Journal of Science Education, 4(2), 73-76, 2003.
Schmoker, M. Results now. How we can achieve unprecedented improvements in teaching and learning. Alexandria, VA: Association for Supervision and Curriculum Development. 2006.
Towns, M. H. Guide to developing high-quality, reliable, and valid multiple-choice assessments. *Journal of Chemical Education*, 91(9), 1426-1431, 2014.
Vernon, D., & Blake, R. Does problem-based learning work? A Meta-analysis of evaluative research. *Academic Medicine*, 68(7), 550-563, 1993.
Vyas, R., & Supe, A. Multiple choice questions: A literature review on the optimal number of options. *National Medical Journal of India*. (3):130-3, 2008.
Zimmerman, B. B., Sudweeks, R. R., Shelley, M.F., & Wood, B. How to prepare better tests: Guidelines for university faculty. Provo, UT: Brigham Young University Testing Services. 1990.

Received 22-01-2015 /Approved 30-04-2016

A rethinking of assessment practice: an experience with a stage test

El replanteamiento de la práctica de la evaluación: una experiencia con una prueba en etapas

André Luis Trevisan, Regina Luzia Corio de Buriasco

Department of Mathematics , Federal Technological University of Paraná, Londrina, PR, Departament of Mathematics, State University of Londrina, PR, Brazil, andrelt@utfpr.edu.br, reginaburiasco@gmail.com.

Abstract

This article presents a study based on an experience involving a different assessment instrument (a stage test) in a second-year high school class. The work involving this use was developed over a semester, in order to take the written test as a learning task, and was developed under the perspective of assessment as a formative instrument present in the educational process both as a teaching and students' learning processes diagnostic tool and as a way to investigate pedagogical practice. Reflections originating from the written production of students in one of the questions of the test as well as a critical analysis of the instrument itself and the attitudes as a teacher are presented. The approach adopted is qualitative/ interpretative in the light of Content Analysis.

Keywords: *school learning assessment, stage test, written production analysis.*

Resumen

Este artículo presenta un estudio sobre una experiencia con un instrumento diferenciado de evaluación (una prueba en etapas) en una clase de la escuela secundaria de segundo año. El trabajo relacionado con este uso se desarrolló a lo largo del primer semestre, con el fin de tomar la prueba como una tarea de aprendizaje, y bajo la perspectiva de la evaluación como un instrumento de formación presente en el proceso educativo, tanto para el proceso de enseñanza y aprendizaje de los alumnos, como para herramienta de diagnóstico y una manera de investigar las prácticas pedagógicas. Las reflexiones se originaron a partir de la producción escrita de los estudiantes en una de las preguntas de la prueba, así como un análisis crítico del instrumento en sí y las actitudes que como docente se presentan. El enfoque que se adopte es cualitativa / interpretativa a luz del análisis de contenido.

Palabras clave: *evaluación del aprendizaje escolar, prueba en etapas; análisis de la producción escrita.*

INTRODUCTION

In this article, we present excerpts of a study about the use of a six-stage test in Math lessons. Although it is an investigation in the teaching of Mathematics, it can be applied to education of the sciences, and it brings contributions using modern active methods into teaching and assessment practices.

In opposition to the concept of Math as "an erudite discipline whose teaching is provided to all ages", Freudenthal (1979, p. 318) understands Math as a natural and social activity, the evolution of which follows that of the individual and meets the needs of an expanding world.

For Freudenthal (1979), Math is a both natural and social human activity, just like the speaking, drawing and writing. It is included among the first known cognitive activities to be taught. However, it evolved and changed, including its Philosophy and method, under the influence of social changes.

Under the Realistic Mathematics Education, a movement that gained power in Holland in the late 1950s and had as its forefather the mathematician Hans Freudenthal, students must be seen as active participants in the educational process. Situations that demand math organization should be proposed to the students. From these situations math concepts will arise as well as opportunities to reinvent math through a process of reality *mathematization* (De Lange (1987, 1999, 2003), Freudenthal (1979), Gravemeijer (2008), Gravemeijer & Terwel (2000), Van Den-Heuvel Painhuizen (1996)).

An assessment consistent with the RME must as an educational activity be formative and treat Math as a human activity, focused on meaningful activities. It should take into account that, during their development process, students go through several levels of mathematization and "create" their own math, offering them (imaginable) realistic contexts. Several other authors (Buriasco (2000); De Lange (1987, 1999); Esteban (2001, 2009); Hadji (1994); Van Den Heuvel-Panhuizen (1996); Viola dos Santos, Buriasco & Ciani (2008)) have referred to assessment as a formative instrument in the education process, both as a means to diagnose teaching and learning processes and as an instrument of pedagogical practice investigation. Along these lines, analyses involving the written production of students developed at GEPEMA (Group of Studies and Research in Mathematics Education and Assessment (http://www.uel.br/grupo-estudo/gepema/) are carried out under the perspective of assessment as an investigation practice and learning opportunity.

Besides being aware of this, we, as teachers, have often wondered how we could "operationalize" an assessment perspective to help us interpret, include, regulate and mediate teaching and learning processes. Barlow (2006, p.165) gives us a hint: it is necessary to kill the imaginary evaluator, by questioning and rejecting myths and rites and false appearances as well as to know how to revive it, by preserving or recreating "that which carries meaning and is rich in potential efficacy".

In an attempt to "kill" my own imaginary evaluator when I started my Doctoral studies I found myself in the position of a teacher/ researcher trying to reconceptualize assessment practice. So, at that time, my idea of assessment came down to "taking tests", thus changing assessment practice would imply modifying the instrument. Accordingly, I decided to experiment with a different written test format with my classes, inspired by some studies that used two-stage tests. It included a written test accomplished in two phases: in the classroom and with limited time (first phase) and generally at home, with more time (second phase). According to De Lange (1987), the two-stage test gives students the opportunity to reflect upon their work: after being taken at school for the first time, the test is corrected and commented on by the teacher and then returned to the students for additional work.

Menino and Santos (2004) and Santos (2004) report experiences about the use of the two stage test as an assessment tool applied to different levels of education. For Menino and Santos (2004), the second stage is based on "runs" offered by the teacher at the end of the first phase.The student performs the second stage in a period agreed to beforehand, working especially with open questions. According to Santos (2004), the second stage must include test questions of an open nature, such as exploration and research tasks. In these questions that allow for any degree of development of the response in the first stage, the student has the possibility to deepen their understanding in the second stage. In addition, this instrument constitutes a new moment of learning and contributes to the fundamental student self-assessment due to the comments that the teacher writes. In making these comments, "the teacher needs to be (careful and) extremely self-critical with his or her evaluation comments" (Santos, 2004, p.6).

More recently, the work carried out Gepema (Pires, 2013; Trevisan, 2013; Mendes, 2014) have presented proposals to split the test into more phases, which has been called the *stage test*. This work relates an experience with a stage test, and was attempted in order to make the written test be a learning task and was developed with the perspective of assessment as a formative instrument present in the educational process both as a teaching tool and as a diagnostic tool for the students' learning processes.In addition, it is a way to investigate pedagogical practice. We present some reflections originating from the written production of students in one of the questions of the test as well as a critical analysis of the instrument itself and of the attitudes of the teacher.

METHOD

The assessment tool was used with a group of second-year high school students (age: 15-17 years) from a public institution in Brazil, where the first author works as a Math teacher. The test comprised 28 questions (taken from textbooks and tests used in previous years), on the content selected for the first semester, and it was organized to be taken in six phases rather than two, all in the classroom. The option for the number "six" was based on an analogy to a common assessment model used in Basic Education classrooms, which includes two bi-monthly tests and a "retake" test. The difference is that, instead of being given six isolated tests during the semester, questions were compiled in a single notebook to be answered during regular school hours and on pre-established dates.

Students could choose the questions they wanted to answer in each phase (considering that a single grade would be given at the end of the semester), and solutions to the problems could be altered in the following phases, whenever needed. Thus, as the semester went on and the contents were explained in the classroom, students were thought to be able to solve the problems as they received the test. At the end of the third phase I wrote a question on the side of each item of the test regardless of whether the answer was right or wrong, intending to motivate the students to reflect on what they had done up to that point.

The study involving the written production of 25 students was done in the light of Content Analysis (Bardin, 1977), the corpus comprising the set of solutions of each test question, from the second phase onwards. The constitution of this corpus complies with the selection rules noted by Bardin (1977): all documents used in the analysis included different solutions (representation) from the same test (homogeneity), resolved by different students, all from the same class (completeness), and were adequate as a source of information for the research in question (pertinence).

As investigators, we were interested in finding signs in the students' written production that would allow us to understand whether the intervention adopted was closer to correcting (which would allow the student to recognize and correct their own mistakes) and regulating (enabling the student to recognize his solution strategies) proposal. Such characteristic are inherent to a formative assessment (Hadji, 1994; Barlow, 2006). In order to codify and categorize the students' written production, we used an identification code formed by the letter P (test) and followed by an arbitrary number sequence with two digits (01, 02, ..., 25), organizing the groups into G1, G2, and so forth, using the procedure adopted by the student to solve the question as a cut off point.

ANALYSIS OF WRITTEN ANSWERS TO A TEST QUESTION

For this paper, we present a study involving one of the test questions. Although it is a specific example, it illustrates the format and characteristic of test questions (which, possibly, teachers usually choose when preparing written tests). Here is the question: *If an arc measures 3780 degrees, which is its first positive determination?*

We understand as the first determination of an arc the smallest arc congruous to it (i.e. with the same image in the trigonometric cycle). Usually, if the arc is positive, the measure of arc is divided by 360 degrees, and the remainder of this division is taken as the first determination; the quotient indicates the number of complete turns in the trigonometric cycle.

Analysis of the written production showed that all students had used the strategy (highly "discussed" in class) of dividing 3780 by 360 and taking the remainder of this division as the answer. However, we wondered whether they understood the algorithm and could interpret the meaning of the quotient and the remainder of this division. Hence, we added the following question: *What does the result of this division represent?*.

Groups organized according to the answers given by the students, up to and after the third test stage are shown in Table 1, as well the question that I wrote on the side of each answer, intending to motivate them to reflect on what they had done up to that point.

The group G1 represents the production of students who used the "vertical form" as the procedure to calculate the division of 3780 by 360, getting the quotient 10 and the remainder 180. Next, they recognize the 180-degree arc as the first positive determination for the 3780 degrees arc. However, when questioned on the meaning of quotient 10, only two of them said it refers to the number of complete turns in a circumference that corresponds to a 3780 degrees arc. Two students answered that it is the first positive determination of the arc. Figure 1 shows a P2 solution.

Table 1 – Groups organized according to the written production.

Group		**Analysis**		
	Test	Up to 3rd stage	Question	After 3rd stage
G1	P1, P2, P6, P10, P13, P14, P17, P19, P20	Shows the "vertical form" of the division (3780 divided by 360), getting 10 as the quotient and 180 as the rest. Gives 180 degrees as the answer.	What does the result of this division represent?	P19 recalculates the algorithm, getting the same results, showing zero as the answer. P1 and P2 answered that the result of the division represents the number of complete turns. P6 and P17 answered that the result represents the first positive determination of the arc. The others maintained their solutions.
G2	P9	Shows the "vertical form" of the division (3780 divided by 360), getting 10 as the quotient and 180 as the rest. Gives "10" as the answer.	What does the result of this division represent?	Maintains the solution.
G3	P3	Shows the "vertical form" of the division (3780 divided by 360), getting 10 as the quotient and 180 as the rest. No answer is provided.	What does the result of this division represent?	Answers that the result represents the first positive determination of the arc.
G4	P11, P12, P15, P16	Shows 10 as the result for 3780 divided by 360. Next, multiplies 3780 by 360 and subtracts the result from 3780, getting 180 as the answer. Gives 180 as the answer.	What does the result of this division represent?	Maintains the solution. P11 and P15 answer that the result of the division represents the number of complete turns.
G5	P22, P23, P25	Shows 10.5 as the result for 3780 divided by 360. Next, multiply 3780 by 10 and subtract the result from 3780, getting 180. Gives180 degree as the answer.	What does the result of this division represent?	Maintains the solution. .
G6	P18	Shows 10.5 as the result for 3780 divided by 360. Gives as the answer: "No,it's neutral."	What does a "neutral" arc mean	Answers the question by saying that " it is in the 0 degrees, 90 degrees, 180 degrees, 270 degrees,360 degrees points".
G7	P24	Shows the algorithm for the 3780 divided by 360 by "cutting" the zeros and getting 10 as the quotient and 18 as the rest. Provides 18 degrees as the answer.	What does this "can-cellation" mean?	Maintains the solution.

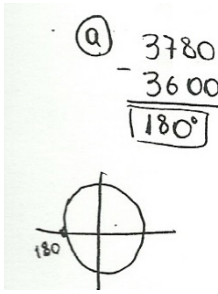

Figure 1 – solution in P2.

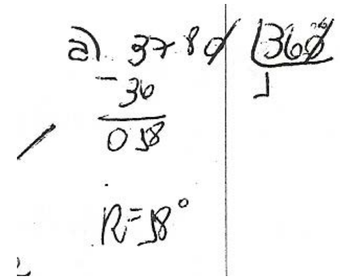

Figure 4 – solution in P15.

Both G2 and G3 are formed by only one student each, both using the same G1 procedure. In the case of G2, however, the student provides 10 as the answer, and, even after being questioned on the meaning of the result of the division, he maintains the original answer. In the case of G3, the student may have thought that presenting the algorithms of the accomplished operations would answer the question. After being questioned, the student answers that the "result" of the division represents the first positive determination of the arc, though it is unclear whether the result refers to the quotient or the remainder of the division.

G4 and G5, on the other hand, differ from each other since the first shows 10 as the quotient for the division 3780 by 360 while the second shows 10.5. Both adopt the strategy of "recovering" the rest of the division by multiplying 360 by 10 and subtracting this result from 3780. In both cases, the 180 - degrees measure is the answer to the question. When questioned on the meaning, both two students say they represent the number of complete turns of the 3780 degrees arc. Figure 2 shows the solution in P11.

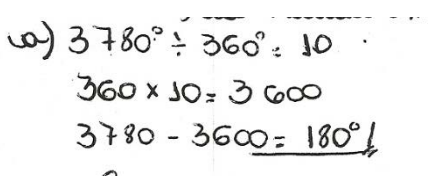

Figure 2 – solution in P11.

In G6, we find the production of only one student who, after having divided 3780 by 360 obtained the quotient 10.5, concluded that the first positive determination is neutral. When questioned on the meaning of "neutral", the student informs that they are in the "0 degrees, 90 degrees, 180 degrees, 270 degrees and 360 degrees", referring to arcs whose extremities lay on some of the Orthogonal Cartesian axes. Figure 3 shows this student's solution.

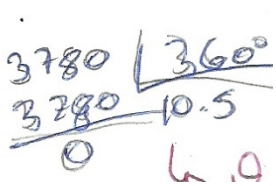

Figure 3 – solution in P18.

Finally, G7 corresponds to a production in which, by using the "vertical form" to resolve the division 3780 by 360, the student realizes a cancellation by "cutting" the zero from the units order in both the dividend and divisor. By mistake, the student presents 1 as the quotient for the division 378 by 36, and 18 as the rest, forgetting that the cancellation of the zero would imply in multiplying the remainder by 10 and takes the 18 degrees value as the answer. Figure 4 shows this student's solution, which was maintained even after the questioning.

RETHINKING ASSESSMENT TASKS

My expectation as a teacher was that, after having given students the possibility to change their solutions in stages of the test, they would really do it effectively. I also expected that my questionings would contribute to their improvement or changes. The analysis of the written production, however, did not point in this direction. Yet, I noticed a series of "flaws", both in the preparation and implementation of the instrument, as I reviewed the literature to search for a reference that would support my practice. Although I had modified the instrument, the questions and, more importantly, my own attitude towards it continued being traditional.

In the light of the constituted theoretical background, we proposed to reevaluate the questions of the test. In our opinion, none of the questions of the test offers students the possibility to mathematize situations. The solution of all of them involved the use of routine procedures which prioritized mechanisms instead of math concepts and reflected my excessive preoccupation with "covering the program"; the objectives that I intended to meet. I did not view clearly the intended learning outcomes of the different topics in the description of the discipline. The topics were introduced to the students simply because they were there, "frozen" in the test questions.

According to Oliveira e Pacheco (2008), this content has become an automatic part of the schooling process. It is neither questioned by us, who are used to seeing the topics where they are, nor are the objectives that we want to reach when working with the students very clear. So, how can we assess them? We end up repeating assessment schemes despite knowing that these classic mechanisms are often inadequate to the innovation we try to incorporate in our daily work.

The test posed and discussed in this article illustrates that fact. The problem solution involved the application of standard algorithms to obtain the first positive determination of an arc outside the first turn; most students used the same strategy in their solutions: divide the arc measure by 360 degrees and take the remainder of the division as the answer. However, when questioned on the meaning of the quotient of this division, many were unable to interpret it.

Below we suggest a reformulation, based in the Realistic Mathematics Education ideas, within a realistic context (a spinning wheel representing the trigonometric cycle) so that the problem demands more than just remembering a fact or reproducing a technique and turns into something attractive and stimulating[1]:

In a contest, there is a circle divided into six geometrically equal sectors. Around the center of the circle runs a pointer that, after being spun, indicates the amount in dollars each player has to pay or receive (Figure 5).

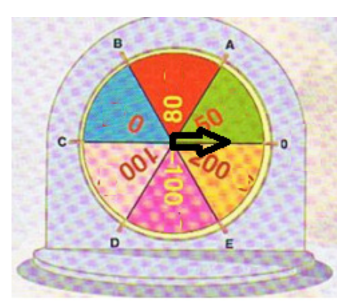

Figure 5 – Proposal for the reformulation of the question.

1 Proposal adapted from http://www.esaas.com/grupos/matematica/estagios/exerciciossite/FichasTrabalho/FichaTrabalho2_MatA.pdf.

Game rules:

- The pointer starts to spin at 0, in a positive or negative direction.
- The move is valid only when the pointer spins at least two complete turns; otherwise, the move is repeated.
- Whenever, the pointer stops at the division of two sectors, the move is repeated.

a) John made a move. In each of the following cases, determine the result of the move, given the amplitude of the arc described by the pointer, justifying your answer:

i) 3780 degrees

ii) – 1043 degrees

iii) $35\pi/6$ *rad*

b) Mary made only one move in the positive direction and won 80 dollars. Write possible amplitudes for the arc described by the pointer, knowing that it turned less than six times.

According to the assumptions of RME, problems should be presented to students in contexts that are realistic. The teacher should identify situations that can be used to explore informal strategies of students and thus to serve as starting points for the reinvention process. Freudenthal (1979) suggests looking for applications that can meet phenomena to be organized by concepts, procedures and mathematical tools. We recognize that the reworded question contained the presence of this rich context of meaning, conducive to mathematization, in which students are able to imagine something and make use of their own experiences and knowledge. Moreover, in its reformulated version, the question has characteristics pointed out by Van Den Heuvel-Painhuizen (1996) as desirable in evaluating problems:

- in addition to making situations "recognizable" and easily imaginable, the chosen context can provide a pleasant and inviting environment, increasing accessibility to the problem;
- compared to the previous formulation (a task of "numbers"), the question now offers more opportunity for students to demonstrate their skills;
- the chosen context can incite the students to formulate strategies, expanding the ability to solve a problem by their own means and mathematical insights.

FINAL CONSIDERATIONS

Along a research path filled with concern, doubts, questionings, disappointments, as well as many moments of learning, I realized that the assessment act had other meanings besides the one which I was used to.

Thinking not only about an assessment's certifying function but also about its guiding and regulating perspectives demanded going beyond "verifying" whether the students had learned the content and finding alternatives in order to guide them constantly in their learning processes. Rethinking assessment under an investigation practice and learning opportunity perspectives depended on a change in the concept of Math from a ready and self-contained science to a more dynamic Math that reflects the organization processes of reality.

The use of a stage test challenged the assessment model to which both I, the teacher, and the students, had already been accustomed. Firstly, the test was already familiar. As they felt uncomfortable with it, since they did not know how to study for a test that they already knew, I ended up planning my lessons to "prepare them" to take the test. The possibility of reviewing the questions as many times as needed, a genuine opportunity to provide feedback in a formative assessment context, proved to be highly limited at that moment.

The "rereading" of the test questions, carried out through the analysis of the students' written answers, showed that the questions written on the side of their solutions were highly limited and contributed very little to help them recognize and correct their errors. Improving of this "art of making questions" is a constant exercise in the practice of a teacher who seeks to turn assessment into an investigation practice as well as a learning opportunity for the students.

BIBLIOGRAPHY

L. Bardin. *Análise de conteúdo*. 3ed. Lisboa: Edições 70, 1970.

M. Barlow. *Avaliação escolar*: mitos e realidades. Porto Alegre: Artmed, 2006.

R. C. Bogdan; S. K. Biklen. *Investigação qualitativa em educação*. Portugal: Ed. Porto, 1994.

R. L. C. de Buriasco. Algumas considerações sobre avaliação educacional. *Estudos em Avaliação Educacional*, v. 22, p. 155-177, 2000.

J. De Lange, J. *Mathematics, Insight and Meaning*. Utrecht: OW &OC, 1997.

J. De Lange. *Framework for classroom assessment in mathematics*. Utrecht: Freudenthal Institute and National Center for Improving Student Learning and Achievement in Mathematics and Science, 1999.

J. De Lange. Mathematics for Literacy. In: Madison, B. L., & Steen, L. A. (Eds). *Quantitative Literacy*: Why Numeracy Matters for Schools and Colleges. Princeton, New Jersey: National Council on Education and the Disciplines, 2003, p. 75 – 89.

M. T. Esteban. Avaliar: ato tecido pelas imprecisões do cotidiano. In: Garcia, R. L. (Org.). *Novos olhares sobre a alfabetização*. São Paulo: Cortez, 2001, p. 175 – 192.

M. T. Esteban. Avaliação e fracasso escolar: questões para debate sobre a democratização da escola. *Revista Lusófona de Educação*. Lisboa, v.3, 2009, p. 123-144.

H. Freudenthal. Matemática nova ou educação nova? *Perspectivas*. Portugal, v.9, Issue 3, p. 317-328, 1979.

K. Gravemeijer. RME theory and mathematics education. In: Tirosh, D., & Wood, T. (Eds). *Internacional handbook of mathematics education*: Knowledge and beliefs in mathematics teaching and teaching development. Rotterdam, The Netherlands: Sense Publishers, 2008, p. 283 – 302.

K. Gravemeijer; J. Terwel, J. Hans Freudenthal: a mathematician on didactics and curriculum theory. *Journal of Curriculum Studies*. Basingstoke, v. 32, Issue 6, p. 777 – 796, 2000.

C. Hadji. *A avaliação*: regras do jogo. 4.ed. Portugal: Porto, 1994.

H. Menino; L. Santos. Instrumentos de avaliação das aprendizagens em matemática. O uso do relatório escrito, do teste em duas fases e do portefólio no 2º ciclo do ensino básico. In: SEMINÁRIO DE INVESTIGAÇÃO EM EDUCAÇÃO MATEMÁTICA. 15o. Lisboa, 2004. *Actas*... Lisboa: APM, 2004. p. 271-291.

I. B. Oliveira; D. C. Pacheco, D. C. Avaliação e currículo no cotidiano escolar. In: Esteban, M. T. *Escola, currículo e avaliação*. 3.ed. São Paulo: Cortez, 2008, p.119 – 136.

L. Santos. *As actuais orientações curriculares no ensino e aprendizagem da Matemática*: a avaliação e os seus desafios. 2004.

M. Van Dehn Heuvel-Panhuizen. *Assessment and Realistic Mathematics Education*. Utrecht: CD-ß Press/Freudenthal Institute, Utrecht University, 1996.

J. R. Viola dos Santos; R. L. C. de Buriasco; A. B. Ciani. A Avaliação como Prática de Investigação e Análise da Produção Escrita em Matemática. *Revista de Educação*, v. 25, p. 35 – 45, 2008.

Received 06-09-2015 /Approved 30-04-2016