Journal Plus Education, ISSN: 1842-077X, E-ISSN (online) 2068-1151 Vol XII (2015), No. 2. pp. 167 - 176

WHAT GRADE SHOULD UNDERGRADUATE EDUCATION HAVE?

Claudiu Schwartz
Aurel Vlaicu University of Arad
claudiu.schwartz@gmail.com

Abstract: *The article addresses the problematic of educational assessment in Romania, and tries to illustrate through a speculative study the flaws of the current state of assessment methodology in our country. By making a small study with the statistical accuracy similar to the statistical accuracy of current assessment methods used by education, certain aspects about the accuracy of it will be subtly revealed, and questions remaining unanswered by our study are similar to questions remaining unanswered by our formal assessment system.*

Key words: *Standardized assessment, methods, skills,content acquisition*

Assessment is a topic widely debated upon in the United States where standardized tests are used on average 15 times a year between grades 3 and 8. Romania however does not seem to abuse the scores of standardized tests. Aside from changing variations on "capacity" exam, there is the Baccalaureate. Thus, standardized assessment in Romania is performed invariably – depending on the number of education reforms – from two up to six times during undergraduate studies. The rest of the assessment is based on subjective interpretation of the student's performances by the teacher in class.

Although the Romanian Law reads "*The assessment focuses on skills, provides real feedback and underpins pupils' individual learning plans.*" (Law of Education). It seems provides rather interpretable terms on assessment. Not just that it doesn't define "feed-back" in this context, but also lends a property to the term – "real" – a term which is again interpretable. Thus " real feedback" doesn't mean precisely something clear but it's rather left at one's judgment of how the two terms are defined, and this can be seen in practice: for instance, a grade of 7 on chemistry provides as much feedback as much it is perceived so by the assessed, and to what extent this feedback is also real or not depends on the rater. Thus, if the student considers the grade good bad – according to his/hers expectations – but the teacher considers the grade differently, the feedback misses its purpose altogether. Everything is subject to perspective, and the "real feedback" makes no exception, being perceived by the rater with

a certain way, perceived by the student differently and perceived completely different by the parents. There are three points of view on a single note which is supposed to provide real feedback. Some questions remain: who is the feedback intended for? And what is the purpose of it?

The Law of Education regulates assessment through "*[...] a unique bank of assessment instruments, having indicative function, in order to help the teachers rate in class*". Again, the Law indicates a base of instruments which work as guidelines meant to help raters. Obviously, it means that assessment results will also be indicative, if anything. Regulating assessment using guidelines which can only produce indicative results should dismiss high-precision rating.

From the above-mentioned aspects, the following conclusion can be drawn: the capacity, progress and performance of every student are determined by the judgment of the rater, that is, the teacher, through his own perception regarding the real feedback his grades provide and through the placement of the student's performances within an indicative rating system, according to guidelines. To what extent this approach is a negative or positive one depends on the perspective from which the matter is looked at.

Thus, relying on the ethical attitude, the impartiality and moral conduct of the teacher, this type of evaluation is very accurate, due to his/hers deep knowledge of students he/she is teaching. They are in the best position to assess the progress and performance of their students whom they know closely. Therefore, the assessment is accurate and the approach is a positive one.

On the other hand, the objectivity of statistics, math and the absolute impartiality standardized scores, disputes the effectiveness of humanist-approached assessment in the aforesaid manner, emphasizing the impossibility of absolute objectivity of a single individual. If the evaluation is not objective but – at least to some extent – equivocal, under the influence of individual perception, then assessment using only an interpretative guiding framework is inadequate. Using subjective interpretation of guidelines to produce a precise objective grade which will impact the student throughout his/hers educational progress seems a negative approach.

Standardized assessment, like all sciences that underpinned technological progress, relies on the scientific method. The scientific method itself relies on precision. Therefore, there is no "hot" or "cold" since these are subjective notions. According to the scientific method, there is only the thermometer, having 0 as reference point and plus or minus grades as values. In this way, the scientific method ensured that the situated meaning of the words "hot" or "cold" can change accordingly, without influencing the whole of science. The entire process of technological evolution is grounded in a precise numerical measurement system, particularly because it is impartial, universal, and thus objective.

While our national standardized tests have the aim to objectively measure school performance of learners in relation to the qualitative baseline of Romanian education, they tend not to be very accurate. This lack of accuracy seems to originate not from the fact that they aren't standardized, but from the way they are standardized. If rating the grades of this standardized test is being done by human raters, there will be differences, and the more raters involved, the higher the differences. If the differences between them are too high – which in many cases are – then the mean obtained is highly inaccurate from a statistical point of view. Therefore, the subjective judgment of the raters is again the core of the problem, resulting in inaccurate results. (Herman et al. 2005).

The purpose of this article however is not to criticize or evaluative methods used in Romanian formal education at present, but rather to try and assess the results of some aspects of school education, using methods statistically similar in accuracy to those used by it to assess learners.

Within this context, we've taken the liberty to elaborate a multiple choice general knowledge test, composed of 6 items consisting of general knowledge questions, with three possible answers and one correct one, identifiable even through elimination.

We will henceforth refer to it as "speculative study". The reason that we call it speculative is connected to its purpose, that is, to speculate on why the results of it are the way they are, without re-testing the claims made. The claims will be just educated guesses which encourage further study in order to determine what is really happening, and explain many of the scores resulted from the testing. In addition, the difficulty of the questions may be debatable and will be debated on. Either way, as far as we are concerned, the information questioned is of basic level, general-purpose knowledge. We will assess the formal education system in this way which will leave many questions unanswered, underlining the many questions the assessed are facing when evaluated in such a way.

The point is to realize that there is much uncertainty in the current assessment system, and using the same type of accuracy when assessing the educational system itself the results are debatable, which will prove the lack of accuracy within hybrid statistical-human rated assessment systems.

The test was performed at a state university – whose name cannot be published for ethical reasons – on first year law students, during March 2015. Population sample: 300 students, both genders; content sample: "Multiple choice trivia".

General knowledge test contains questions from two curriculum areas: "Mathematics and Sciences" and "Man and Society". Thus, the questions are related to the following subjects: mathematics, physics, chemistry, geography, history and biology. The reason behind choosing only two subjects in the

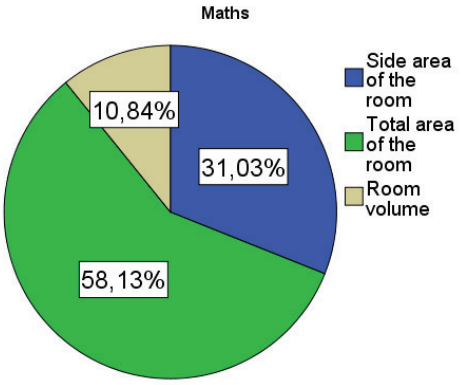
curriculum area "Man and Society" is due to the perspective of the study: for disciplines focused on problem-solving, which predominantly rely on logical-mathematical intelligence (Gardner 1989) three are tested: mathematics, physics and chemistry; and for disciplines focused on content acquisition, which predominantly uses mnemonic capacity, other three: geography, history and biology, having thus parity between disciplines.

The study tracks only performance indices. Incomplete tests were canceled, which reduced the actual population sample just 203. The performance indicator was the score to the test, which granted 1 point for a correct answer and 0 for wrong answers. The data were processed using SPSS Statistics v. 17.0, and reports focused only on frequencies and averages.

Data were obtained both from responses to each item, and from averages and average comparisons between the two categories of disciplines, namely problem solving and content acquisition.

The first item consisted of questioning involving a mathematical way:

Fig. 1 – item 1: *"You need to paint the walls of a room but do not know how much paint you are going to need. What should you calculate in order to find out?*



Statistics

	Math	Score
Valid	203	203
Mean		,31

Score

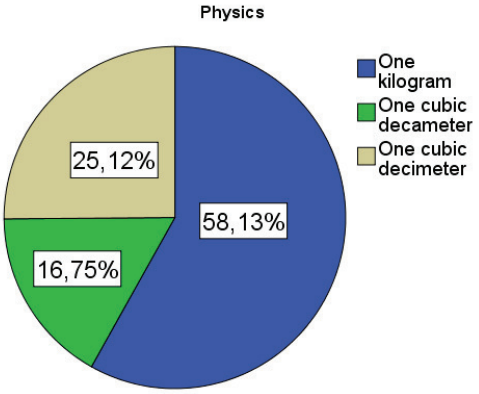
	Frequency	Percent
0	140	69,0
1	63	31,0
Total	203	100,0

Of the 203 participants, only 63 knew what computing is necessary to solve the problem. A percentage of 58.13% of participants 118 respectively, considered calculating the total area and 10.84% respectively 22 participants felt that the volume could be necessary calculation.

These responses indicate a complete lack of comprehension of the notions of area and/or volume. If a fundamental concept of space geometry is not understood, it means that it is not actually acquired. A successfully acquired notion implies the possibility of using it in different contexts,

meaning it must be embedded into the vocabulary. If it cannot be used logically, it misses the purpose of its acquisition.

The second item verifies the understanding of basic measurement units. Fig. 2 – item 2: ”*How much is one liter?*”



Statistics

	Physics	Score
Valid	203	203
Mean		,25

Score

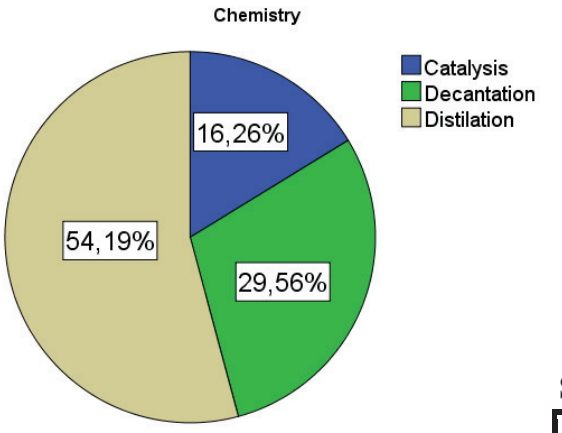
	Frequency	Percent
0	153	75,4
1	50	24,6
Total	203	100,0

Of the 203 participants, 50 answered correctly – 25.12% and a total of 85 participants, representing 41.87% of the total, have correctly identified what the liter is intended to measure. Assuming that is a common mistake to confuse multiples with subs within the units, one aspect remains alarming: 58.13% of participants, respectively 118, which got the answer completely wrong, prove that the unit of measurement is not understood, by confusing a unit of mass – gram – with one of volume.

Of course, one can say that it is a common mistake. But the essence of the problem is that these students have made calculations on physics problems using measurement units and converting them while they had no idea what they were actually doing. It remains hence questionable the efficiency of conducting problems whose results may be accurate, but are not understood.

The third item concerns completely different concepts of chemistry, namely: decantation, catalysis and distillation.

Fig 3 – item 3: "*Which of these processes is used in order to clarify liquids containing suspended solids?"*



Statistics

	Chemistry	Score
Valid	203	203
Mean		,30

Score

	Frequency	Percent
0	143	70,4
1	60	29,6
Total	203	100,0

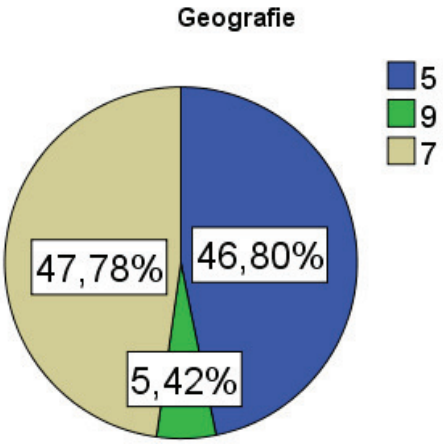
In this case, the notions should not be confused in any way. There is no question of flaws in logic, misunderstanding or gaps in content. Of the 203 participants, 60 or 29.56% knew what decantation or settling means. Any other choice involves not knowing at least two terms out of the three: the wrong answer chosen and the correct one. While catalysis is a chemical process specific to the oil industry, distillation is a common process even used in household activities. However, 54.19% – 110 participants thought of distillation as the correct answer.

The fourth item contains a simple trivia question

Fig. 4 – item 4: *" Conventionally, how many continents are on Earth?"*

Statistics

	History	Score
Valid	203	203
Mean		,82



Score

	Frequency	Percent
0	106	52,2
1	97	47,8
Total	203	100,0

A percentage of 46.80% – 95 participants left out two continents, and a very small percentage of 5.42% – 11 participants, adds two extra.

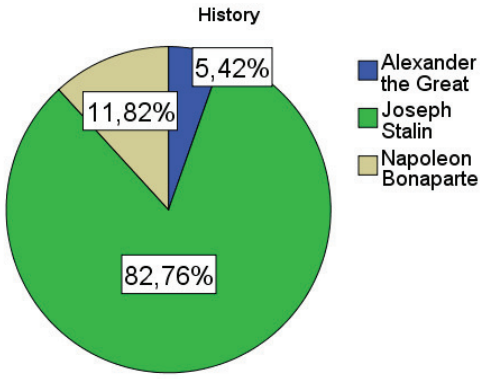
It cannot be said that this question has any practical echo, or that the information is widely usable, but nonetheless it is a basic information on the planet we live on and its general conventions. A share of 52.20% and 106 first-year law students who do not know how many continents are on Earth could mean that their attendance to geography lessons was at least partially, in vain.

The fifth item contains a universal history question.

Fig 5 – item 5: ”Which of the mentioned leaders was contemporary with Adolf Hitler?”

Statistics

	Geography	Score
Valid	203	203
Mean		,48



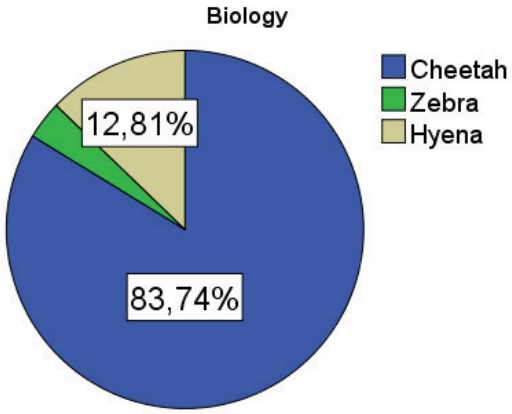
Score

	Frequency	Percent
0	36	17,7
1	167	82,3
Total	203	100,0

To this question 82.76% – 167 participants responded correctly and 36, 17.24% respectively, did not. We assume that some respondents have deliberately answered wrong, either in protest or for other reasons, because this question is likely to be one of those that had the potential to "insult the intelligence" of participants, leaving the possibility of a wrong answer as a form of protest.

The last question involves testing some very general notions from biology.

Fig. 6 – item 6: *” Which of the following animals is a feline?”*



Statistics

	Biology	Score
Valid	203	203
Mean		,83

Score

	Frequency	Percent
0	34	16,7
1	169	83,3
Total	203	100,0

Again, this question has been answered correctly by 169 participants 83.74% respectively, zebra is considered a feline by only 7 participants and hyena by 26. This information in turn could ever be obtained from the multitude of television documentaries on specialized channels, which address mostly topics regarding felines and reptiles.

Comparing the averages of the two categories of questions indicates a big difference: the average score is 0.28 for the problem-solving category and 0.73 on content acquisition category.

Let us presume that we have a grand total of the two and apply an arithmetic average calculation on them. The result will be 0.50. It might lead to the conclusion that the grade of the questioned students is on this particular test is thus, 5,00. Instead of having a grade of 2,80 in problem-solving and 7,30 on content acquisition, we made a mean out of two, very different scores which resulted from the same students.

Obviously the problem is not only reflected by the manipulation of the grades which is by far misleading, but also from the low number of students participating – although we have reasons to believe that in this particular case, average numbers would have indicated the same regardless of the population sample size – and by the fact that there was only one question addressing each of the disciplines.

Many would consider this study a rude joke, but in doing so, considering the statistical accuracy similarities between our study to assess undergraduate education and the way undergraduate education actually assesses are quite many. And while the flaws of this study are many according to the scientific method, and its relevancy is only indicative, the same can be said about formal assessment. While the averages we calculated reveal, basically nothing in statistical terms, the same type of rating behavior is being used in educational assessment.

The changes of a student who has obtained the grades of 7,50; 7,35; and 7,16 on a national test to be admitted into graduate school are the same with the students whose grades were 10; 6,05; and 6. The entry system disregards the fact that maybe the student wants to specialize in mathematics, excels in mathematics but performs poorly on the other subjects of the exam. Being both “at the same level” with an average 7,33 they are competitors while

things should be clear from this point of view: the one that is mediocre is just mediocre, and the other is below mediocre on two of the subjects but excels in one.

In the case depicted above, poor handling of means and averages can cause a lot of confusion, frustration and counter-productivity and from a statistical point of view, terribly flawed and inaccurate.

At the same time, the data we've revealed tends to only confirm a suspicion about the formal educational system: it is centered on content delivery. Delivery of content though is not learning, but again, since the tests are so statistically poor in accuracy and present many vices, they can't actually be taken into account.

To conclude, while the educational evaluation and assessment system still runs on approximate information which fuel precise numbers with a permanent impact on the student, a question remains: is this right? Or something should be changed, in either addressing the problem of properly standardized assessment or to keep the traditional methods and remove the precise noting with an indicative system, since assessment cannot be precise and indicative at the same time.

References:

Kamenetz, Anya (2014). *Testing: How Much Is Too Much?* retrieved 11.06.2015 http://www.npr.org/sections/ed/2014/11/17/362339421/testing-how-much-is-too-much

Legea educației naționale nr. 1/2011 cu modificările și completările ulterioare, retrieved 11.06.2015 http://www.edu.ro/index.php/base/frontpage

Dicționarul explicativ al limbii române, versiunea pentru desktop

Herman Joan L., Webb Noreen M., Zuniga Stephen (2005). CSE Report 653, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), *Measurement Issues in the Alignment of Standards and Assessments: A Case Study,* Graduate School of Education and Information Studies University of California, Los Angeles. retrieved https://www.cse.ucla.edu/products/reports/r653.pdf

Gardner, H., & Hatch, T.; Hatch (1989). *Multiple intelligences go to school: Educational implications of the theory of multiple intelligences* (PDF). Educational Researcher 18 (8): 4. retrieved http://www.sfu.ca/~jcnesbit/EDUC220/ThinkPaper/Gardner1989.pdf