What is Wrong with Teaching to the Test: Uncertainty Techniques Help in Understanding (and Hopefully Resolving) the Controversy

Olga Kosheleva
Department of Teacher Education
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
Email: olgak@utep.edu

***Abstract*—In the USA, in the last decade, standards have been adapted for each grade level. These standards are annually checked by state-wide tests. The results of these tests often determine the school's funding and even the school's future existence. Due to this importance, a large amount of time is spent on teaching to the tests.**

Most teachers believe that this testing approach is detrimental to student education. This belief seems to be empirically supported by the fact that so far, the testing approach has not led to spectacular improvements promised by its proponents. While this empirical evidence is reasonably convincing, the teacher community has not yet fully succeeded in clearly explaining their position to the general public – because the opposing argument (of the need for accountability) also seems to be reasonably convincing.

In this paper, we show that the situation becomes much clearer if we take uncertainty into account – and that, hopefully, a proper use of uncertainty can help in resolving this situation.

What is "teaching to the test"? In the last few decades, in the US school education, state-wide tests have been developed for testing the mathematical knowledge of students at the end of each grade. Student performance on these state tests (and on similar tests in other disciplines) has become the most important criterion of how the performance of schools and teachers are gauged:

- Funding of individual schools is largely determined by the test results.
- In some cases, schools are disbanded and teachers are fired if the test results are unsatisfactory several years in a row.

Because of the importance of the test results, schools are understandably paying a large amount of attention to making sure that the students pass these tests. In other words, instead of spending most of the time teaching the material – as it was in the past – teachers now spend a significant amount of time teaching "to the test".

The results of teaching to the test are not as spectacular as the proposers hoped. The main idea behind the tests sounds reasonable:

- if we do not gauge how well students are doing in a way that will enable us to compare different schools and different teachers,
- then how will we know which schools are doing better and which schools need improvement?

The authors of this idea expected that with well-established testing, the students' knowledge will drastically improve. Alas, these expectations turned out to be too optimistic:

- In some states and some school districts, there has been some improvement.
- However, overall, this program has not been a spectacular success as its proponents hoped.
- Moreover, there is anecdotal evidence that in many cases, with the introduction of state-wide testing, the students' knowledge actually decreased.

Teaching to the test: a current controversy. There is a current controversy, both in the media and in the scholarly publications (see, e.g., [2], [3], [14]) about the state-wide tests and teaching to the test.

In a simplified form, the controversy is fought along the following lines:

- on the one hand, many politicians believe that tests are a good idea (while they acknowledge the limitations and drawbacks of the existing tests, and the need to improve the tests);
- on the other hand, most teachers strongly believe that the entire approach of teaching to the tests is flawed and detrimental to education.

In the media, this controversy gets personal and nasty:

- Some politicians accuse the teacher community and the teacher unions of defending weak under-performing teachers at the expense of the student success.
- Some teachers accuse the politicians of ignorance-motivated over-simplified populist interference with a complex teaching process.

While there may be some anecdotal evidence in support of such accusations, the situation is definitely more complex than the simplified picture one may get from the media coverage:

- on the one hand, several knowledgable politicians, with successful teaching experience, are in favor of the state-wide tests,
- on the other hand, many very good teachers, teachers who are extremely successful in teaching, are strongly against the current emphasis on these tests.

Population is somewhat confused. One of the frustrating aspects of the current controversy is that the general population is somewhat confused about it:

- on the one hand, it is reasonable to require accountability, and this accountability logic naturally leads to the current testing program;
- on the other hand, respected teachers are against this program, and empirical evidence also shows that it has not led to spectacular successes – contrary to natural expectations motivated by accountability.

What we do in this paper. In this paper, we argue that the confusion – and, to some extent, the controversy itself – is largely due to the simplification of the complex pedagogical process. Specifically, we argue that if we properly take uncertainty into account, then the situation becomes much clearer.

In this explanation, we follow up on our previous research [1], [4], [6], [7], [8], [9], [10], [11], [13] on taking into account uncertainty (in particular, fuzzy uncertainty; see, e.g., [5], [12]) when gauging the pedagogical process.

The background of our main idea. In general, it is assumed that learning comes from repetitions:

- once a student has repeated a certain procedure certain number of times,
- the student has mastered it.

This is why an important part of learning each idea of high school mathematics is practice. For example:

- unless students do a lot of exercises where they have to add fractions,
- they will not master this skill well enough to be able to easily add two fractions, and
- this will hinder their progress in the following mathematical topics like dealing with polynomials (where the ability to add fractions is already assumed).

In general:

- the only way to learn to write is to practice writing,
- the only way to learn to spell is to practice spelling,
- the only way to learn a foreign language is to practice it, and
- the only way to learn the multiplication table is to practice multiplication.

The required number of repetitions depends:

- on the complexity of the topic,
- on the match between this particular topic and the student's individual interests and prior skills,
- etc.

However, the fact remains:

- for every topic and for every student,
- there is a number of iterations after which the student will master this topic.

From this viewpoint, let us analyze both the traditional teaching process and the new situation of teaching to the test.

Analysis of the traditional teaching process. The main objective of a school mathematics program is that after graduation, students should have certain skills. These skills often build on each other, so that one skill requires another one.

For example:

- In order to be able to solve quadratic equations, we need to know how to add, how to subtract, how to multiply, etc.
- In order to be able to handle polynomials with rational (fractional) coefficients, a student needs to be able to perform arithmetic operations with fractions, etc.

To illustrate our main idea, let us consider a simple sequence of two skills A and B, for which the skill B requires that the student also has learned skill A.

For example, we can take,

- as A, the ability to add, subtract, multiply, and divide fractions, and
- as B, the ability to process polynomials with rational coefficients.

For simplicity, let us assume that

- skill A is learned in one year, and
- skill B is learned in the following year.

At the end of the A-B sequence, a student should have mastered both skills A and B. Let us assume that the student needs

- n_A iterations to master skill A, and
- n_B iterations to master skill B.

This means that by the end of the school education, a student should have done

- n_A iterations of skill A and
- n_B iterations of skill B.

Since the skill B is only taught in Year 2, we should have all n_B iterations in Year 2. However, since practicing the skill B often involves practicing skill A, a student practices the skill A also when she learns skill B.

Not every problem related to learning skill B necessarily involves skill A. For example, when the students learn that $(a+b)\cdot(a-b) = a^2 - b^2$ or that $(a+b)^2 = a^2 + 2\cdot a\cdot b + b^2$, these polynomial problems do not include any operations with fractions.

Let us denote by r the proportion of problems of type B that involve using skill A. Then, during n_B exercises needed to master skill B, the student, in effect, performs $r \cdot n_B$ exercises in which she practices skill A as well.

In the traditional teaching approach, without annual checks, where the only objective is mastery at the end, all we therefore need is to be have a total of n_A exercises in skill A by the end of Year 2.

- In Year 2 we, in effect, have $r \cdot n_B$ exercises in this skill.

- This means that it is sufficient to have $n_A - r \cdot n_B$ exercises in skill A in Year 1.

Yes, this number $n_A - r \cdot n_B$ is smaller than n_A, so by the end of Year 1, the students have not yet fully mastered skill A, but this is normal in education – the skills come with practice.

How situation changes when we teach to the test. What happens when we teach to the test?

- According to the school program, Year 1 is devoted to teaching skill A.
- So, to test how well the students learned after this year, it is reasonable to design a test with questions about skill A.

Once we give this test, the results are far from perfect – because, as we have mentioned, by the end of Year 1, the students only had $n_A - r \cdot n_B < n_A$ exercises, and they have not yet mastered this skill. The argument "Is this how much we want our graduates to know about A?" sounds convincing, so a pressure is placed on schools to improve the score on the test at the end of Year 1.

The only way to do it is to increase the number of skill-A-related exercises in Year 1 to n_A.

Teaching to the test: a seemingly positive result. Now, the test grades for Year 1 go up – because:

- in the past, the students did not have enough exercises to master skill A, while
- now, they have enough exercises, so they do master skill A at the end of Year 1.

The progress is visible, results are good. But are they?

Teaching to the test: while there is a drastic improvement in test scores, there is no significant improvement of the school graduates knowledge. Let us see how teaching to the test affects the main school objective – to make sure that the graduates learn both skills A and B at the end of both years. Let us show that with respect to this criterion, we should not expect any significant improvement. Indeed:

- in the past, we had a total of n_A exercises in skill A:
- now, the students have $n_A + r \cdot n_B$ exercises in skill A.

In both cases, we have enough exercises to master skill A. So, in both cases, we should have the same reasonably positive result.

Teaching to the test: a serious problem. At first glance, it may seem like teaching to the test may not be as spectacularly successful as its proponents claim, but overall, its results are positive – at least the test grades improved.

However, there *is* a serious problem. The problem is that school time is limited. The time that now schools use to practice additional $r \cdot n_B$ repetitions of skill A in Year 1 has to come at the expense of something else. Clearly, it comes at the expense of other topics that are not explicitly included in the statewide test.

As a result,

- while students' knowledge of the topics included in the test (like skills A and B) does not decrease (and many even increase),
- the students' mastery of some other skills will necessarily drastically decrease.

This is what teachers object to when they object to "teaching to the test".

We clarified the problem – but what is a solution? In the above text, we explained the problem. This explanation also helps to find a solution to this problem.

In order to compare different schools, to compare different teachers, we need to have some objective way of gauging the student success. But we need to make sure that this comparison does not hinder the students' progress.

In the ideal world, we should design better tests – this is one of the few things with which everyone agrees. However, even with the existing tests, we can drastically improve the situation if we *no longer require* that

- at the end of each school year,
- students should have a perfect knowledge of all the topics that they learned during this year.

This requirements comes from the "crisp" thinking, thinking that does not take uncertainty into account – a student either mastered the skill or did not. In reality, after a few exercises of the skill A, a student usually achieves mastery *to a degree*.

As a result, in the traditional approach,

- the student will have an imperfect score on the test for skill A at the end of Year 1;
- this is OK, as long as this score is what we should expect after $n_A - r \cdot n_B$ exercises, so that we will be sure that
  - after additional $r \cdot n_B$ exercises involving skill A in Year 2
  - the student will achieve the true mastery of skill A.

This should be the guideline to developing required satisfaction level on annual exams. Any increase of this satisfaction level should be *discouraged* because

- it would indicate that the teachers are over-emphasizing skill A in Year 1, while
- they could use fewer exercises of this particular skill and spend the remaining time teaching the students some other useful skills.

How fuzzy logic can help. Fuzzy logic has been explicitly designed to handle situations in which some property is true to a degree – so fuzzy logic seems to be a perfect tool for this analysis.

Our idea is more general than teaching-to-the-test controversy. Our main objective is to help in understanding and resolving the "teaching to the test" controversy. However, the same idea can be applied to all levels of education as well. We should not aim for perfect knowledge on intermediate classes.

For example, college students taking a computer science sequence may be somewhat shaky about programming at the end of the first class, but their basic skills are reinforced in the following classes.

We used this idea in [1], [7] to plan an optimal teaching schedule, and it worked. We hope that this idea can lead to further improvement of the teaching process.

Acknowledgments. The author is thankful to the anonymous referees for valuable suggestions.

REFERENCES

[1] R. Aló and O. Kosheleva, "Optimization Techniques under Uncertain Criteria, and Their Possible Use in Computerized Education", *Proceedings of the 25th International Conference of the North American Fuzzy Information Processing Society NAFIPS'2006*, Montreal, Quebec, Canada, June 3–6, 2006, pp. 593–598.

[2] A. Bell, "Teaching for the test", In: M. Selinger (ed.), *Teaching Mathematics*, The Open University Press, London, UK, 1999, pp. 41–46.

[3] U. Boser, "Teaching to the test?", *Education Week*, June 7, 2000.

[4] O. Castillo, P. Melin, E. Gamez, V. Kreinovich and O. Kosheleva, "Intelligence Techniques Are Needed to Further Enhance the Advantage of Groups with Diversity in Problem Solving", *Proceedings of the 2009 IEEE Workshop on Hybrid Intelligent Models and Applications HIMA'2009*, Nashville, Tennessee, March 30–April 2, 2009, pp. 48–55.

[5] G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and +Applications*, Upper Saddle River, New Jersey: Prentice Hall, 1995.

[6] O. Kosheleva, "Potential Application of Fuzzy Techniques to Math Education: Emphasizing Paradoxes as a (Seemingly Paradoxical) Way to Enhance the Learning of (Strict) Mathematics", *Proceedings of the 27th International Conference of the North American Fuzzy Information Processing Society NAFIPS'2008*, New York, New York, May 19–22, 2008.

[7] O. Kosheleva and R. Aló, "Towards Economics of Education: Optimization under Uncertainty", *Proceedings of the Second International Conference on Fuzzy Sets and Soft Computing in Economics and Finance FSSCEF'2006*, St. Petersburg, Russia, June 28 – July 1, 2006, pp. 63–70.

[8] O. M. Kosheleva and M. Ceberio, "Processing Educational Data: From Traditional Statistical Techniques to an Appropriate Combination of Probabilistic, Interval, and Fuzzy Approaches", *Proceedings of the International Conference on Fuzzy Systems, Neural Networks, and Genetic Algorithms FNG'05*, Tijuana, Mexico, October 13–14, 2005, pp. 39–48.

[9] O. Kosheleva and V. Kreinovich, "What is the Best Way to Distribute Efforts Among Students: Towards Quantitative Approach to Human Cognition", *Proceedings of the 28th North American Fuzzy Information Processing Society Annual Conference NAFIPS'09*, Cincinnati, Ohio, June 14–17, 2009.

[10] O. Kosheleva and V. Kreinovich, "Towards Optimal Effort Distribution in Process Design under Uncertainty, with Application to Education", *Proceedings of the 4th International Workshop on Reliable Engineering Computing REC'2010*, Singapore, March 3–5, 2010, pp. 509–525.

[11] O. Kosheleva and F. Modave, "Asymmetric Paternalism: Description of the Phenomenon, Explanation Based on Decisions Under Uncertainty, and Possible Applications to Education", *Proceedings of the 27th International Conference of the North American Fuzzy Information Processing Society NAFIPS'2008*, New York, New York, May 19–22, 2008.

[12] H. T. Nguyen and E. A. Walker, *First Course on Fuzzy Logic*, CRC Press, Boca Raton, Florida, 2006.

[13] P. Pinheiro da Silva, A. Velasco, and O. Kosheleva, "Degree-Based Ideas and Technique Can Facilitate Inter-Disciplinary Collaboration and Education", *Proceedings of the Annual Conference of the North American Fuzzy Information Processing Society NAFIPS'2010*, Toronto, Canada, July 12–14, 2010, pp. 388–393.

[14] W. J. Popham, "Teaching to the test?", *Educational Leadership*, 2001, Vol. 58, No. 6, pp. 16–20.