Commentary for Teaching and Learning Secondary School Mathematics: Canadian Perspectives in an International Context, Springer Berlin.

On Measures of Measurement and Mismeasurement: A commentary on Planning and Assessment

Bharath Sriraman
Department of Mathematical Sciences
University of Montana

1. Introduction

There is an old Indian fable which goes along the following lines:

> A student who has memorized all night long for an examination the next day is seen walking very carefully along a cobble stoned path. On encountering a friend along the way, the student is asked why he is stepping so carefully. The student replies "I am afraid if I fall down, I will not remember all the information I have memorized all night long, it could get easily jumbled up."

This story has many morals and interpretations- and it is up to the reader, in this case the mathematics educator holding this book to decide what it really is. Before doing so, the six chapters in this section could very well serve as a compass or a guide to our interpretation because they delve deeply into issues surrounding planning and assessment in a mathematics classroom keeping the triage of the student, the lesson and the teacher in mind.

Several themes emerge from the six chapters which can be subsumed under the categories of homework and testing, alternative assessment, student understanding, and problem solving. Then there is also the issue of "scale"- which relates to the title of this commentary, i.e., what are we measuring through planning and assessment? an individual student, an entire classroom, entire schooling systems, or an entire country? The question of scale is important since it relates to recommendations made in each of the six chapters on what is possible at a particular micro-level. More often than not, recommendations for planning and assessment that work at a classroom level are often not implementable at a school level or even a district level. For now, we leave these larger considerations aside and focus on the contents of the six chapters.

2. The Six Chapters

Holm's chapter presents a necessary provocation- namely can't we plan a lesson with an end in mind? Is she suggesting we teach to the test? No! The end in this case is to further student understanding of concepts in mensuration. After a teacher implements an inquiry based lesson that relates to provincial standards in Ontario, Canada, Holm argues based on experience that great tasks simply remain "tasks" if no effort is made to consolidate and discuss what students have learned (and understood) from engaging with the task. Allusions are made to what this consolidation and discussion might look like- in this case to specifically target understanding of the relationship between volumes and surface areas of right cylinders. And even more specifically basic understanding of how the radii of circles form a building block for 3-dimensional circular objects, and using the proper units of measurement. The chapter closes with a specific extension task that conceivably "measures" all these aforementioned understandings. As an example of a lesson (task) that had a specific end in mind, it achieves the goals of measuring understanding by having students express their reasoning in words, and reveal their technical knowledge about surface areas and volumes of 3-D circular objects. However, what is not clear is ways in which a teacher might assign a score or a grade to this extension task, and whether or not is can be reformulated in a way that it can even be used as question that generated several related multiple choice item?

This question can be understood when seen through the lens of Correa's chapter which follows Holm's because it analyzes the relationship between ability and demonstrated work. In other words, can the numerous formative tasks that led to the summative extension task be used to evaluate the entire spectrum of ability in the demonstrated mathematical work? Arguably a multiple choice assessment offends the reader's sensibilities in some ways as it is tacitly assumed that open-ended items are better than multiple choice items. And this tacit assumption is further rooted in associating the former with "conceptual understanding" and the latter, viz., multiple-choice items with procedural understanding. We will come back to this at a later stage in the commentary in the context of situation the notion of "memory".

Correa presents two very interesting tasks and uses Kilpatrick, Swafford and Findell's (2001) mathematical proficiency model as an assessment tool. The first task, namely the distance medley relay task is used as an example where a coding scheme is revealed to the reader to gauge student understanding of proportional reasoning. I find this of interest as proportional reasoning is one area of mathematics education where many diagnostic instruments are available to pinpoint student errors (and deficiencies in understanding). Several well validated theories suggest that with age and experience students can tackle more complex proportional problems. In a nutshell, students move beyond the use of constant differences strategies to a building up strategy, and then to a multiplicative approach such as the unit-rate method. The development of these proportional reasoning concepts is not immediate, but rather a gradual process based on continual growth and progress coming from more progressive problems and strategies (Lamon & Lesh, 1992). Taking a Piagetian stance, this development happens at adolescent stages of maturity where additive approaches turn into multiplicative strategies that do not necessarily allow for the generalization to all cases. However, a formulation of a law concerning proportions becomes established that fits the scheme for solving proportions in its various situations (Lesh, Post and Behr, 1988). This may explain why a student will utilize different approaches for individual tasks (Inhelder and Piaget, 1958). The coding of the mathematical work of Rick on the distance medley task using Kilpatrick, Swafford and Findell's (2001) model reveals the interplay between procedural skill, conceptual understanding and elements of the problem solving cycle. While this is useful for the reader to appreciate both the student's thought processes and the teacher's ability to understand the student's thinking, the question of whether such an approach to assessment can be scaled up or not is left unaddressed. While it does suggest that Kilpatrick's mathematical proficiency model is a useful assessment tool, the constraints of time make it impractical for a teacher to be able to use this on an entire class homework set. This first caveat is important for the reader to note. Single cases are relevant as examples of good assessment practices, but become complicated when implementation is asked for on a larger measurement scale. The second caveat is that of inter-rater reliability- would teachers from the same school teaching a similar classroom, and using similar tasks, code the student work in the same way?

These caveats offer the perfect segue to delve into Rapke, Hall and Marynowski's chapter which scales things up for summative assessments that involve elements of mathematical problem solving. The question posed by these authors, is whether we can reframe testing to work within a classroom that emphasized problem solving. The mismatch between classroom assessment and classroom practice is one of the holy grails of the discipline of mathematics assessment. Without having technology as a confounding variable in this debate, it is well known that while activities in the classroom can emphasize student thinking, it is difficult to capture and thereby assess this on both formative and summative measurements (unless portfolios with fair rubrics are used). These authors make the bold suggestion of involving students in the development of tests- which can sound like heresy to the measurement orthodoxy. The second suggestion is to involve students as peer reviewers in assessing other responses on tests. In a sense what is being proposed harks back to the days of oral testing, where students were posed problems by the teacher in front of the entire classroom, and provided feedback on their solution in the spirit of constructive criticism. While the claim of these authors is that these strategies are, "re-castings of the traditional paper-and-pencil test [w]hich, teachers can use …[t]o promote deep approaches to learning and as a result help students to perform better on assessments", in a sense it is an appeal to a very old tradition of assessment going back to the Socratic method of *elenchus-proof-refutation*, which is found both in the Moore Method, as well as the Lakatosian heuristic (Sriraman & Dickman, 2017).

Suurtamm calls for us to move beyond the century old notion of aligning models of learning that emphasized facts and procedures easily measurable through end-of-unit tests, into acknowledging that students learn differently in this century, and are able to develop and convey their mathematical understandings in multi-modal ways. She writes that "current perspectives of mathematics teaching and learning value mathematical understanding through student engagement in problem solving and argumentation." The upshot is a need to focus on alternatives to traditional paper-pencil testing which can "provide multiple opportunities for students to show what they know and can do."

In a similar vein, the following chapter from McFeetors calls for customizing student learning and alternative approaches to a one-size-fits-all approach to homework and classroom routines. It is obvious that this chapter calls for differentiating instruction when possible and for being cognizant to multi-modal ways of representation available for lesson delivery, when students are able to work in customized classroom settings. Some limitations are also provided to such an approach.

Pai argues for expanding our current notions of the very idea of assessment, by calling on us to go beyond the idea of "measuring". An argument is put forth for a more holistic process that provides "formative, summative, and interpersonal functions, depending on the circumstances of the classroom." In other words, it calls for a boutique like attention to the needs of specific classroom milieus in the context of the material that has been learned. This chapter contains a nostalgic undertone of the teacher being able to mathematically journey through the different stages of sophistication of the student with mathematics. It also presumes a Lakatosian like ideal to a classroom where tangential questions and topics can become objects of curiosity and mathematical attention. Alas, as we all know- the day to day reality in a public school classroom is quite different from the lofty ideals on this chapter.

3. What is assessment, really?

So, we now return back to the fable and examine what it means in light of the six chapters that have addressed assessment and planning of mathematics lessons. In essence assessment is really about aligning instructional outcomes with (correct) responses on tests. It is also a reflection of the teacher, the teaching methods, and the learner and their learning methods. A class which performs poorly on a test can pose the following questions:

Is the test designed and written to assess (and thereby measure) what I have learned?
Or
Is the test designed and written to assess (and thereby measure) what I have not learned?

This is a fundamental question that is addressed to the writer of a test (or assessment). Are learners being assessed based on outcomes that can be fairly measured on a test. i.e., are they posed problems at various levels of "sophistication" (e.g., based on Bloom's taxonomy) on the mathematical material that has been covered? Is there a balance between items that call for recalling information (factual memory), and those that ask for applying and synthesizing information (deep memory)?

An adaptation of a game theoretic diagram illustrates what the basic constituents of an assessment look like. Assuming an assessment is construed as a "test" or "game" or "battle" between the teacher and student, then four basic win-loss combinations can be generated from the constituents. In our case it is important to define what the terms "win" and "loss" mean in this situation to establish our own norm of "Nash equilibrium", and the reader is urged to do so, based on the recommendations of the six chapters.

Teacher

Win Loss

Aligned to learning (A)	Factual Memory (fM)
Non-Alignment to learning (Na)	Deep Memory (dM)

Student

Loss Win

For instance an assessment that is aligned to learning and also calls for deep memory results in a win-win situation for the student, and the teacher (reflected in the measured score), as opposed to complete non-alignment to learning and reliance on factual memory which is NOT a win-win combination for neither student nor teacher (which will also reflect in the measured score). The former results in the equilibrium we desire and the latter in disequilibrium.

Scenario 1:	A-fM	Win-Loss
Scenario 2:	A-dM	Win-Win
Scenario 3:	Na-fM	Loss-Loss
Scenario 4:	Na-dM	Loss- Win

Scenarios 2 and 3 are in need of no explanation to the reader and can serve as tautological cases of the diagram (see explanation above). Scenarios 1 and 4 are Win-Loss and Loss-Win situations which are interesting to unpack as the reader would undoubtedly want to know who the winners and losers are in these situations.

Scenario 1 is a win for the teacher and a loss for the student because it results in a situation where students are taught behaviorally (think back to the push for mathematical reform in the 1970s in the U.S) and assessed exactly for each subskill or specific content covered. Think of a timed test with 20 problems that all call for the "skill" of addition. The scores might reflect well on the teacher as this form of learning does result in a "winning" score on a test, but is a loss for a student who has compartmentalized addition as the repeated invoking of factual memory- e.g., align the numbers from right to left in columns, carry over when the sum of the digits is over 9, etc.

Scenario 4 is a loss for the teacher and a win for the student because a situation has arisen where the teacher has failed to design an assessment that can provide a student an opportunity to demonstrate deep memory but the latter is sufficient to do fairly well on any kind of assessments. Having introduced a rather old-fashioned term "memory" into this commentary on chapters that extol the virtues of problem-solving based classrooms and alternative modes of assessment, an apology (defense) is needed. So this commentary concludes with this defense.

4. An Apology on Memory

> *The power of memory is great, very great, (my God). It is a vast and infinite profundity. Who has plumbed its bottom? This power is that of my mind and is a natural endowment, but I myself cannot grasp the totality of what I am (St. Augustine of Hippo,* Book X.8.15).

It is fashionable these days to confuse memory with memorization or rote learning. Mathematical memory is not this at all. To paraphrase St. Augustine of Hippo and much later Vadim Krutetskii, in the language of modern cognitive psychology, mathematical memory is that associated with an explicit memory system, thereby representing information which can be consciously recalled and explained. It is that which has resulted in deep structural insights into the nature of mathematics, at having experienced its fluidity in abstraction and generalization processes in the context of the mathematics being learned. It is a deeper kind of memory that allows for persons to retrieve relevant information when needed and be adept at using it or adapting it. For instance not knowing the derivative of a function can be overcome with the mathematical memory that contains its definition as the limit of a difference function evaluated at zero- and being able to apply this to derive canonical derivatives. It is not the knee-jerk recall of information that is simply memorized for a particular day or test, but one that is retrievable due to the cultivation of all previous mathematical structures within the memory.

The moral of the story based on the six chapters and this interpretation of what mathematical memory means, is that the student who memorized all night long was about to take a test for which his deeper mathematical memory would not be accessed, and his performance on the test would be based on the ability to recall different pieces of information (un-related or even dystopic from one another). We are of course assuming the student is taking a math test and is not about to give a recitation or performance requiring memorized information. In other words, the test would not measure his true knowledge, one that has accumulated over time, but that which is superficial and easily cast from the mind. The six chapters provide ideas, suggestions, tasks and reflections about the different assessments that have worked, and ways in which these target deeper learning objectives, as opposed to simple recall of information or process. Mass testing for deeper student understanding and problem solving is not even at its infancy yet, as we have learned from tests like PISA, item development and item coding is riddled with problems of cultural incongruence, relevance, and mismatch with local or even national curricular goals. In mathematics, it makes more sense to adopt alternative modes of assessment starting at the grass roots level as many of the chapters in this section advocate. In time, we can hope to achieve this on a much larger scale provided this is one of our goals as researchers.

5. References

Inhelder, B., & Piaget, J. (1958). *The Growth of Logical thinking from Childhood to Adoloscence.* New York: Basic Books.

Kilpatrick, J., Swafford, J., & Findell, B. (2001). The strands of mathematical proficiency. In J. Kilpatrick, J. Swafford, & B. Findell (Eds), *Adding it up [electronic resource]: Helping children learn mathematics* (pp. 115-155). Washington, DC: National Academy Press.

Lamon, S.J.,& Lesh, R. (1992). Interpreting responses to problems with several levels and types of correct answers. In R.Lesh & S.J.Lamon (Eds.), *Assessment of Authentic Performance in School Mathematics.*(pp.320-335). AAAS Press.

Lesh, R., Post, T., & Behr, M (1988). Proportional reasoning. In J. Hiebert & M. Behr (Eds.), *Number Concepts and Operations in the Middle Grades* (pp.93-118). National Council of Teachers of Mathematics, Inc., Reston, VA

Sriraman,B., & Dickman, B. (2017). Mathematical pathologies as pathways into creativity. *ZDM Mathematics Education*, 49, 137-145.