Measuring Conceptual Understanding:
The Case of Teaching with Abstract and Contextualised Representations

Ian Jones, Matthew Inglis, Camilla Gilmore and Marie-Josée Bisson
Mathematics Education Centre, Loughborough University

Executive Summary.

Introduction.
The difficulty of measuring conceptual understanding presents a barrier to progress in the development and practice of high-quality mathematics education interventions. Conceptual understanding is commonly defined as deep knowledge of the underlying concepts of mathematics and how they relate to one another (Crooks & Alibali, 2014). Innovative methods for teaching mathematics are commonly claimed to impact positively on students' conceptual understanding; yet if conceptual understanding cannot be measured efficiently and reliably then robust evidence cannot be established. A recent and high-profile example of this problem is the debate over whether it is better to teach mathematical topics using abstract or contextualised representations. Some scholars have concluded that abstract representations are preferable (e.g. Kaminski et al., 2008) while others have come to more equivocal conclusions (e.g. Brown, McNeil & Glenberg, 2009). Key to these disparate conclusions is the lack of agreed and trustworthy measures of conceptual understanding (De Bock et al., 2011). As such, the current trend towards grounding mathematics curricula in real-world scenarios (ACME, 2012; MEI, 2012; Gowers, 2012; Truss, 2012) lacks an evidence base.

In the research reported here we developed a measure of conceptual understanding using a Comparative Judgement (Pollitt, 2012) approach, and demonstrated its application to the abstract vs. contextualised debate. Comparative Judgement (CJ) is a way to assess open-ended and creative mathematical work. It involves no mark schemes and no marking because such traditional methods cannot reliably be applied to assessing open-ended work (Laming, 1990). Instead two pieces of student work are presented on a screen and the assessor is asked to decide which is "better". The decision may be based on a specific objective, such as "the better understanding of fractions", or may be general, such as "the better mathematician". This is a binary decision. There is no need to decide how much better one piece of work is than the other. When many such pairings are shown to many assessors the decision data can be statistically modelled to generate a score for each student. The statistical modelling also produces quality control measures, such as checking the consistency of the assessors. Previous research has shown the comparative judgement approach produces reliable and valid outcomes for assessing the open-ended mathematical work of secondary and undergraduate students (Jones & Alcock, 2014; Jones, Inglis, Gilmore & Hodgen, 2013).

Objectives.
There were two objectives to the research reported here.
1. To apply CJ to measuring the learning outcomes of randomised controlled trials in which students are taught key concepts.
2. To provide valid and reliable evidence on the relative benefits of abstract and contextualised representations for introducing key concepts to students.

To achieve these objectives we undertook five studies. The first three studies investigated the feasibility of using CJ to measure understanding of key concepts across a range of contexts. The final two studies applied CJ to determining whether abstract or contextualised representations are superior for introducing two key concepts to students.

Studies 1a, 1b and 1c. Measuring understanding of key concepts.
Secondary school and university students completed open-ended tests on three concepts: the role of letters in simple algebra; derivatives in calculus; and *p*-values in statistics. These concepts were chosen because, unusually, validated measures have been developed in these areas and so provided a yardstick for evaluating the CJ approach. We found that student scores based on expert pairwise judgements of the open-ended tests correlated with the traditional test scores and with students' general mathematics achievement. This suggests that the CJ approach enabled the valid assessment of students' understanding of the three concepts.

Study 2. Abstract vs. contextualised representations: The case of algebra.
We first investigated whether CJ could be used to detect group differences in a randomised controlled trial. The focus was on the abstract vs. contextualised debate for the case of introducing letters in algebra to primary school students. A total of 189 students were randomly assigned to two groups and each received a series of three specially designed algebra lessons. One group was taught algebra using the MiGen software (Noss et al., 2012), which offers a broadly contextualised approach to learning mathematics; the other group was taught using the Grid Algebra software (Hewitt, 2014), which offers an abstract approach. Following the intervention, the students' understanding of the role of letters in algebra was tested using an open-ended test, which was then assessed by experts using CJ, and a traditional test. We found that the Grid Algebra group outperformed the MiGen group on both measures, although the difference between groups was larger for the open-ended test. In conclusion then, for the case of introducing algebra to primary children, the abstract approach, as exemplified by Grid Algebra, produced measurably greater learning gains. Moreover, the open-ended CJ-based test was slightly more sensitive than the traditional test at detecting this difference.

Study 3. Abstract vs. contextualised representations: The case of calculus.
We then investigated whether CJ could be used to detect group differences under more tightly controlled conditions. The focus was again on the abstract vs. contextualised debate, this time for the case of introducing differential calculus to high-achieving secondary students. 189 students were randomly assigned to

two groups and each received a series of three calculus lessons. Unlike for Study 2, the lessons were identical except that the materials drew on real-world examples (e.g. accelerating vehicles) for one group, and used only abstract representations (mathematical symbols and graphs) for the other group. Following the intervention, open-ended CJ-based post-tests and traditional post-tests were administered to measure the students' understanding of the concept of derivative. We found no difference in overall performance between the two groups on either of the measures. Thus, for the case of introducing calculus to high-achieving secondary students, neither abstract nor contextualised representations produced measurably greater learning gains.

Findings.
There are two main findings from the research. First, CJ can be used to evaluate students' conceptual understanding, and to evaluate the relative effectiveness of different teaching approaches. As such, a significant barrier to progress in the field of mathematics education can now be overcome; namely, the paucity of effective measures of students' conceptual understanding in different domains. Our contribution will enable researchers to evaluate and understand the effectiveness of various educational resources and approaches more quickly and validly than has been possible to date. This in turn will provide policy-makers and teachers with better evidence about the relative effectiveness of educational interventions.

Second, we have informed the abstract vs. contextualised representations debate by providing evidence on relative effectiveness in two contexts. For the case of algebra we compared two technology-specific approaches to teaching using abstract and contextualised representations. We found that an abstract approach using the Grid Algebra software was more effective for learning about letters in algebra than a contextualised approach using the MiGen software. For the case of differential calculus we compared two specially-designed sets of teaching resources. We found that an abstract approach, using formal representations such as symbols and graphs, and a contextualised approach, using real-world representations such as accelerating cars, were equally effective for learning about the concept of derivative. We conclude that the role of abstraction and contextualisation when teaching mathematics is nuanced, and effectiveness depends on the concept being taught, the approach used, and perhaps the age and prior achievement of learners. Importantly, the CJ approach enabled us to overcome the measurement problem that has limited the findings of previous research.

References.
ACME. (2012). *Post-16 Mathematics*: *A Strategy for Improving Provision and Participation*. London: Advisory Committee on Mathematics Education.

Brown, M. C., McNeil, N. M., & Glenberg, A. M. (2009). Using Concreteness in Education: Real Problems, Potential Solutions. *Child Development Perspectives, 3*, 160–164.

Crooks, N. M., & Alibali, M. W. (2014). Defining and measuring conceptual knowledge in mathematics. *Developmental Review, 34*, 344–377.

De Bock, D., Deprez, J., Van Dooren, W., Roelens, M., & Verschaffel, L. (2011). Abstract or concrete examples in learning mathematics? A replication and elaboration of Kaminski, Sloutsky, and Heckler's study. *Journal for Research in Mathematics Education, 42*, 109–126.

Gowers, T. (2012). *How Should Mathematics be Taught to Non-Mathematicians?* Blog post, https://gowers.wordpress.com/2012/06/08/how-should-mathematics-be-taught-to-non-mathematicians/

Hewitt, D. (2014). A symbolic dance: the interplay between movement, notation, and mathematics on a journey toward solving equations. *Mathematical Thinking and Learning, 16*, 1–31.

Jones, I., & Alcock, L. (2014). Peer assessment without assessment criteria. *Studies in Higher Education, 39*, 1774–1787.

Jones, I., Inglis, M., Gilmore, C., & Hodgen, J. (2013). Measuring conceptual understanding: The case of fractions. In A. M. Lindmeier & A. Heinze (Eds.), *Proceedings of the 37th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 113–120). Kiel, Germany: IGPME.

Laming, D. (1990). The reliability of a certain university examination compared with the precision of absolute judgements. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 42*, 239–254.

MEI. (2012). *Integrating Mathematical Problem Solving: Applying Mathematics and Statistics Across the Curriculum at Level 3. End of Project Report*. London: Mathematics in Education and Industry.

Noss, R., Poulovassilis, A., Geraniou, E., Gutierrez-Santos, S., Hoyles, C., Kahn, K., … Mavrikis, M. (2012). The design of a system to support exploratory learning of algebraic generalisation. *Computers & Education, 59*, 63–81.

Pollitt, A. (2012). The method of Adaptive Comparative Judgement. *Assessment in Education: Principles, Policy & Practice, 19*, 281–300.

Truss, E. (2012). *Elizabeth Truss Calls for a Renaissance in Maths*. Norfolk: Speech to the National Education Trust. Retrieved from https://www.gov.uk/government/speeches/elizabeth-truss-calls-for-a-renaissance-in-maths