Exploring formative assessment and peer feedback in technology-enhanced mathematics learning environments using bar model virtual manipulatives

Poh Hwee Sim, Claire

Charles University, Faculty of Education, Prague, Czech Republic; clairepohhs@gmail.com

The uptick in the adoption of digital assessment, driven by increased technology integration in classrooms, not only transforms the assessment approach but also holds crucial implications for how teachers assign tasks and shape the way students engage in mathematical reasoning. This paper explores how educators leverage technology to enhance mathematics assessment and feedback. Using screencast (or screen recording) as a primary method of data collection, interactions of students solving word problems utilising the bar model, a web-based virtual manipulative, are recorded. Analysis of data collected may offer insights into students' specific competencies and deficiencies and inform teaching practices to meet their' learning needs. Digital assessment is broadened to include peer feedback and self-evaluation, facilitated by real-time interaction and idea-sharing through screen mirroring, another innovation supported by classroom connectivity.

Keywords: Digital assessment, peer feedback, bar model.

Introduction

Technology has significantly impacted the assessment and feedback processes in teaching mathematics for both summative and formative purposes. The shift from offline to digital assessment introduces fresh possibilities for evaluating mathematics learning and modifying task structures as well as the scope of assessed abilities and skills (Drijvers et al., 2016). The concept of a connected classroom where teachers and students can exchange digital information has persisted for many years (Stacey & Wiliam, 2013). In recent years, technological advancements have rendered this vision more attainable. In the study by Clark-Wilson (2010), using classroom aggregation technology for mathematics was found to have promoted peer assessment as well as self-evaluation. Moreover, it was observed that teachers used feedback from students to inform the planning of future activities. These findings show that classroom connectivity offered fresh possibilities for formative assessment, providing teachers with insights into students' mathematical thinking. Leveraging classroom connectivity, multiple devices may be screen mirrored on the class display for side-by-side comparison and discussion among students. Studies suggest that students adjust their responses when comparing their work with peers, fostering increased opportunities for peer assessment and self-evaluation (Stacey & Wiliam, 2013). Drijvers et al. (2016) outline two crucial steps in formative assessment: collecting data on student achievements and devising strategies to enhance performance. Screen recordings serve as primary data. Teachers review these recordings and formulate appropriate measures to improve performance.

Theoretical Background

Framework for technology-mediated feedback

Mayer's (2002) Cognitive Theory of Multimedia Learning posits that learning is enhanced when information is presented through multiple modalities, such as visual and auditory channels. Figure 1 illustrates a cognitive theory of multimedia learning. Mayer asserts that multimedia messages that engage these cognitive processes are more likely to promote meaningful learning. His findings support a social agency extension of the cognitive theory of multimedia learning, suggesting that social cues within multimedia messages activate a conversational schema in learners, prompting deeper cognitive engagement. This holds significant implications for how the theory will shape peer feedback and assessment practices. This approach resonates with established conversational theories like Grice's (1975) conversational norms, which underscore the dedication to comprehending the other speaker's communication. Two approaches exist for assessing learning: retention tests and transfer tests (Mayer & Wittrock, 1996). Retention tests assess the capacity for memory recall. Transfer tests evaluate how effectively learned knowledge is applied to novel situations. He asserts that transfer tests offer the best assessment of learner understanding. Mayer's hypothesis proposes that better transfer is facilitated through interaction and conversation.

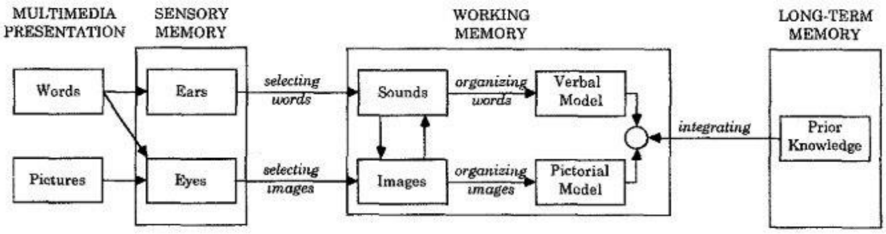


Figure 1: A cognitive theory of multimedia learning (Mayer, 2002, p. 103)

Findings suggest that side-by-side comparison techniques, such as Visual Analysis for Image Comparison (VAICo), offer advantages in terms of speed, clarity, and accuracy in identifying differences in image data (Schmidt et al., 2013). Figure 2 shows various image sets from diverse domains, showcasing the method's adaptability across datasets. A crucial aspect of effective data analysis involves selecting suitable similarity metrics. We suggest teachers adopt this selection method when choosing screenshots of students' solution for side-by-side screen comparison to teach bar model strategies. Rittle-Johnson et al. (2017) underscore the significance of comparison in conceptual learning, with their classroom-based research supporting its efficacy in algebra instruction. Mayer (2002) suggests expanding the cognitive theory of multimedia learning to include social factors affecting learners' engagement in deep cognitive processing, such as combining visual (e.g., selective screenshots) and verbal (e.g., feedback interaction) models. Building on the body of research, we argue that when teachers present students' solutions using side-by-side screen for comparison, they foster real-time sharing and collaboration among students, contributing to multimedia learning through peer feedback.

Framework for the model method

A distinctive pedagogy of Singapore mathematics, the model method is inspired by Greeno's part-whole and comparison schemas (Nesher, Greeno & Riley, 1982; Kintsch & Greeno, 1985). Students use rectangular bars to visualize mathematical relationships, facilitating comprehension of abstract quantities. For discussion in this study, consider the following illustrations of part-whole and comparison models.

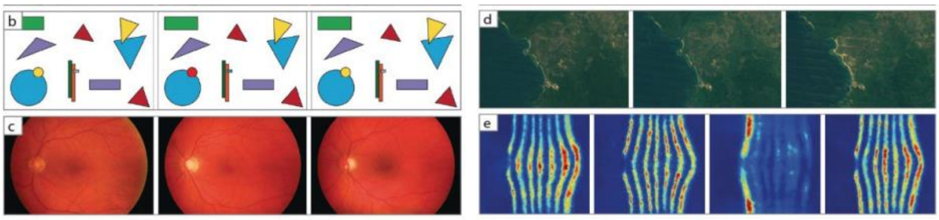
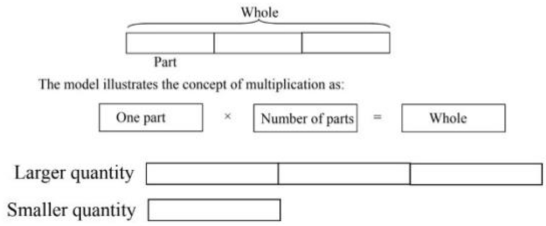


Figure 2: Image datasets (Schmidt et al., 2013, p. 6) (b) shapes disappear, re-appear or change their colour (c) retina images from different patients (d) satellite images of a coastline affected by tsunami in Indonesia (e) images with colour coded gene expression information

Table 1: Part-whole model for multiplication and division & multiplicative comparison models (Kho et al., 2014, p. 227)

	Part-whole model: The total is determined by the multiplication of one part and the number of parts. Conversely, if we know the total and one factor, we can find the other factor through division. Comparison model: The larger quantity is three times the smaller quantity, and conversely, the smaller quantity is one-third of the larger quantity. For example, if the larger quantity represents three units, the smaller quantity represents one unit. Together, they total four units, with a difference of two units between them.

Research Questions

The study seeks to examine digital assessment and feedback through an extension of mathematical experience using integrated technology. The research questions guiding the study are as follows:

RQ1. How can teachers leverage classroom connectivity to effectively analyse students' conceptual deficiencies in word problem solving utilising bar model virtual manipulatives?

RQ2. How can side-by-side screens be utilised for peer feedback in a technology-enhanced classroom?

Methods

This study analysed students' digital experiences within technology-enhanced mathematics learning environments, focusing on formative assessment and feedback. For this research, a programme was piloted with an elementary school in the Czech Republic involving a cohort of nine Grade 8 (age 14) participants, during which they engaged in word-problem solving, utilising bar model virtual manipulatives apps provided within their tablet devices. The study leveraged classroom connectivity through digitally accessing and analysing students' mathematical thinking via screencast, i.e., recordings of their on-screen interactions while using bar model manipulatives. The students were given a series of tasks during which on-screen activities were recorded. The pilot programme, led by the researcher who also served as the teacher, consisted of two 3-hour sessions, one-week apart.

Data collection and analysis

Screencast was utilised as the primary means of collecting data, capturing visual information of students' digital interactions. Following data collection, the teacher-researcher commenced the analysis process by describing and interpreting visual cues within the datasets in the form of analytic memos. The following sections outline the data collection process, which generated samples for analysis. This is followed by sorting and selecting samples for class discussion utilising side-by-side screen. Figure 3 is a flowchart illustrating this process.

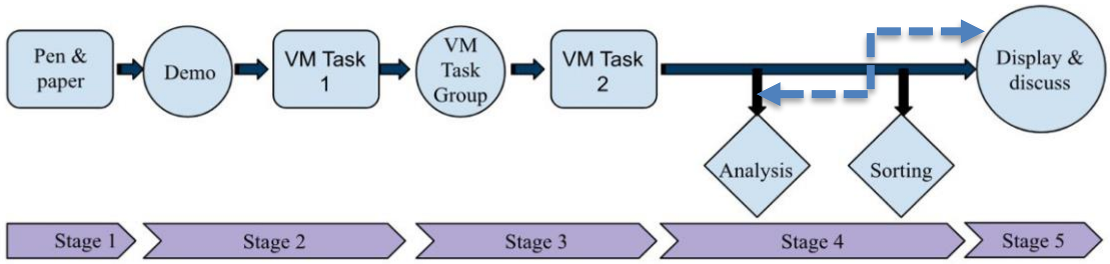


Figure 3: Instrumentation, data collection and analysis

Stage 1: Pen and paper task – Baseline establishment

Stage 2: Demo and Task 1 – Standardised demo and Task with virtual manipulatives

Stage 3: Group activities (VM Task Group) and Task 2 (VM Task 2) - Participants collaboratively solved word problems in groups of four or five. The aim was to assess the impact of group activities on students and determine if peer feedback is evident thereafter. Task 1, Group Activities and Task 2 were recorded for analysis, concluding Day 1 of the study.

Stage 4: Analysis and sorting - This involved systematically reviewing screen recordings of Task 1, Task 2 and Group Activities and writing analytic memos to capture emerging themes and pattern:

Initial Observations – The recordings were viewed multiple times for initial understanding.

Identifying Patterns – Patterns, themes and significant moments in the interactions were identified (see Coding). The identification process also included selecting among models with correct solution and incorrect solutions and sorted according to similarity metrics (Schmidt et al., 2013).

Coding – Codes were applied to segments of the screen output to represent moments of interactions. Interactions contributing to conceptual understanding: 1) use of the appropriate concept (Table 1), 2) accuracy in partitioning 3) alignments of parts 4) correct labelling, 5) application of operations

Memo Writing – Specific moments in the videos were annotated, documenting observations, interpretations and insights corresponding to coded segments.

Interpretation I – Coded segments and memos were analysed to identify broader themes related to research objectives. For RQ 1, data on interactions highlighting both strengths and weaknesses in the topic was collected and interpreted. Next steps were formulated to address weaknesses.

Stage 5: Display and discuss

Interpretation II – The selected screenshots served as instructional material and were displayed in class for side-by-side comparison and discussion. Observations were documented and redirected to Stage 4 for analysis (Figure 2). For RQ 2, reflective prompts (cognitive transfer and image comparison) guided reflections on students' engagement and comments. Cognitive transfer focused on identifying their ability to apply learned concepts in new situations. Image comparison related to their perception of accuracies/inconsistencies in screenshots and observed cross-referencing and editing of their own model construction.

Results

Interpretation I

This section addresses RQ1, examining interactions to identify strengths and weaknesses in the topic. We analyse Jakub's on-screen data to understand his grasp of the word problem in detail.

Question: 3/5 of the students in Grade 8 and 2/3 of the students in Grade 7 are girls. Both classes have the same number of girls. Grade 8 has 4 more boys than Grade 7. How many students are there in Grade 8?

For clarity, sequential screenshots (Transformations) of significant interactions are provided (Table 2b) with analytic notes for each moment. Table 2a outlines interpretations and reflections for these interactions. The documentation includes: 1) Appropriate Concept: Yes 2) Partitioning Accuracy: Yes 3) Alignment of Parts: No 4) Correct Labelling: No 5) Application of Operations: No.

The applied codes effectively pinpoint locations in the video, facilitating analysis. A thorough screencast examination reveals precise areas where Jakub struggled, with analytical notes explaining underlying reasons. Despite using the appropriate model, Jakub's solution was incorrect. The on-screen data not only clarifies specific points of struggle but also underscores his potential to engage with a concept he correctly selected but has not yet mastered in its application. Tailored remediation strategies can address Jakub's conceptual gaps, guiding future instructional strategies.

Interpretation II

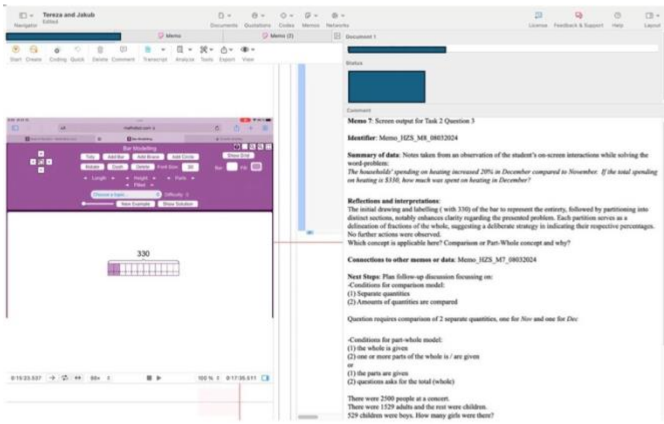
We use screenshots from Tereza's and Jakub's screen output (Figure 5 and Figure 6) to address the research objective that side-by-side screen effectively support peer feedback in a technology-enhanced classroom. The screenshots were selected using similarity metrics, aligning with the notion that selecting suitable similarity metrics is crucial for effective data analysis (Schmidt et al., 2013). This adds validity to their utilization in the study. Exchanges between teachers and students primarily reflected prompts and interactions related to the word problem. No individual attribution was recorded regarding which participant commented or posed each question during the study.

Table 2a: Reflections and interpretations of significant interactions

(a) The visual appears to be an effort to understand whether the number of girls aligns with the specified conditions outlined in the given word problem.
(b) The transformation shows that the appropriate bar model concept has been applied, i.e. the comparison model. This grasp of selecting the suitable application of the concept may potentially be attributed to the peer collaboration or peer learning experienced during the preceding group activity.
(c) The consistent use of the 'x' notation across all units led to confusion, resulted in him reaching an impasse.
(d) This visual representation highlights a discrepancy in logic, where the number of units drawn is unequal despite being depicted in the same size.

In addressing how side-by-side screens are utilised for peer feedback, we integrate Mayer's (2002) method of employing transfer tests to assess learner understanding. Questions and comments are famed and analysed from the perspective of cognitive transfer. Furthermore, we utilise similarity metrics to improve the effectiveness of data analysis (Schmidt et al., 2013) and to foreground inconsistencies between bar model constructions. When students identify these inconsistencies, peer feedback is engaged. Details of the analysis are provided in Table 3a and Table 3b.

Table 2b: Sequential screenshots of Jakub's data



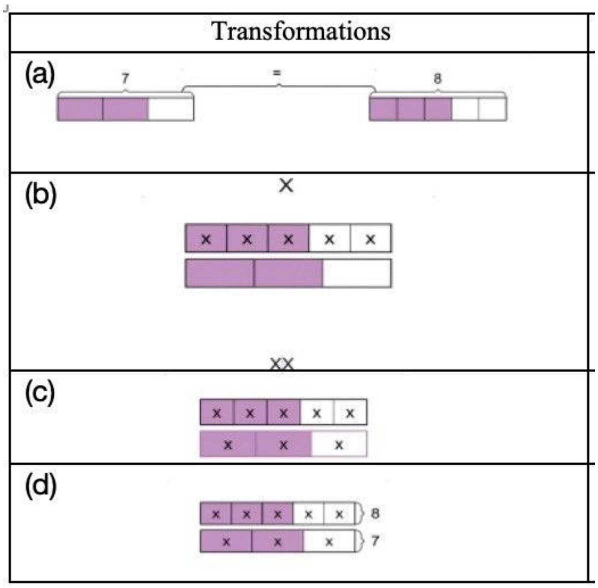
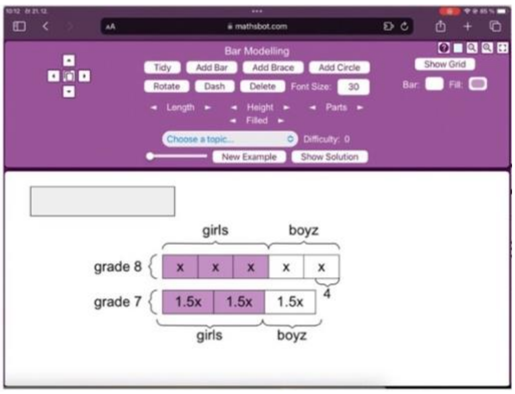


Figure 4: Jakub's data and analytic memo

This illustrates on-screen video data uploaded for analysis using qualitative research tool. The qualitative tool utilised enables thorough capturing of significant interactions and facilitates the creation of corresponding analytic memos



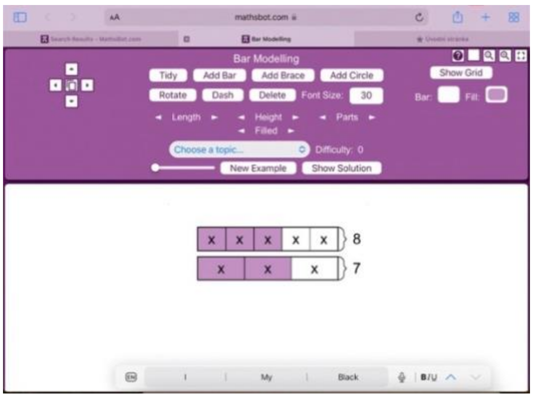


Figure 5: Tereza's screenshot **Figure 6: Jakub's screenshot**

Note. From MathsBot.com (Hall, 2013)

Table 3a: Analysis of students' contributions

Student's contributions:	Cognitive Transfer	Image Comparison
That (pointing to Jakub's model) is not correct.	By identifying errors or inconsistencies in peer's work, the student demonstrates understanding of the subject matter and capacity to critically analyse problem-solving approaches.	Further discussion about *why* he/ she thought it was incorrect triggers a number of debates among the students. Some were not agreeable there were any errors, others asserted that there were errors which prompted the comments that follow.
5x is not equal to 3x	This observation indicated the student's ability to apply his/ her understanding of mathematical concepts, specifically the relationship between quantities represented by variables (5x and 3x)	This comment shifted some students' attention to Tereza's model, with one pointing out the shaded regions indicated equal values. Someone brought up this might mean Tereza's model was wrong, since 5x was not equal to 3x. Several others disagreed.
One bar should be longer and one bar should be shorter	This statement reflected the student's observation of a discrepancy in the lengths of bars depicted in a diagram, showcasing his/ her ability to apply mathematical concepts to evaluate visual information presented in the diagram.	There were some initial confusion which model or bar was being referred to. This led to a deliberation that concluded that one class has 4 more boys than the other.. A number of students started turning to their devices to rework their constructions, others were seen cross-referencing each other's model

Table 3b: Analysis of teacher's contributions

Teacher's contributions:		
There are the same number of girls in both classes (restating the important information).	Highlighting key words to restate the problem prompting students to re-address the problem in a structured way. It brought focus to the problem-solving transfer, promoting meaningful learning.	Several students started to point out the same number of girls were represented by two perfectly aligned shaded parts on Tereza's models.
How do we show that 3/5 of the students in Grade 8 and 2/3 of the students in Grade 7 have the same value? Show this on your bar model.	This is a critical juncture in the learning process. The students' attention was drawn to a seemingly different but equivalent values. This is an opportunity for meaningful learning as they grapple with complex ideas and develop strategies	Some students noticed that Tereza's model fitted the description. A number of them cross-referenced their own models with Teresa's model and were making changes.
Everyone, draw (the bars) on your screen. How do you make the rectangles equal?	This was an opportunity to be seized upon as the bar model approach lends itself to effectively convey abstract concepts through its visual representation.	There were significantly more discussions, some revising their constructions, shading, re-labelling to emphasise this aspect of the information discussed

They were then asked to write down the algebraic equation. Eventually, they arrived at the equation: $5x - 4 = 4.5\,x$

They then solved for x by transposition and arrived at the answer: $x = 8$; Answer 40

Discussion and initial findings

Drijvers et al. (2016) resonate with our study's findings, indicating a shift in mathematics pedagogical practices from teacher-led demonstrations to student-led modelling and discussions facilitated by technology integration. The analysis outcome satisfies the research goals of using technology to help teachers effectively analyse students' conceptual gaps. Screencasts enabled the teacher to monitor students' math activity unobtrusively, facilitating authentic feedback to offer tailored support more effectively. The gradual capture of Jakub's screen output showcased his potential capacity in applying a concept, an aspect that would not be evident if viewed solely as a finished product, e.g., on pen and paper or static display. Further, the analysis suggests that side-by-side screen shows promise in supporting peer feedback. The images featuring similarity metrics sparked discussions among the students, fostering cognitive transfer as they exchange perspectives while examining the images they are comparing.

Conclusion

While the arguments presented strongly advocate for utilising digital means for assessing mathematics, in practice, implementing the features discussed remains challenging. In our research, we recognised a constraint regarding the efficiency of using screencasts for assessment, particularly in larger classroom settings. While our study involves a modest sample size of 9 students, the scalability of screencasts becomes challenging when applied to classrooms with more students. Managing, reviewing, and evaluating a large volume of screencasts poses logistical and practical hurdles for educators. This limitation highlights the need to explore alternative assessment strategies or technological solutions to ensure effective assessment practices in larger classroom environments. Another significant constraint is the lack of robust tools to cater for authentic mathematical practices such as sketching and scribbling within the app. Certainly, resorting to paper and pen can circumvent these constraints. However, this approach proves impractical when the objective is assessment through technology, as only a portion of the student's work would be visible within the assessment system. As a result, students may struggle to demonstrate their full problem-solving abilities, leading to a misalignment between their mathematical competence and the assessment's practice domain (Drijvers et al., 2016).

References

Clark-Wilson, A. (2010). Emergent pedagogies and the changing role of the teacher in the TI-Nspire Navigator-networked mathematics classroom. *ZDM–The International Journal of Mathematics Education*, *42*(7), 747–761.

Drijvers, P., Ball, L., Barzel, B., Heid, M. K., Cao, Y., & Maschietto, M. (2016). *Uses of Technology in Lower Secondary Mathematics Education: A Concise Topical Survey*. ICME-13 Topical Surveys, 1–34. http://doi.org/10.1007/978-3-319-33666-4

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Speech acts* (pp. 51–58). Brill. https://doi.org/10.1163/9789004368811_003

Hall, J. (2013). *MathsBot.com*. https://mathsbot.com/

Kho, T. H., Yeo, S. M., & Fan, L. (2014). Model method in Singapore primary mathematics textbooks. In K. Jones, C. Bokhove, G. Howson, & L. Fan (Eds.), *Proceedings of the International Conference on Mathematics Textbook Research and Development* (pp. 275–282). University of Southampton.

Kintsch, W., & Greeno, J. G. (1985). Understanding and solving arithmetic word problems. *Psychological Review*, *92*(1), 109–129. https://doi.org/10.1037/0033-295X.92.1.109

Mayer, R. E. (2002). Multimedia learning. *Psychology of Learning and Motivation*, *41*, 85–139. https://doi.org/10.1016/S0079-7421(02)80005-6

Mayer, R. E., & Wittrock, M. C. (1996). Problem-solving transfer. In D. Berliner and R. Calfee (Eds.), *Handbook of educational psychology* (pp. 45–61). Macmillan.

Nesher, P., Greeno, J. G., & Riley, M. S. (1982). The development of semantic categories for addition and subtraction. *Educational Studies in Mathematics*, *13*(4), 373–394. https://doi.org/10.1007/BF00366618

Rittle-Johnson, B., Star, J. R., & Durkin, K. (2017). The power of comparison in mathematics instruction: Experimental evidence from classrooms. In D. C. Geary, D. B. Berch, and K. M. Koepke (Eds.), *Acquisition of complex arithmetic skills and higher-order mathematics concepts* (pp. 273–295). Elsevier. https://doi.org/10.1016/B978-0-12-805086-6.00012-6

Schmidt, J., Gröller, M. E., & Bruckner, S. (2013). VAICo: Visual Analysis for Image Comparison. *IEEE Transactions on Visualization and Computer Graphics*, *19*(12), 2090–2099. https://doi.org/10.1109/TVCG.2013.213

Stacey, K., & Wiliam, D. (2012). Technology and assessment in mathematics. In M. A. Clements, A. J. Bishop, C. Keitel, J. Kilpatrick & F. K. S. Leung (Eds.), *Third international handbook of mathematics education* (pp. 721–751). Springer. https://doi.org/10.1007/978-1-4614-4684-2_23