A Novel Approach to formulae production and overconfidence measurement to reduce risk in spreadsheet modelling

Simon Thorne
University of Wales Institute
Cardiff, UK

UWIC

Motivation

- EuSpRIG is well aware that spreadsheets are a problem!
  - Countless instances of erroneous spreadsheets
- And have proposed methodologies, auditing methods and software solutions.
- Our aim, at UWIC, is to make spreadsheets accurate from the very start, without damaging the flexibility that they posses.

UWIC

EuSpRIG 2003

- EuSpRIG 2003, Dublin
  - Ray Panko suggested that we should be considering the problem in a wider context.
    - In particular, Human Factors
- Hence my aim has been to consider the synthesis of a novel approach based upon greater understanding of human factors

Human Factors

- Several Contributing issues
- Base Error Rate (Panko, 1998)
  - Pre-disposal to error on simple repetitive tasks
- Overconfidence
  - Too much faith in one's abilities, leading to complacency
- Miller's threshold (Human Working Memory limit)
  - Errors caused by handling greater than 9 concepts simultaneously

Example Driven Modelling

- Taking example user provided input as a basis for formulae production
- The philosophy of this approach concerns human and machine learning (Michie, 1979)

Comparing Humans and Computers (conventional)

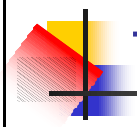
	Pattern matching	Generating real-world examples	Manipulating mathematics	Logical deduction
Human	Y	Y	?	?
Computer	N	N	Y	Y

Comparing Humans and Computers (conventional)

- Humans are good at giving real-world examples, weak at generating formulae
- Computers are good at manipulating mathematics (ALU), weak at generating real-world examples

Comparing Humans and Computers (conventional)

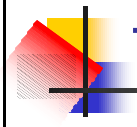
	Produce formulae	Generate real-world examples
Human	WEAK	STRONG
Computer	1 STRONG	2 WEAK



The Experiment

- Designed using Stanley and Campbell (1969)
- Series of 10 short tests
  - 5 using a 'Traditional' approach (control)
  - 5 using Example Driven Modelling (EDM)

UWIC



The Participants

- Undergraduate (final year) and Masters level students
  - Business Information Systems and Information Systems respectively
  - Random selection from within these groups
  - Students have experience of spreadsheets throughout their courses
    - End User Computing package module

UWIC

The Task

- Participants were given a refresher lecture and document
  - The document contained the construct of various statements in Excel
  - The document contained all the code necessary to complete a working answer for all questions
- Five scenario questions for which a working answer was required.
  - Working answers were determined from visual inspection and testing using known data

UWIC

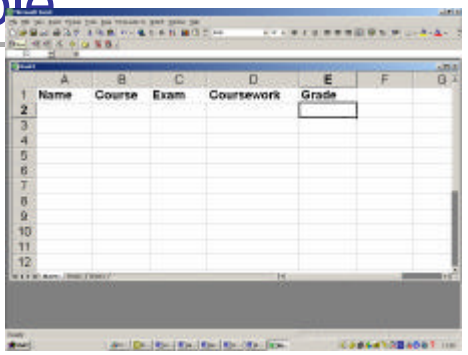
Traditional approach example

- **Question 1**
  - "Create a formula that will output grade as pass or fail in the spreadsheet below. To gain a pass the average of exam and assignment percentage marks must be greater or equal to 40, otherwise the grade is a fail."
- Traditional approach answer (One possible solution)
  - =IF(AVERAGE(C2:D2)>= 40, "Pass", "Fail")

UWIC

EDM example

- Same question



- Answer:
  - Pass: Coursework 40, Exam 40 (Average 40)
  - Fail: Coursework 20, Exam 30 (Average 25)

Complexity and error

- Relative question complexity was assessed using Halstead's Complexity.
  - Adapted from Halstead's Difficulty

$$Complexity = \frac{2*n1}{n2*N2}$$

n1 = the number of distinct operators
n2 = the number of distinct operands
(N1 = the total number of operators)
N2 = the total number of operands

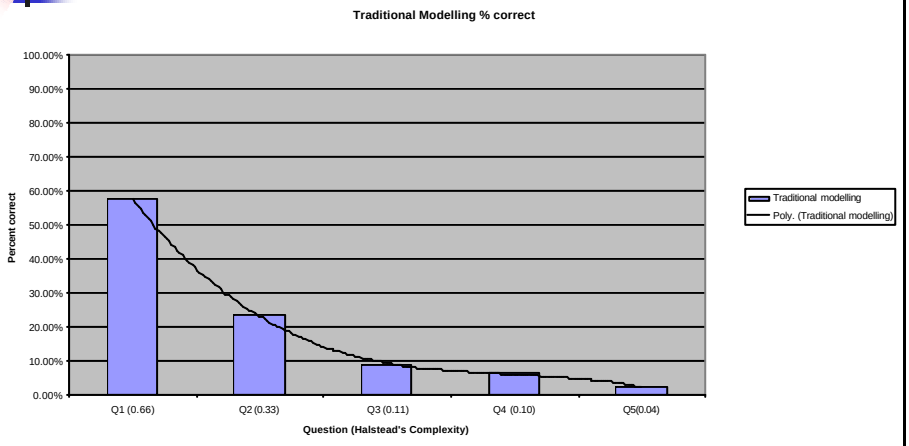
Confidence

- Construction of a new method for statistically analysing confidence.
  - Comparison of expected and actual, gained from a confidence questionnaire.
  - Gives a result of 0-5, 0 indicates a perfect match between expected and actual
  - Above 0 indicates overconfidence by varying degrees
  - Value can also range to one fifth below 0, results below 0 indicate **under confidence**

Overconfident?

- During one of the sessions being run, a student finished notably early.
- I commented on how quick he was (15mins)
- His reply was: "Yeah, it was easy! I used to be a spreadsheet programmer for a Bank in Hong-Kong"
- This participant's percentage of models with error was a terrifying **70%**

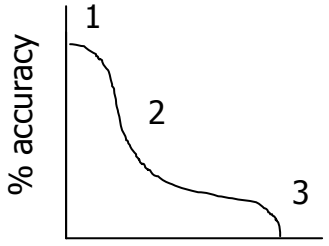
Traditional modelling



20% accuracy

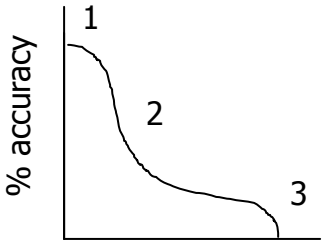
Human Factors evident?

- Using Panko's notion of Base Error Rate (Panko, 1998), we can theorise the percentage accuracy will never reach 100%. This causes a 'kink' in the otherwise negative exponential graph. (see 1)



Human Factors evident?

- The middle sector is a negative exponential. (see 2). A similar kink, to that of base error rate occurs at high complexity. See 3



UWIC

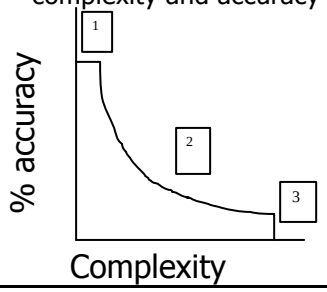
Why so kinky?

- I believe that the upper complexity kink can be explained by Miller's threshold.
  - Miller threshold (Miller, 1965) describes the well known limitations of Human Working Memory (HWM)
  - He states that HWM can comfortably handle "seven concepts, plus or minus 2 simultaneously"

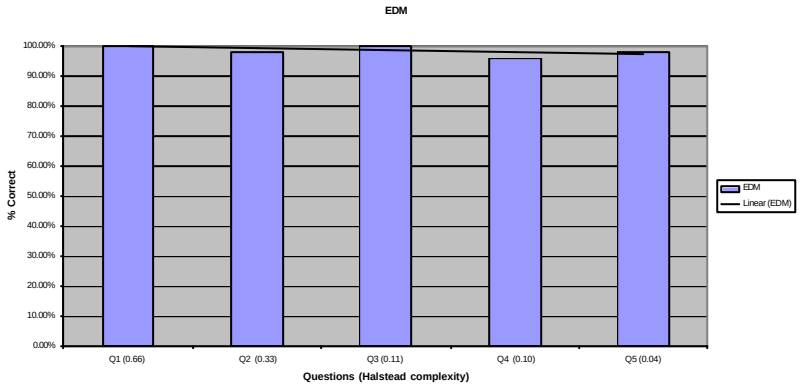
UWIC

Graph adjustment

- The kink at high % accuracy can be attributed to Base Error Rate (BER).
- The kink at high complexity can be attributed to Miller's threshold.
  - Adjusting the graph to represent this, the true nature of the complexity and accuracy model is revealed.



EDM

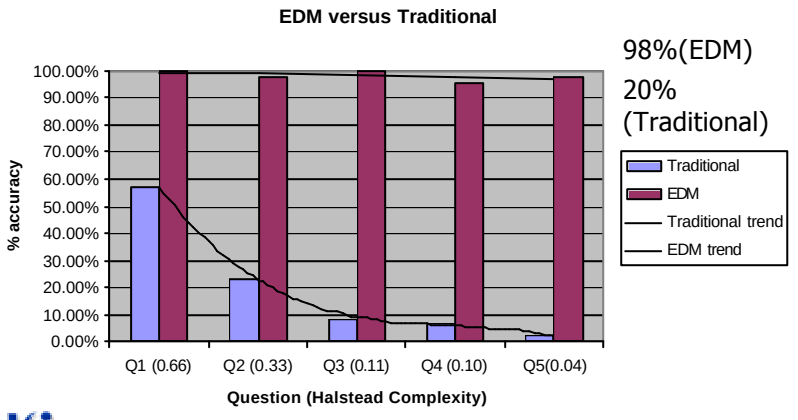


98% accuracy

Applicability of EDM

- So far, only one domain has been considered
  - Not a business domain
- In order to get a feel for the wider applicability of this technique, more environments need to be tested
  - Spreadsheet building environment in HMCE

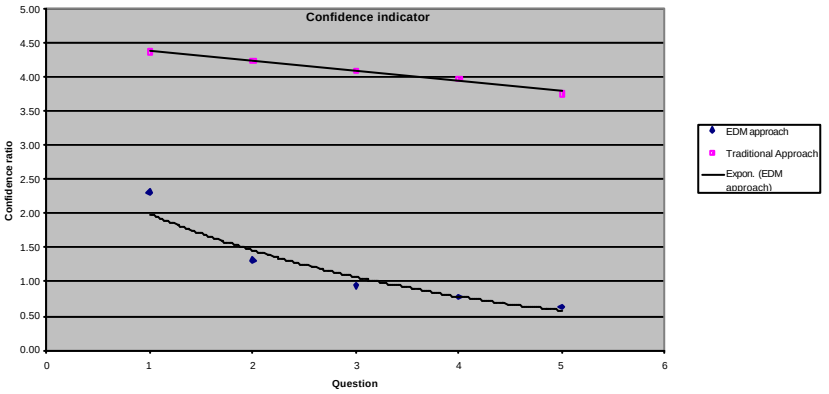
EDM and Traditional modelling



98%(EDM)

20% (Traditional)

Overconfidence



Conclusions

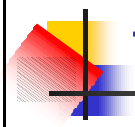
- The Error and Complexity relationship is a negative exponential
  - Limited at either extremity by BER (Panko, 1998) and Miller's threshold (Miller, 1956)

Conclusions

- EDM demonstrates far superior accuracy than traditional methods
- Modellers were significantly overconfident for the traditional modelling tasks
  - EDM confidence results indicate that modellers were much better at matching their perceived performance against actual resulting, on occasions, in under confidence.

What next?

- This is a R.I.P. (Research In Progress)
- A slightly adjusted experiment is planned in a spreadsheet building environment.
  - HMCE
  - We need you! (And your money!!!!)



Thank you.

Any questions?

Sthorne@uwic.ac.uk