

To the International Association for the Evaluation of Educational Achievement (IEA) (**makers of TIMSS**) and to the US Common Core State Standards (National Governors Association (NGA Center) and the Council of Chief State School Officers (CCSSO)) (**users of TIMSS**)

IEA executive director Dirk Hastedt [... @]

NGA Center director R. Kirk Jonas [... @]

CCSSO executive director Carissa Moffat Miller [... @]

[anonimised June 27 2018]

May 31 2018

Concerning: When TIMSS may be invalid and when TIMSS is incomplete

Dear Mr Hastedt, Mr Jonas and Ms Moffat Miller,

Thank you for all your good work.

The following is intended to help improve TIMSS and the CCSS.

(1) Introduction

I am writing as an econometrician (Groningen 1982) and teacher of mathematics (Leiden 2008). It is fortunate that Mr Hastedt has a masters in mathematics and a Ph.D. in education research. Mr Jonas has a masters in political science and a Ph.D. in public policy. Ms Moffat Miller has a masters in sociology and a Ph.D. in education research. Let me advise you to consider reading Pierre van Hiele (1986), "*Structure and insight: A theory of mathematics education*", Academic Press. Van Hiele's theory is often misunderstood as relevant for geometry only, but Van Hiele presented it as a general theory for any subject, with geometry only as (aptly) *demonstration*.¹

The impetus to write you precisely today is that I completed the paper *Arithmetic with $H = -1$* ,² that highlights the issues below that are relevant for you more generally. The paper clarifies a.o. that negative numbers are important for number sense, understanding of the world, and pupil competence in mathematics. The paper deconstructs a curious confusion in traditional didactics of math. CCSS puts negative numbers very late in Grade 6, while Holland and the UK have them in Grade 7 (junior high). Would CCSS allow for change, and would TIMSS be able to record improvement, or, might reference to TIMSS make it harder to change at CCSS ?

(2) When TIMSS may be invalid

Arithmetic in elementary school prepares kids for later life. The highest aim is for algebra in highschool. If you do not master the classic algorithms in arithmetic (e.g. $1/7 + 1/8 = \dots$) then you will not be able to do proper algebra (e.g. $1/a + 1/b = \dots$). When there is a diagnosis for some pupils that they really cannot do better than with a calculator, and some trial and error, then these kids should get the education that serves them for later life, and they should not be subjected to later teaching of algebra that they would not be able to understand. I presume that these notions are so obvious that you will agree with them.

It has appeared in Holland that psychometricians (at Leiden University but also at Dutch CITO) can lack expertise in mathematics education (ME) and its research (MER). Some of them also explicitly say that they are not interested in ME & MER, and that it is sufficient for them that they would be competent on "testing" itself. They presume that the sums in K12 are so simple that they, with their expertise in Item Response Theory (IRT), would be competent to deal with those. This however is an unwarranted presumption. These psychometricians *count the correct outcomes of sums* and neglect *how* the outcome was found, by technique or calculator or trial and error. This kind of measuring runs against the very purpose of teaching mathematics: that it matters (for later stages in the curriculum) how an answer was found.

If this attitude and/or phenomenon also occurs at TIMSS, TIMSS would be invalid.

¹ <https://boycottholland.wordpress.com/2015/11/24/a-general-theory-of-knowledge/>

² <https://zenodo.org/record/1251687>

Around 1970 there was a “math war”³ w.r.t. what was called the “New Math”.⁴ Hans Freudenthal and later the Freudenthal Institute at Utrecht University advocated an approach to didactics of mathematics, baptised as “realistic mathematics education” (RME), and likely also known in the USA as “reform math”, that indeed allows for such trial and error and the use of the calculator. But trial and error and the use of the calculator are not the way to do algebra in highschool. Proper testing should expose RME as inferior for such purpose. The Freudenthal Institute should be abolished as unscientific, and motivated by ideology, like homeopathy.

However, there now is the “perfect storm”, that invalid testing by the psychometricians allows this teaching philosophy to survive. Currently, Holland features somewhat high in TIMSS because many kids use trial and error and the calculator, but universities set up remedial teaching since students are lacking in technique. Given this criticism, the KNAW / Royal Dutch Academy of Sciences in 2009 supported a report on arithmetic education.⁵ The committee was dominated by mathematicians without a background in empirics or K12 itself, and they relied upon psychometricians to provide for “the empirics”. Their report is invalid, for above reason.

The issue might be rephrased in this manner: whether the test criterion is the *correct outcome* for some types of sums or whether proper tests should be developed to measure student competence in technique. This rephrasing however would change the subject. The latter namely should be obvious, see the first paragraph in this section. Highschool teachers grade exams by checking on technique. The true problem is that psychometricians are lacking in competence on ME & MER and that they don't care about this.

This “perfect storm” may somewhat be rephrased in the following manner: That RME has succeeded in advocating that “sums with context” would provide an excuse for allowing the use of a calculator. Pierre van Hiele developed the theory that context would be important for the early levels of insight, but he also pointed to the need of abstraction for advancement. Basically, the methods (technique & algebra / trial and error / calculator) and the situations (context / no context) are independent of each other. It is only a confusion by RME to suggest that a context justifies the use of a calculator. It is true that education should also involve the use of the calculator, including when there are awkward numbers, but we should make sure that its use does not prevent the learning of technique required for later algebra (when you might avoid such awkward numbers). In this view on the issue, it are advocates of RME who have dominated with their confusion that context justifies the calculator, with the decision to include such sums in the tests, such that the psychometricians are not primarily responsible. The problem remains that the psychometricians support something that they do not understand, and that they are instrumental in tests that fail to expose an inferior didactics of mathematics.

There is a huge scandal here in Holland, with an experiment on kids in elementary school without the proper protocol on experimentation on human beings. A teaching method was introduced that allowed trial and error and the use of the calculator, and it survived in psychometric testing because those tests are invalid. Psychometricians neglect critique on their failure. This constitutes a breach of scientific integrity.⁶ My reason to write you (the reason, and not the impetus above) is that the Dutch system on Research Integrity apparently is failing as well. Leiden University regards this issue as a “scientific dispute”. KNAW, supervising itself on research integrity, allows this to happen. We live in a “knowledge society” but the safeguards on what “knowledge” is are underdeveloped. Research integrity however requires scientists to correct an error when it is clarified to them, and to first study a field before meddling with it.

(3) When TIMSS is incomplete

In ME & MER there is a distinction between (i) those who regard K12 “mathematics” as given (tradition) and who only look for better ways of teaching tradition, and (ii) those (me) who hold that mathematics would be clear in itself, so that problems in didactics are caused by the empirical fact that tradition is not clear but rather crooked. Improvement of didactics is another term for that mathematics education must be re-engineered.

Mathematics may have a reputation of being clear but the reality is that it has come about in a process of 5000 years of, indeed, trial and error. A critical mind will find a paradise for improving

³ https://en.wikipedia.org/wiki/Math_wars

⁴ https://en.wikipedia.org/wiki/New_Math

⁵ <https://www.knaw.nl/nl/actueel/publicaties/rekenonderwijs-op-de-basisschool>

⁶ <http://thomascool.eu/Papers/Math/CWI-Leiden/2016-09-30-Letter-to-CWI-anonimised.pdf>

didactics (re-engineering of mathematics education). A main criterion is that we must remain *practical*. What could be handled by the system of standards, teacher training, and so on? Would IEA and CCSS e.g. support a new ISO standard on the pronunciation of integers? ⁷

Perhaps the matter might be rephrased in this manner: What does TIMSS want to measure: whether kids learn the current crooked “mathematics” or whether they learn mathematics? However, this opens the can of worms who decides what math would be, with presumably a key role for mathematicians themselves who are blind to the empirics of didactics. However, it is proper to look at this way: When TIMSS encodes current tradition in concrete, so that countries with a low score try to copy the higher-up countries, and when TIMSS does not allow for improvement beyond this, then TIMSS indeed turns current tradition into a prison, on the assumption that the last 5000 years of history indeed would have created a perfect mathematics.

Let me invite you again to read the paper mentioned above on arithmetic with $H = -1$ on the treatment of negative numbers and rational numbers in elementary school. ⁸ This deconstructs some misconceptions amongst teachers and educators in elementary school. Apparently mathematicians have been so focused on their abstract theories since John Wallis in 1673 that they could not understand those misconceptions by teachers and educators in elementary school. (I am open for better explanations.) Subsequently, CCSS give standards that put fractions first and that postpone negative numbers to Grade 6. However, this curriculum appears to be based upon a confusion, and, apparently, deliberate efforts to eliminate contradictory evidence from discussion. Read the paper and wonder how this ever could have happened.

Well, the likely explanation is: Within this group (i) above, who take tradition as given, there are (i-a) mathematicians without a background in didactics and empirical research, who mistake mathematics education for mathematics itself, and (i-b) pedagogues without a background in mathematics (who might hold that a calculator does a good job for daily life). The latter implies that these pedagogues have no access to empirical methods that use techniques based in mathematics and statistics. There might be (i-c) mathematically competent people who employ empirical techniques, like those psychometricians, but still within said tradition. The latter means that they have not been trained on proper empirics, since proper *observation* of mathematics education leads to (ii) with its rejection of tradition. Researchers who are more competent on statistics may also have agendas of their own, like those psychometricians.

CCSS states with a reference to also TIMSS: “Fact: The mathematical progressions presented in the Common Core State Standards are coherent and based on evidence.” ⁹ But, this “evidence” is based upon input that has been generated by the very tradition that ought to be tested. You will only find what you put in there before. CCSS thus sets up a circular argument.

In answer to this, TIMSS and CCSS should better allow for a decent degree of experimentation at the frontier of innovation and re-engineering of mathematics education. TIMSS might say that countries are free to experiment and that TIMSS will duly record the results. I doubt whether such response would really fit IEA. IEA started with the assumption that countries could learn from each other, but the focus now shifts from *countries* to *factors* that drive success. One of the factors for future success will be proper experimentation at the frontier. Subsequent improvements in countries could be traced to the adoption of methods that appeared successful at the frontier. TIMSS might hold that it looks at *current* and not *future* success: “(...) to investigate how the participating countries are providing educational opportunities in mathematics and science to students, and the factors related to how students are using these opportunities”. However, also a *current* success *now* might be explained by investments in experiments in the past. Also, some of the opportunities provided to students are the very experiments at the frontier, that allow some students to escape from the shackles of tradition.

This warrants the statement: TIMSS would be incomplete.

TIMSS needs a frontier for redesign of mathematics education and its research.

⁷ <https://doi.org/10.5281/zenodo.774866>

⁸ <https://zenodo.org/record/1251687>

⁹ <http://www.corestandards.org/about-the-standards/myths-vs-facts/>

(4) Supplementary 1. Mathematics by computer

Let me alert you to an issue for State regulators w.r.t. computers and computer languages: “Everyone will be served by clear distinctions between (i) what is in the common domain for mathematics and education of mathematics (the language) and (ii) what would be subject to private property laws (programs in that language, interpreters and compilers for the language) (though such could also be placed into the common domain).”¹⁰

(5) Supplementary 2. Statistics education and government classes

Let me alert you to that something goes horribly wrong¹¹ in apparently many countries in the world, with both (i) education in statistics and (ii) government classes on democracy w.r.t. the application of statistics in the political science of electoral systems and votes and seats. The link to IEA / TIMSS and CCSS is: this branch of political science mistakes *words* (the language program) for proper *observation* (mathematics program on measurement and data).

On these latter two alerts: I can only observe that it is relevant for education but I have not elaborated how it could be relevant for IEA and NGA and CCSSO. I assume that you would spot the relevance when you would consider the given links.

(6) Warning on Holland

I started writing about didactics of mathematics in 2008, and was quite surprised by both my findings and reactions by others to those. My background is econometrics, with is a generalisation and not specialisation, and perhaps this helped to maintain common sense and roots in science and observation. Let me warn about the pervasive influence of ideology. Mathematics concerns thinking and people have a close attachment to how they think, or are trained to think, and mathematics even comes with some claim of being a better way of thinking. Regrettably I must report that Holland apparently since about 2000 has a “math war” to such extent that you cannot trust anyone from Holland except me on the issues in this letter. Each person might be correct for 99% but then there is always this 1% that subverts it, while it takes scrutiny to pinpoint and deconstruct that 1%. You can check how my analysis since 2008 has been treated in Holland: a wonderful opportunity to improve mathematics education and its didactics is maltreated by RME and traditionalist mathematics and psychometric “testing” and whatever 1% of bottleneck confusion. IEA has its headquarters in Amsterdam, and some of its staff may have heard about this “math war”. This letter should provide you with a reality check.

(7) Closing

I have tried to alert you to the notion of “re-engineering of mathematics education”. Let me invite you to read said paper on the negative and rational numbers and also my book *Elegance with Substance* (PDF online),¹² and also *A child wants nice and no mean numbers* (PDF online),¹³ (with now above [amendment] on the pronunciation of numbers) and allow me to suggest that you invite the participants in TIMSS and CCSS-Math to also give the issue a chance. Best is that you set up workshops on the many examples given, and that you invite people to write reviews on those examples and the general issue. When you see emerge some critical mass then you could proceed from there. I hope that this works for you.

Sincerely yours,

Thomas Cool / Thomas Colignatus
Econometrician (Groningen 1982) and teacher of mathematics (Leiden 2008)
<https://zenodo.org/communities/re-engineering-math-ed>
<http://econpapers.repec.org/RAS/pco170.htm>
[.... Scheveningen, Holland ...]
<http://thomascool.eu>

¹⁰ <https://boycottholland.wordpress.com/2018/05/30/terminology-of-mathematics-by-computer/>

¹¹ <https://zenodo.org/record/1228640>

¹² <https://zenodo.org/record/291974>

¹³ <https://zenodo.org/record/291979>