the students did the medicine calculations first and the other half did the IV rates first. Students were given a half-hour break between the different tasks. The process was invigilated by three researchers; one dedicated to the IV section: the others covering the medicines section. Students’ results were
recorded as they finished each task or shortly afterwards if trays were ‘stacking up’.

The IV fluid calculations had to be demonstrated using bags of fluid and IV giving sets with differing drop factors (delivering different numbers of drops per millilitre) or attached to a volumetric pump. An IV prescription sheet was provided per student with 5 prescriptions for different fluids to be delivered over differing time periods. As only one type of giving set was available, each IV set-up was labelled with a drip factor relating to the calculation required in the on-line test. Students were required to calculate two rates per prescription – one in millilitres per hour for entry into the volumetric pump and the other for drops per minute. When they were ready to set the rate, they signalled to the invigilator who observed them setting the rate in ‘practice’ by adjusting the manual roller on the giving set and in-putting the correct numbers on the volumetric pump. This model had been tested during the pre-pilot materials testing exercise and it was considered impractical for the invigilator to observe and count the drops each time. Students were therefore asked to tell the invigilator how many drops they were setting the rate at and this was recorded on the ‘mark’ sheet. Another problem was many students’ lack of experience in setting rates on different volumetric pumps and so the invigilator was allowed to key in the numbers dictated by the student.

Calculators

Student participants were not allowed to use calculators. This is not because we object in principle to the use of calculators by nurses or nursing students – see Hutton (1998) for a discussion of this issue, but banning the use of calculators would enable us to standardize the testing procedures across all participants.

Data collection and analysis

The computer-based assessment data were automatically entered into a spreadsheet while the practical scores were entered by hand and transcribed for analysis by the psychometrics specialist member of our team. The transcription was done by 2 people to avoid transcription errors. The same answer given on
both computer and in the practical was recorded as congruence (whether correct or not). Any deviation was considered incongruent.

**Results**

**Tablets and Capsules**

Calculations were correct and concordance between performance in the computer assessment and in the practical activity was 100% for all but one of the calculations involving tablets and capsules (mean congruence = 98%). The incongruent results involved 2 students using a half tablet which was available in the computer model but not in the practical.

**Liquid medicines**

Congruence for this section varied from 44–89% (mean congruence = 74%). The majority of students gave the correct answer on the computer but mistakes were made in measuring the actual amount in practice.

**Injections**

Congruence for these items again varied from 56-89% (mean congruence = 73%). Inaccuracies occurred in both the computer and practical assessments, but most errors were made in measuring the amounts in practice.

**IV rates**

Congruence for these items ranged 56-100% (mean congruence = 77%). Overall student performance was generally poorer with more incorrect answers. Congruence was demonstrated several times between wrong answers in both models which suggested that the student was making the same error in both situations. This had been demonstrated to a limited extent in the other sections.
Table 1: Comparison of students’ results by section (n=9)

<table>
<thead>
<tr>
<th>Prescription item</th>
<th>Correct</th>
<th>Incorrect</th>
<th>Measurement of Congruence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Computer Practice</td>
<td>Computer Practice</td>
<td>Between performance on computer/practice tests</td>
</tr>
<tr>
<td>Tablets and capsules</td>
<td>1</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>Minimum</td>
<td>9/9</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Maximum</td>
<td>9/9</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>8/9</td>
<td>89%</td>
<td></td>
</tr>
</tbody>
</table>

| Liquid medicines | 1 | 9 | 7 | 0 | 2 | 7/9 | 78% |
|                  | 2 | 9 | 7 | 0 | 2 | 7/9 | 78% |
|                  | 3 | 9 | 8 | 0 | 2 | 7/9 | 78% |
|                  | 4 | 8 | 5 | 1 | 4 | 5/9 | 56% |
|                  | 5 | 9 | 5 | 0 | 4 | 5/9 | 56% |
|                  | 6 | 8 | 6 | 1 | 3 | 7/9 | 78% |
| Minimum          | 5/9 | 56% |
| Maximum          | 7/9 | 89% |
| Mean             | 7/9 | 74% |

| Injections | 1 | 5 | 8 | 4 | 1 | 6/9 | 67% |
|            | 2 | 7 | 8 | 2 | 1 | 6/9 | 67% |
|            | 3 | 8 | 8 | 1 | 1 | 9/9 | 100% |
|            | 4 | 8 | 9 | 1 | 0 | 8/9 | 89% |
|            | 5 | 7 | 9 | 2 | 0 | 7/9 | 78% |
| Minimum    | 6/9 | 67% |
| Maximum    | 9/9 | 100% |
| Mean       | 7/9 | 73% |

| IV rate (pump) | 1 | 6 | 4 | 3 | 5 | 7/9 | 78% |
|               | 2 | 2 | 5 | 7 | 4 | 6/9 | 67% |
|               | 3 | 4 | 3 | 5 | 6 | 8/9 | 89% |
|               | 4 | 6 | 6 | 3 | 3 | 7/9 | 78% |
|               | 5 | 6 | 4 | 3 | 5 | 5/9 | 56% |
| Minimum       | 6/9 | 67% |
| Maximum       | 7/9 | 89% |
| Mean          | 7/9 | 74% |

Discussion

From the original analysis of the data, it was apparent that performance in the two tests of medicines calculations was reasonably well correlated overall.
However, if minor errors, such as failing to round as instructed, (e.g., for IV infusion rate per unit time calculations entering 166.6 or 166 instead of 167 mL per hour) were excluded and apparent sloppiness in practical measurement were excused, then a greater degree of congruence was reached. See Table 2. The small number of participants precluded any formal statistical testing at this point.

Table 2: Summary of Pilot Data Results allowing for factors outside the calculation itself.

**Congruence between Authentic Computer Assessment & Simulated Practice Assessment**

<table>
<thead>
<tr>
<th></th>
<th>Minimum Congruence</th>
<th>Maximum Congruence</th>
<th>Mean Congruence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tablets &amp; Capsules</td>
<td>78%</td>
<td>100%</td>
<td>98%</td>
</tr>
<tr>
<td>Liquid Medicines</td>
<td>44%</td>
<td>89%</td>
<td>74%</td>
</tr>
<tr>
<td>Injections</td>
<td>56%</td>
<td>89%</td>
<td>73%</td>
</tr>
<tr>
<td>IV Infusions</td>
<td>56%</td>
<td>100%</td>
<td>77%</td>
</tr>
</tbody>
</table>

Extraneous factors were observed which demonstrated that although the actual calculation of drug dosage or infusion rate was tested adequately by both models, neither model could capture all the issues around safe drug calculation in practice. Given our original definition of numeracy, we felt that these needed expansion.

1. Drawing up the correct amount into a syringe

In the computer model, the syringe filled as the plunger was drawn down but there was no facility for air to be drawn in as well. During the practical activity, several students drew in air with the ‘drug’ and failed to expel it before presenting the completed task, thereby preparing a wrong volume of drug in spite of a ‘correct’ reading of the level of the plunger. In discussion with the
computer programme designers, it was decided that although air entry could be built-in to the program in future, it was not possible to change it for the impending main study and so we decided to record both the plunger level and the actual amount of fluid in a syringe where these differed. Feedback to the student would include this error but if the plunger level was at the same level as the computerised plunger for that calculation, we would record the data as concordant. It was decided that the medicines management issues which these errors illustrated needed to be included in the generalised report to the participating Institution.

2. Amount of drug available

In the practical activity we had supplied an appropriate sized ampoule containing placebo drug for each prescription, but by doing this, the students were given an extra cue to make their answer no more than that amount. The computer model allowed such a mistake to be made. For example, on the computer assessment, a student could provide an answer of 4 mL when the answer should have been 0.4 mL. In the practice situation, they were only presented with one 2 mL ampoule and so could not make that mistake. We decided that for the main study we would provide 10 mL plastic ampoules of every placebo drug. This would allow students to make this error in practice and at the same time remove the danger of injury from the 2 mL glass ampoules.

The non-concordant result for tablets and capsules involved students using the half tablet facility on the computer model. There was a reluctance to provide placebos which could be broken, as this was considered poor practice, but again the computer programme reflected the situation in reality where some tablets are scored and half tablets may on occasion be given. After some discussion, it was decided not to alter this for the main study as the correct answer did not require a half tablet.

Evaluation from the students involved in the pilot was in most cases very positive and agreed that these were the sorts of calculations which they had
encountered in practice. Negative comments were associated with the drip rate calculations which some students thought were irrelevant to their practice because they had not yet been given responsibility for intravenous infusions.

**Conclusion**

The pilot study was invaluable in resolving practical issues such as those discussed above. It also provided early indications that the two methods of assessing medicine dosage calculations produced similar results (mean congruence between student performance within the two environments = 80%). This would suggest that this particular computer assessment of drug calculations should give a good indication of a student’s performance in practice. What it also showed was that while the computer model was able to assess conceptual and calculation skills and to an extent technical measurement skills, it could not assess all of the wider numeracy issues involved with safety in medicines measurement and medicines management (for example dispelling air from a syringe).

The pilot tested a small number of students from one school of nursing and so the evaluative comments, although informative gave a limited view. The main study will be carried out on a larger sample from a number of different schools of nursing and will provide a much larger data set, both from the assessments and from the evaluations.
References


The role of constructivist learning approaches and a computer based 'Authentic World' learning environment. Unpublished PhD, University of Glamorgan, Pontypridd.

