Shall We Place More Advanced Students in a Separate Class?

Shahnaz Shahbazova[1], Olga Kosheleva[2], and Vladik Kreinovich[2]

[1]Department of Digital Technologies and Applied Informatics
Azerbaijan State University of Economics UNEC
Baku, AZ1001, Azerbaijan, shahnaz_shahbazova@unec.edu.az
and Institute of Control Systems
Ministry of Science and Education of the Republic of Azerbaijan
Baku, Azerbaijan, AZ1141, shahbazova@cyber.az

[2]Olga Kosheleva and Vladik Kreinovich
University of Texas at El Paso, 500 W. University
El Paso, Texas 79968, USA, olgak@utep.edu, vladik@utep.edu

1. Formulation of the problem

- Usually, in the same class section, we have both students who are more advanced and students who are struggling more.
- In view of this difference, in a class that has several sections, a seemingly reasonable idea is to separate students by their level:
  - more advanced students form one section,
  - students who are struggling form another section.
- Interestingly, empirical data shows that this seemingly reasonable idea does not work:
  - when we place more advanced students in a separate class,
  - the average knowledge level decreases.
- This empirical observation seems to contradict our intuition.
- So, a natural question is: why is it happening?
- How can we provide a convincing explanation for this empirical phenomenon?

2. Word of caution

- It is important to emphasize that it is about students from the same class.
- These students who have passed the same pre-requisite courses and are at approximately the same level of knowledge.
- We are not talking about placing students who are taking high-school algebra and students who are taking calculus in the same class.
- And we are not talking about schools like the ones that existed in small villages in the old times:
  - when students from the first grade all the way to the tenth grade would study together in the same classroom,
  - with the same teacher.
- We are talking about students who, e.g., all have taken algebra and who are now taking calculus.

3. Word of caution (cont-d)

- In the process of taking calculus:
  - some students are somewhat faster to grasp the main concepts, and
  - some struggle more tying to better understand these concepts.
- Another point: while division by level is counter-productive, other divisions can help.
- For example:
  - when teaching calculus or introduction to computer science to different majors,
  - it is a good idea to have, e.g., a section for physics majors and a section for biology majors.

4. Word of caution (cont-d)

- This makes perfect sense, since:
  - the usual way of teaching calculus is
  - to give physical concepts of velocity and acceleration as examples of what is a derivative.
- To physics majors, for whom these notions are natural, this helps.
- However, for many biology majors these examples are only confusing.
- For them, it is better to give biological examples such as growth rate, etc.

5. Our personal experience

- One of us (VK) went to a special mathematical high school in St. Petersburg, Russia – which was then part of the Soviet Union.
- There, math was taught by one of the city's best math teachers.
- We had ten different classes on each grade level.
- The teachers could have easily formed a class consisting of more advanced students and classes consisting of less advanced students.
- However, instead, our classes contained both.
- For example, in our class, we had both students who won National High School competitions and students who barely succeeded.
- Some of them even had to hire private tutors to catch up.
- One may think that the reason was the communist emphasis on equality – not true in our case.

6. Our personal experience (cont-d)

- The teachers formed the classes the way they liked it:
  - our school Principal was regularly reprimanded for the fact that
  - our school did not follow the government recommendations on equal representation of different social and ethic groups.
- No, this was the teachers' decision, motivated by their experience.
- Experience showed them that not placing advanced students in a separate class leads to better learning results.
- The same thing happened when VK:
  - after graduating form high school,
  - became a student at the Mathematics Department of St. Petersburg University.
- The incoming students were divided into groups (cohorts), each of which took all the classes together.

7. Our personal experience (cont-d)

- And again, with more than a dozen such groups, it could be possible for the administration to form:
  - groups with more advanced students and
  - groups with less advanced ones.
- Instead, each group had both more advanced and less advanced students.
- And again, this was not the result of the outside social pressure.
- Our department strongly resisted this pressure.
- One day, a KGB officer came to announce that one of the professor was arrested for reading Samizdat.
- These were books like Orwell that were proclaimed illegal by the authorities.
- The only question professors asked was when the professor will be released – so that we will be able to hire him again.

8. Our personal experience (cont-d)

- No, this division into groups was done by the professors themselves, who realized that this leads to better educational outcomes.
- In short, the experience of our own very good teachers confirmed the above-mentioned empirical fact. I
- It should be mentioned that we, the students, were not always very happy with this arrangement.
- Those who were more advanced believed that placing them in a separate group would be better for them – but the teachers knew better.

9. Let us take into account that students help each other

- To better understand the phenomenon in question, let us take into account that:
  - learning is not just an interaction between a teacher and a student;
  - a significant part of learning occurs when students work together, both in class and when studying together.
- This helping each other is an important part of learning.
- Students usually interact when one of the students is somewhat more advanced than the other one.
- In this case:
  - a student who is somewhat less advanced asks questions to – and solicits other type of help from
  - a student who is more advanced.

10. Students help each other (cont-d)

- It is clear that in this exchange, the less advanced student gains additional knowledge.
- It is somewhat less clear – but still true – that the more advanced student also benefits from this exchange:
  - first, when answering questions, he/she may find an unexpected gap in their knowledge and thus fix it;
  - second, he/she also learns how to better explain to others – which is an important part of knowledge.

11. Personal experience

- One of VK's first papers was when he – while still a student – solved a problem formulated by Professor Aleksandr F. Timan.
- It was completely rewritten by Professor Timan before publication, to make it clearer.
- VK became upset by the low clarity of his original text.
- Professor Timan calmed him down by saying that the ability to write clearly comes with teaching.
- And students teaching other students is an important part of this teaching.

12. Let us come up with a mathematical model of the phenomenon

- Fuzzy techniques were originally designed to translate imprecise natural-language phrases into a precise model.
- In this spirit, let us translate the above imprecise description into a precise model of th above phenomenon.

13. Towards a precise model of the corresponding phenomenon

- How does communicating between two students affect their knowledge?
- Let us consider the communication between two students.
- Let us denote their original levels of knowledge by a and b.
- This level can be measured, e.g., by the grade they would have got on a comprehensive final exam (if it was given at this time) –
- Without losing generality, we can assume that $a \le b$.
- After the communication, the students' knowledge levels increase, to some values $\ell(a,b) > a$ and $u(a,b) > a$.
- These new levels depend on their original levels a and b.
- If the students are at the exact same level of knowledge, there is not need for any communication.

14. Towards a precise model of the corresponding phenomenon (cont-d)

- So, when $a = b$, communication does not change anything:

$$\ell(a, a) = a \text{ and } u(b, b) = b.$$

- As we have mentioned earlier:
  - while there is a difference between the knowledge levels,
  - these are still students on the same level of their overall degrees, students with the same pre-requisite knowledge.
- From this viewpoint, the difference $b - a$ between their levels of knowledge is relatively small.
- We can explicitly express the desired functions in terms of this difference, if we take into account that

$$b = a + (b - a) \text{ and } a = b - (b - a).$$

- Thus, we have

$$\ell(a, b) = \ell(a, a + (b - a)) \text{ and } u(a, b) = u(b - (b - a), b).$$

15. Towards a precise model of the corresponding phenomenon (cont-d)

- Since the difference $b-a$ is small, terms proportional to the square of this difference are much smaller that the difference itself.
- For example, if the difference itself is 10%, its square is 1%, which is much smaller than 10%.
- Terms which are cubic or even higher order in terms of this difference are even smaller.
- Thus, we can use the general technique common in physics – we can:
  - expand the dependence on $b-a$ in Taylor series and
  - keep only linear terms in this expansion.
- Thus, we conclude that

$$\ell(a, a+(b-a)) = \ell(a, a) + c_\ell(a) \cdot (b-a) \text{ and}$$

$$u(b-(b-a)) = u(b, b) + c_u(b) \cdot (b-a).$$

16. Towards a precise model of the corresponding phenomenon (cont-d)

- Here we denoted

$$c_\ell(a) \stackrel{\text{def}}{=} \frac{\partial \ell(a, a+h)}{\partial h}_{|h=0} \text{ and } c_u(a) \stackrel{\text{def}}{=} -\frac{\partial u(b, b-h)}{\partial h}_{|h=0}.$$

- Since $\ell(a, a) = a$ and $u(b, b) = b$, we get

$$\ell(a, a+(b-a)) = a+c_\ell(a)\cdot(b-a) \text{ and } u(b-(b-a)) = b+c_u(b)\cdot(b-a).$$

- So, the knowledge of a increases by $c_\ell(a) \cdot (b-a)$ and the knowledge of b increases by $c_u(a) \cdot (b-a)$.
- Thus, the summary knowledge of both students increases by the amount $c_\ell(a) \cdot (b-a) + c_u(b) \cdot (b-a)$.
- Usually, we have several pairs of communicating students (a_i, b_i), $i = 1, 2, \ldots,$.

17. Towards a precise model of the corresponding phenomenon (cont-d)

- Then the overall increase I in knowledge is equal to the sum of such terms:
$$I = \sum_i (c_\ell(a_i) \cdot (b_i - a_i) + c_u(b_i) \cdot (b_i - a_i)).$$
- We can simplify this expression even further if we again use the fact that quadratic terms:
  - are much smaller than linear ones and
  - thus, can be safely ignored.
- Indeed, all the values a_i and b_i are close to each other.
- This means, in particular, that these values are close to some average knowledge value a_0.
- In other words, the differences $a_i - a_0$ and $b_i - a_0$ are small.

18. Towards a precise model of the corresponding phenomenon (cont-d)

- Thus, we can:
  - expand the dependence of $c_\ell(a) = c_\ell(a_0 + (a - a_0))$ and $c_u(b) = c_u(a_0 + (b - a_0))$ in Taylor series and
  - keep only linear terms in this expansion:

$$c_\ell(a) = c_\ell(a_0) + (a - a_0)) = c_\ell(a_0) + c'_\ell(a_0) \cdot (a - a_0) \text{ and}$$

$$c_u(b) = c_u(a_0 + (b - a_0)) = c_u(a_0) + c'_u(a_0) \cdot (b - a_0).$$

- Here c', as usual, denotes the derivative.
- Substituting these expressions into the above formula, we conclude that

$$\sum_i (c_\ell(a_0) \cdot (b_i - a_i) + c_u(a_0) \cdot (b_i - a_i)) +$$

$$\sum_i (c'_\ell(a_0) \cdot (a_i - a_0) \cdot (b_i - a_i) + c'_u(a_0) \cdot (b_i - a_0) \cdot (b_i - a_i)).$$

19. Towards a precise model of the corresponding phenomenon (cont-d)

- The products $(a_i - a_0) \cdot (b_i - a_i)$ and $(b_i - a_0) \cdot (b_i - a_i)$ are quadratic in terms of the small differences $b_i - a_i$.
- Thus, we can safely ignore such terms, and come up with the final expression that we will use in this talk:

$$I = \sum_i c \cdot (b_i - a_i), \text{ where } c \stackrel{\text{def}}{=} c_\ell(a_0) + c_u(a_0) > 0.$$

20. Natural question and our answer

- Now, we have a precise model describing how knowledge increases when we divide students into communicating pairs.
- The natural question is: how to divide students into communicating pairs so as to maximize the overall knowledge increase.
- For simplicity, let us assume that we have an even number of students, so they *can* be divided into pairs.
- The answer comes from the following proposition.
- *Let us assume that we have* $2n$ *values* $v_1 < \ldots < v_{2n}$.
- *Then, for each division of these values into* n *pairs* (a_i, b_i) *with* $a_i < b_i$, *the following two conditions are equivalent to each other:*
  - *every lower element* a_i *is smaller than every upper element* b_j, *and*
  - *the value* I *corresponding to this division is the largest possible.*

21. Natural question and our answer (cont-d)

- What if we place more advanced students in a separate class, i.e.:
  - have a separate class formed by values $v_k < \ldots < v_{2n}$ for some k, and
  - keep other students in the original class.
- In this case, we cannot have pairs (a_i, b_i) in which one of the values if smaller than v_k, and another one is larger than or equal to v_k.
- Then, the following proposition holds:
- *Let us assume that we have $2n$ values $v_1 < \ldots < v_{2n}$ and that for some k, we do not allow pairs (v_i, v_j) in which $i < k \le j$.*
- *Then, for this division into pairs, the value I cannot be the largest possible.*

22. But why do people often place advanced students in a separate class?

- We have shown that:
  - from the pedagogical viewpoint – from the viewpoint of maximizing the amount of gained knowledge.
  - it is better *not* to place advanced students in a separate class.
- So why do people often place them?
- The answer is that while this separate-class placement decreases the amount of knowledge, it makes many students happier.
- Indeed, many people feel uncomfortable when they always have to interact with folks who are "smarter" (more advanced) than they.
- So, in a class in which students vary by their level, students with the lowest expected grade feel most uncomfortable.
- They will feel more comfortable if they are placed into a separate class, while all higher achievers form their own class.

23. But why do people often place advanced students in a separate class (cont-d)

- How can we fight this tendency?
- This tendency is caused by the fact that we are considering a typical US situation.
- In this arrangement, in each topic, in each class, instructors can form different sections.
- This negative tendency disappears (or at least decreases) when – like VK had in his university studies – students form groups.
- Then, for each topic, we have exactly the same sections.
- In each of the subjects, some members of the group are better, but in other subjects, other folks are better.

24. But why do people often place advanced students in a separate class (cont-d)

- One student may be the best in algebra, but not that good in geometry or history.
- Another student is good in a foreign language but not in PE, etc.
- This way:
  - each student is comfortable in some classes and not so comfortable in other classes,
  - and no rearrangement of the groups can change that.

25. Proof of Proposition 1

- To prove our equivalence result, we need to prove two things:
  - that if the first condition is *not* satisfied, then the maximum is *not* achieved, and
  - that if the first condition *is* satisfied, then the maximum *is* achieved.
- Let us prove this two statements one by one.
- Let us assume that the first condition is *not* satisfied, i.e., that we have $a_i > b_j$ for some elements a_i and b_j.
- Since we have $a_i < b_i$ and $a_j < b_j$, we thus have $a_j < b_j < a_i < b_i$.
- Let us show that in this case, we can increase the sum I by an appropriate rearrangement of pairs.
- Namely, instead of the pairs (a_i, b_i) and (a_j, b_j), we can take the new pairs (a_j, b_i) and (b_j, a_i).

26. Proof of Proposition 1 (cont-d)

- In this rearrangement, all other pairs remain intact; thus:
  - to show that the sum I increases after this rearrangement,
  - it is sufficient to show that the sum of the terms corresponding to the two changed pairs increases.
- Before the rearrangement, this sum was equal to $c\cdot(b_i-a_i)+c\cdot(b_j-a_j)$, i.e., to $c\cdot(b_i-a_i+b_j-a_j)$.
- After the rearrangement, this sum will be equal to $c\cdot(b_i-a_j)+c\cdot(a_i-b_j)$, i.e., to $c\cdot(b_i-a_j+a_i-b_j)$.
- Thus, the desired inequality has the form

$$c\cdot(b_i-a_i+b_j-a_j)<c\cdot(b_i-a_j+a_i-b_j).$$

- This is equivalent – if we divide both sides by a positive number c – to:

$$b_i-a_i+b_j-a_j<b_i-a_j+a_i-b_j.$$

27. Proof of Proposition 1 (cont-d)

- By subtracting identical terms in both sides, we get an equivalent inequality $-a_i + b_j < a_i - b_j$.
- This last inequality is clearly true, since we consider the case when $a_i > b_j$.
- In this case, the left-hand side of the inequality is negative and is, thus, smaller than the right-hand side – which is positive.
- The first statement is thus proven.
- Let us now assume that the first condition *is* satisfied.
- In this case, all upper elements b_i are larger than all lower elements a_j.
- Thus, the n smaller value $v_1, \ldots, v_n$ are lower elements and the n larger values $v_{n+1}, \ldots, v_{2n}$ are upper elements.
- The sum I can be equivalently described as

$$c \cdot \sum_i b_i - c \cdot \sum_i a_i.$$

28. Proof of Proposition 1 (cont-d)

- Thus, in this case, we have

$$I = I_0 \stackrel{\text{def}}{=} c \cdot \sum_{i=1}^{n} v_{n+i} - c \cdot \sum_{i=1}^{n} v_i.$$

- This value is the same for all such divisions into pairs.
- We have proved that the maximum value of I cannot be attained at any other division into pairs.
- So, the maximum is attained at one of these division, so this maximum is exactly I_0.
- The second statement is prove, and so is the Proposition.

29. Proof of Proposition 2

- We do not allow pairs formed by values from the two groups:
  - the smaller-valued group $v_1 < \ldots < v_{k-1}$ and
  - the larger-valued group $v_k < \ldots < v_{2n}$.
- Because of this, since the value v_k belongs to the larger-valued group, it can only be paired with elements from the larger-valued group.
- Since v_k is smaller than all elements of the larger-valued group, it cannot be one of the upper elements.
- So it must be one of the lower elements a_i.
- Similarly:'
  - since the value v_{k-1} belongs to the smaller-valued group,
  - it can only be paired with elements from the smaller-valued group.
- Since v_{k-1} is larger than all elements of the smaller-valued group, it cannot be one of the lower elements.

30. Proof of Proposition 2 (cont-d)

- So it must be one of the upper elements b_i.
- Here, one of the upper elements, namely v_{k-1}, is smaller than one of the lower elements, namely, v_k.
- And we have already proved, in Proposition 1, that in this case, the value I cannot be maximal.
- The proposition is proven.

31. Acknowledgments

This work was supported in part by:

- National Science Foundation grants 1623190, HRD-1834620, HRD-2034030, and EAR-2225395;
- AT&T Fellowship in Information Technology;
- program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478, and
- a grant from the Hungarian National Research, Development and Innovation Office (NRDI).