How to Divide Students into Groups so as to Optimize Learning: Towards a Solution to a Pedagogy-Related Optimization Problem

Olga Kosheleva[1] and Vladik Kreinovich[2]
[1]Department of Teacher Education
[2]Department of Computer Science
University of Texas at El Paso
El Paso, Texas 79968, USA
olgak@utep.edu, vladik@utep.edu

Abstract—To enhance learning, it is desirable to also let students learn from each other, e.g., by working in groups. It is known that such groupwork can improve learning, but the effect strongly depends on how we divide students into groups. In this paper, we describe how to optimally divide students into groups so as to optimize the resulting learning.

Index Terms—optimization, groupwork, uncertainty

I. INTRODUCTION

Groupwork as a way to teach better. Traditionally, students mostly learn from their instructor. The instructor presents the new material, asks the students to solve some related problems, and then provides individual feedback to students – explaining, to each student, his or her possible misunderstandings. Such an individual feedback is extremely helpful to the student. However, providing such an individual feedback requires a lot of time – especially in a class of reasonable size. So, if this feedback only comes from the instructor (and a Teaching Assistant), the amount of such feedback is limited. Moreover, in such situations, a significant amount of time is needed to grade the assignments of the whole class, so there is a significant delay between the test time and the time when students get their feedback.

It is well known that we can increase the amount of feedback – and decrease the delay of producing this feedback – if we also ask students from the class to provide useful feedback to each other. Different students have somewhat different misconceptions, so when a small group of students starts solving a problem together, they can often see each other's mistakes and provide corrections – thus, teaching each other.

Groupwork is not a panacea. While in principle, groupwork is efficient, its efficiency depends on how we divide students into groups. If we simply allow students to group themselves together, often, strong students team together and weak students team together. Strong students already know the material, so they do not benefit from working together. Similarly, weak students are equally lost, so having them solve a problem together does not help; see, e.g., [1], [2], [3], [4], [5], [6], [8], [9], [10], [11].

How to divide students into groups? Since the efficiency of groupwork depends on the subdivision into groups, to make groupwork as efficient as possible, it is desirable to find the optimal way to divide students into groups. This is the problem that we study in this paper.

II. DESCRIPTION OF A MODEL

How to describe the current state of learning. In this paper, we assume that for each student, the degree to which this student learned the material can be characterized by a single number – crudely speaking, this student's grade so far. In the following text, we will denote the number of students in the class by n, and we will denote the degree of knowledge of the i-th student by d_i, $i = 1, \ldots, n$. So, we arrive at the following definition.

Definition 1. *Let n be an integer; we will call this integer a* number of students. *By a* state of knowledge, *we mean a tuple $d = (d_1, \ldots, d_n)$ consisting of n non-negative numbers.*

Subdivision into groups. The following describes the general subdivision into groups.

Definition 2. *Let n be a number of students. By a* subdivision into groups, *we mean a subdivision of the set $\{1, \ldots, n\}$ into a finite number of non-intersecting subsets $G_1, \ldots, G_m$ for which $G_k \cap G_l = \emptyset$ for $k \neq l$ and $\bigcup_{k=1}^{m} G_k = \{1, \ldots, n\}$.*

In this paper, we will mostly consider subdivision into groups of equal size.

How groupwork helps students: a description. If two students with degrees $d_i < d_j$ work together, then the degree of knowledge of the i-th student increases. As we have mentioned earlier, if two students are at the same level of knowledge, there is not much that they can learn from each other. The more the j-th student knows that the i-th student doesn't, the more the i-th student will learn. So, it is reasonable to assume that the amount of material that the i-th student learns is proportional to the difference $d_j - d_i$, with some known coefficient of proportionality α. Thus, after the groupwork, the new level of knowledge of the i-th student is equal to $d_i' = d_i + \alpha \cdot (d_j - d_i)$.

If more than two students work together, then each student learns from all the students from the group who have a higher degree of knowledge. For example, if three students, with original degrees of knowledge $d_i < d_j < d_k$, work together, then after the groupwork, their new levels of knowledge are equal to $d'_i = d_i + \alpha \cdot (d_j - d_i) + (d_k - d_i)$, $d'_j = d_j + \alpha \cdot (d_k - d_j)$, and $d'_k = d_k$. In general, we arrive at the following definition.

Definition 3. *Let n be the number of students, and let $\alpha > 0$ be a real number. For each state of knowledge $d = (d_1, \ldots, d_n)$ and for each subdivision into groups $G_1, \ldots, G_m$, the resulting state of knowledge $d' = (d'_1, \ldots, d'_n)$ is defined as follows: for every $k = 1, \ldots, m$ and for every $i \in G_k$, we have*

$$d'_i = d_i + \alpha \cdot \sum_{j \in G_k, d_j > d_j} (d_j - d_i).$$

Dynamic groups. Subdivision into groups varies. Our objective is to find the subdivision which, at this moment of time, leads to the best gain. From this viewpoint, we need to find groups that work for a forthcoming short period of time, during which the change in grades – proportional to the coefficient α – is small. After this brief interaction, we can again gauge the student's knowledge and, of needed, change the subdivision into groups to reflect what students learned. From this viewpoint, it is sufficient to consider small positive values α.

Possible objective functions. Our goal is to find a subdivision into groups for which the overall degree of knowledge is optimal. This optimal subdivision depends on how we gauge the overall degree of knowledge. In this paper, we will consider three possible criteria (see, e.g., [7]):

- first, we will consider the *average grade* $a \stackrel{\text{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^{n} d_i$;
- another reasonable criterion is maximizing retention, i.e., minimizing the number of students who failed the course; in this case, most attention is paid to students who are at the largest risk of failing, i.e., to students with the smallest possible degree of knowledge; the better the knowledge of this worst performing student, the smaller the risk of failing; thus, from this viewpoint, we should maximize the *worst grade* $w \stackrel{\text{def}}{=} \min_{i=1,\ldots,n} d_i$;
- many high schools brag about the number of their graduates who get into Ivy League colleges; from this viewpoint, most attention is paid to the best students; from this viewpoint, we should maximize the *best grade* $b \stackrel{\text{def}}{=} \max_{i=1,\ldots,n} d_i$.

We will consider all three optimality criteria – and their combinations.

III. STRAIGHTFORWARD RESULTS

Let us consider the situation when we have $n = g \cdot m$ students, we know their degree of knowledge $d_1, \ldots, d_n$, and we want to subdivide these students into m subgroups of g students so as to maximize each of the three objective functions. Let us start with the simplest case $g = 2$, when we divide students into pairs.

Proposition 1. *To maximize the average grade a, we divide the students into pairs as follows:*

- *we sort the students by their knowledge, so that*

$$d_1 \leq d_2 \leq \ldots \leq d_n,$$

- *in each pair, we match one student from the lower half*

$$L_0 \stackrel{\text{def}}{=} \{d_1, d_2, \ldots, d_{n/2}\}$$

with one student from the upper half

$$L_1 \stackrel{\text{def}}{=} \{d_{(n/2)+1}, \ldots \leq d_n\}.$$

Comment. For reader's convenience, all the proofs are placed in the special Proofs section.

Proposition 2. *To maximize the worst grade w, we divide the students into pairs as follows:*

- *we sort the students by their knowledge, so that*

$$d_1 \leq d_2 \leq \ldots \leq d_n,$$

- *we pair the worst-performing student (corresponding to d_1) with the best-performing student (corresponding to d_n), and,*
- *if there are other students with $d_i = d_1$, we match them with d_{n-1}, d_{n-2}, etc.*

Other students can be paired arbitrarily.

If we try to optimize the best grade, subdivision is useless: no matter how we subdivide the students, in this model, the best grade does not change; this is true for any group size $g \geq 2$:

Proposition 3. *No subdivision will improve the best grade b.*

For average-grade and worst-grade optimization, similar results hold for groups of general size $g \geq 2$:

Proposition 4. *For every $g \geq 2$, to maximize the average grade a, we divide the students into groups as follows:*

- *we sort the students by their knowledge, so that*

$$d_1 \leq d_2 \leq \ldots \leq d_n,$$

- *based on this sorting, divide the students into g sets:*

$$L_0 = \{d_1, d_2, \ldots, d_{n/g}\}; \ldots$$

$$L_k = \{d_{k \cdot (n/g)+1}, \ldots, d_{(k+1) \cdot (n/g)}\}, \ldots,$$

$$L_{g-1} = \{d_{(g-1) \cdot (n/g)+1}, \ldots, d_n\};$$

- *in each group, we pick one student from each of g sets L_0, L_1, ..., L_{g-1}.*

Editorial comment. For reader's convenience, all the proofs are placed in the special Proofs section.

Comment. If we measure the students' performance accurately enough, then the degrees of knowledge of different students are different. In this case, we have the following result.

Proposition 5. *If all students' degree d_i are different, then, to maximize the worst grade w, we divide the students into groups as follows:*

- *we sort the students by their knowledge, so that*

$$d_1 \leq d_2 \leq \ldots \leq d_n,$$

- *we combine the worst-performing student (corresponding to d_1) with $g-1$ best-performing students (corresponding to d_n, d_{n-1}, ..., d_{n-2}).*

Other students can be grouped arbitrarily.

Comment. The proposition also holds if several students have the same degree – as long as there is only one worst-performing student. If there are several equally low-performing students $d_1 = d_2 = \ldots = d_s$, then we need to divide $(g-1) \cdot s$ top-performing students into s subgroups of $g-1$ so as to maximize the minimum of the overall grade within each subgroups. Then, each of the low-performing students is matched with one of these subgroups.

IV. MORE NUANCED OPTIMALITY CRITERIA

Discussion. When we use the above optimality criteria such as average grade, worst grade, or best grade, we end up with several different subdivisions which lead to the same optimal value of the selected criterion. We can use this non-uniqueness to optimize something else. In other words, instead of the original optimality criterion, we consider a *lexicographic* combination of these criteria. For example, we say that one subdivision is better than another one if either its average grade a is better ($a > a'$), or they have the same average grade $a = a'$ but the first subdivision has a better worst grade $w > w'$. Here are some related results.

Proposition 6. *Let us assume that we perform the following optimization:*

- *first, we optimize the average grade;*
- *if there are several subdivisions for which the average grade is optimal, then, among all the subdivisions, we select the one for which the worst grade is the largest;*
- *if we have several subdivisions with the largest worst grade, we select the one with the largest second worst grade,*
- *etc.*

Then, for $g = 2$, the following subdivision is optimal:

- *we sort the students by their knowledge, so that*

$$d_1 \leq d_2 \leq \ldots \leq d_n,$$

- *we then match d_1 with d_n, d_2 with d_{n-1}, and, in general, d_k with d_{n+1-k}.*

Comment. For group size $g = 2$, we get the exact same subdivision if we first maximize w, then a, and/or then maximize the second-worst grade, etc.

Proposition 7. *Let us assume that we perform the following optimization:*

- *first, we optimize the average grade;*
- *if there are several subdivisions for which the average grade is optimal, then, among all the subdivisions, we select the one for which the worst grade is the largest;*
- *if we have several subdivisions with the largest worst grade, we select the one with the largest second worst grade,*
- *etc.*

Then, for $g \geq 2$, if all the degrees are different, the following subdivision is optimal:

- *we sort the students by their knowledge, so that*

$$d_1 < d_2 < \ldots < d_n,$$

- *based on this sorting, divide the students into g sets:*

$$L_0 = \{d_1, d_2, \ldots, d_{n/g}\}; \ldots$$

$$L_k = \{d_{k \cdot (n/g)+1}, \ldots, d_{(k+1) \cdot (n/g)}\}, \ldots,$$

$$L_{g-1} = \{d_{(g-1) \cdot (n/g)+1}, \ldots, d_n\};$$

- *we match the smallest value $d_1 \in L_0$ with the largest values from $L_1, \ldots, L_{g-1}$,*
- *we match the second smallest value $d_2 \in L_0$ with the second largest values from $L_1, \ldots, L_{g-1}$,*
- *in general, we match $d_i \in L_0$ with the values $d_{(k+1) \cdot (n/g)+1-k} \in L_k$ for $k = 1, \ldots, g-1$.*

V. A MORE NUANCED MODEL

Main idea. In the above analysis, we used a simplified model in which only the weaker students, with $d_i < d_j$, benefit from the groupwork. In reality, stronger students, with $d_j > d_i$, benefit too: when they explain the material to the weaker students, they reinforce their knowledge, and they may see the gaps in their knowledge that they did not see earlier. The larger the difference $d_j - d_i$, the more the stronger student needs to explain and thus, the more this stronger student reinforces his or her knowledge. It is therefore reasonable to assume that the resulting increase in knowledge is proportional to the difference $d_j - d_i$, with a different coefficient $\beta > 0$. Thus, we arrive at the following definition:

Definition 4. *Let n be the number of students, and let $\alpha > 0$ and $\beta > 0$ be real numbers. For each state of knowledge $d = (d_1, \ldots, d_n)$ and for each subdivision into groups $G_1, \ldots, G_m$, the* resulting state of knowledge $d' = (d'_1, \ldots, d'_n)$ *is defined as follows: for every $k = 1, \ldots, m$ and for every $i \in G_k$, we have*

$$d'_i = d_i + \alpha \cdot \sum_{j \in G_k, d_j > d_j} (d_j - d_i) + \beta \cdot \sum_{j \in G_k, d_i > d_j} (d_i - d_j).$$

It turns out that if we maximize either the average grade or the worst grade, then the optimal subdivisions are exactly the same as for the previously used (less nuanced) model:

Proposition 8. *In the model described by Definition 4, to maximize the average grade a, we divide the students into pairs as follows:*

- *we sort the students by their knowledge, so that*

$$d_1 \le d_2 \le \ldots \le d_n,$$

- *in each pair, we match one student from the lower half*

$$L_0 \overset{\text{def}}{=} \{d_1, d_2, \ldots, d_{n/2}\}$$

with one student from the upper half

$$L_1 \overset{\text{def}}{=} \{d_{(n/2)+1}, \ldots \le d_n\}.$$

Proposition 9. *In the model described by Definition 4, for every $g \ge 2$, to maximize the average grade a, we divide the students into groups as follows:*

- *we sort the students by their knowledge, so that*

$$d_1 \le d_2 \le \ldots \le d_n,$$

- *based on this sorting, divide the students into g sets:*

$$L_0 = \{d_1, d_2, \ldots, d_{n/g}\}; \ldots$$

$$L_k = \{d_{k \cdot (n/g)+1}, \ldots, d_{(k+1) \cdot (n/g)}\}, \ldots,$$

$$L_{g-1} = \{d_{(g-1) \cdot (n/g)+1}, \ldots, d_n\};$$

- *in each group, we pick one student from each of g sets L_0, L_1, ..., L_{g-1}.*

Proposition 10. *In the model described by Definition 4, to maximize the worst grade w, we divide the students into pairs as follows:*

- *we sort the students by their knowledge, so that*

$$d_1 \le d_2 \le \ldots \le d_n,$$

- *we pair the worst-performing student (corresponding to d_1) with the best-performing student (corresponding to d_n), and,*
- *if there are other students with $d_i = d_1$, we match them with d_{n-1}, d_{n-2}, etc.*

Other students can be paired arbitrarily.

Proposition 11. *In the model described by Definition 4, if all students' degree d_i are different, then, to maximize the worst grade w, we divide the students into groups as follows:*

- *we sort the students by their knowledge, so that*

$$d_1 \le d_2 \le \ldots \le d_n,$$

- *we combine the worst-performing student (corresponding to d_1) with $g-1$ best-performing students (corresponding to d_n, d_{n-1}, ..., d_{n-2}).*

Other students can be grouped arbitrarily.

Proposition 12. *Let us assume that we perform the following optimization:*

- *first, we optimize the average grade;*

- *if there are several subdivisions for which the average grade is optimal, then, among all the subdivisions, we select the one for which the worst grade is the largest;*
- *if we have several subdivisions with the largest worst grade, we select the one with the largest second worst grade,*
- *etc.*

Then, in the model described by Definition 4, for $g = 2$, the following subdivision is optimal:

- *we sort the students by their knowledge, so that*

$$d_1 \le d_2 \le \ldots \le d_n,$$

- *we then match d_1 with d_n, d_2 with d_{n-1}, and, in general, d_k with d_{n+1-k}.*

Proposition 13. *Let us assume that we perform the following optimization:*

- *first, we optimize the average grade;*
- *if there are several subdivisions for which the average grade is optimal, then, among all the subdivisions, we select the one for which the worst grade is the largest;*
- *if we have several subdivisions with the largest worst grade, we select the one with the largest second worst grade,*
- *etc.*

Then, in the model described by Definition 4, for $g \ge 2$, if all the degrees are different, the following subdivision is optimal:

- *we sort the students by their knowledge, so that*

$$d_1 < d_2 < \ldots < d_n,$$

- *based on this sorting, divide the students into g sets:*

$$L_0 = \{d_1, d_2, \ldots, d_{n/g}\}; \ldots$$

$$L_k = \{d_{k \cdot (n/g)+1}, \ldots, d_{(k+1) \cdot (n/g)}\}, \ldots,$$

$$L_{g-1} = \{d_{(g-1) \cdot (n/g)+1}, \ldots, d_n\};$$

- *we match the smallest value $d_1 \in L_0$ with the largest values from $L_1, \ldots, L_{g-1}$,*
- *we match the second smallest value $d_2 \in L_0$ with the second largest values from $L_1, \ldots, L_{g-1}$,*
- *in general, we match $d_i \in L_0$ with the values $d_{(k+1) \cdot (n/g)+1-k} \in L_k$ for $k = 1, \ldots, g-1$.*

Interestingly, now we can optimize the best grade b. For $g = 2$, the result is the same as for optimizing worst grades, but for $g > 2$, the result is different:

Proposition 14. *In the model described by Definition 4, to maximize the best grade w, we divide the students into pairs as follows:*

- *we sort the students by their knowledge, so that*

$$d_1 \le d_2 \le \ldots \le d_n,$$

- *we pair the best-performing student (corresponding to d_n) with the worst-performing student (corresponding to d_1), and,*

- *if there are other students with $d_i = d_n$, we match them with d_2, d_3, etc.*

Other students can be paired arbitrarily.

Proposition 15. *In the model described by Definition 4, if all students' degree d_i are different, then, to maximize the worst grade w, we divide the students into groups as follows:*

- *we sort the students by their knowledge, so that*

$$d_1 \le d_2 \le \ldots \le d_n,$$

- *we combine the best-performing student (corresponding to d_n) with $g - 1$ worst-performing students (corresponding to d_1, d_2, ..., d_{g-1}).*

Other students can be grouped arbitrarily.

VI. CASE OF UNCERTAINTY

In practice, we rarely know the exact values of d_i, we only know approximately values $\widetilde{d_i}$ – and, e.g., we know the accuracy Δ of these estimates, i.e., we know that $d_i \in [\widetilde{d_i} - \Delta, \widetilde{x_i} + \Delta]$. In this case, we do not know the exact gain, so it is reasonable to select a "maximin" subdivision, i.e., a subdivision for which the guaranteed (= worst-case) gain is the largest. One can prove that the subdivisions obtained by applying the above algorithms to the approximate value $\widetilde{d_i}$ are optimal in this minimax sense as well.

VII. PROOFS

Proof of Proposition 1.

$1°$. First, we note that maximizing the average grade is equivalent to maximizing the sum $n \cdot a = \sum\limits_{i=1}^{n} g_i'$ of the new grades, which is, in turn, equivalent to maximizing the overall gain $\sum\limits_{i=1}^{n} g_i' - \sum\limits_{i=1}^{n} g_i = \sum\limits_{i=1}^{n} (g_i' - g_i)$.

$2°$. Let us take the optimal subdivision, and show that it has the form described in the formulation of Proposition 1.

Indeed, in each pair, with degrees $d_i \le d_j$, we have a weaker student i and a stronger student j. Let us prove that the optimal subdivision into groups, each stronger student is stronger (or of the same strength) than each weaker student. In other words, if we have two pairs $d_i \le d_j$ and $d_{i'} \le d_{j'}$, then $d_i \le d_{j'}$. We will prove this by contradiction. Let us assume that $d_i > d_{j'}$. Let us then swap the i-th and the j'-th students, i.e., instead of the original pairs (i, j) and (i', j'), let us consider two new pairs (i, j') and (i', j). The corresponding two terms in the overall gain are changed from $\alpha \cdot (d_j + d_{j'} - d_i - d_{i'})$ to $\alpha \cdot (d_j - d_{j'} + d_i - d_{i'})$. The difference between the two expressions is equal to $2\alpha \cdot (d_i - d_{j'})$. Since we assumed that $d_i > d_{j'}$, this difference is positive, which means that the above swap increases the overall gain. The possibility of such an increase contradicts to the fact that we have selected the subdivision for which the overall gain is already the largest possible. This contradiction shows that our assumption $d_i > d_{j'}$ is wrong, and thus, $d_i \le d_{j'}$.

Since every weaker-of-pair student is weaker than every stronger-of-pair student, all weaker-of-pair students form the bottom of the ordering of the degrees d_i, while all the stronger-of-pair students form the top of this ordering – exactly as the formulation of Proposition 1 suggests.

$3°$. To complete the proof, we need to prove that every subdivision satisfying the condition of Proposition 1 leads to the optimal average grade. Indeed, we know that one optimal subdivision satisfies this condition. One can check that for each such subdivision, the overall gain is equal to $\sum\limits_{i \in L_1} d_i - \sum\limits_{j \in L_0} d_j$, where L_1 is the set of all the indices i from the upper half, and L_0 is the set of all the indices from the lower half. Thus, the overall gain for all such subdivisions is the same – and it is therefore exactly equal to the gain corresponding to the optimal subdivision. So, all subdivisions satisfying the condition of Proposition 1 indeed lead to the optimal average grade.

The proposition is proven.

Proof of Proposition 2. The worst grade w is the grade of the worst-performing student. For each of these students i, the larger the difference $d_j - d_i$, the more their grade will increase. So, if there is only one such student, we pair him or her with the best-performing one. If there are other students with the same worst grade, we pair them with the best-performing among the not-yet-paired students, etc.

Proof of Proposition 3. One can easily see that in the above model, in each group, the largest grade does not change. So, the best grade b, which is the largest of the group-wide best grades – also does not change.

Proof of Proposition 4.

$1°$. Let us first prove that an optimal group subdivision satisfies the property described in the formulation of Proposition 4. Indeed, let us start with an optimal subdivision. Within each group, we can sort its g students in the increasing order of their grades; thus, every student gets assigned a rank in the corresponding group. Now, we can prove that for every two ranks $r < r'$, a grade of a student of rank r is always less than or equal to the grade of a student of rank r' – even when they are from different groups. Similarly to the proof of Proposition 1, this can be proven by contradiction: if a grade d_i of a student of rank r is larger than the grade d_j of a student of rank r', then we can swap these two students and improve the overall gain.

$2°$. To complete the proof, we must show that for all subdivisions that satisfy the condition from the formulation of Proposition 4, the gain is the same. Indeed, the overall gain is equal to the sum of gains obtained in each group. Let us therefore calculate the gain in each group.

For $g = 3$, once we have $d_i \le d_j \le d_k$, the gain is equal to α times the sum

$$(d_k - d_j) + (d_k - d_i) + (d_j - d_i) = 2d_k - 2d_i.$$

Thus, the overall gain is equal to $2 \sum\limits_{i \in L_2} d_i - 2 \sum\limits_{i \in L_0} d_i$, and hence, it indeed does not depend on the subdivision – as long

as the subdivision satisfies the condition from the formulation of Proposition 4.

For $g = 4$, once we have $d_i \leq d_j \leq d_k \leq d_l$, the gain is equal to

$$(d_l-d_k)+(d_l-d_j)+(d_l-d_i)+(d_k-d_j)+(d_k-d_i)+(d_j-d_i) =$$
$$3d_l + d_k - d_j - 3d_i.$$

Thus, the overall gain is equal to

$$3\sum_{i \in L_3} d_i + \sum_{i \in L_2} d_i - \sum_{i \in L_1} d_i - 3\sum_{i \in L_0} d_i.$$

For a general group size g, one can prove, by induction, that once we have $d_{i_1} \leq \ldots \leq d_{i_g}$, then the gain of this group is equal to

$$(g-1) \cdot d_{i_g} + (g-3) \cdot d_{i_{g-1}} + \ldots +$$
$$(2k - g - 1) \cdot d_{i_k} + \ldots$$
$$-(g-3) \cdot d_{i_2} - (g-1) \cdot d_{i_1}.$$

Thus, the overall gain is equal to

$$(g-1) \cdot \sum_{i \in L_{g-1}} d_i + (g-3) \cdot \sum_{i \in L_{g-2}} d_i + \ldots +$$
$$(2k - g + 1) \cdot \sum_{i \in L_k} d_i + \ldots$$
$$-(g-3) \cdot \sum_{i \in L_1} d_i - (g-1) \cdot \sum_{i \in L_0} d_i.$$

So, the sum does not depend on the subdivision – as long as the subdivision satisfies the condition from the formulation of Proposition 4.

The statement is proven, and so is the proposition.

Proof of Proposition 5 (and of the Comment after Proposition 5) is similar to the proof of Proposition 2.

Proof of Proposition 6 (and of the Comment after Proposition 6). According to Proposition 1, we select between subdivisions in which in each pair, one element is taken from the lower half L_0 and the other element is taken from the upper half L_1. Similarly to the proof of Proposition 2, among such subdivisions, we select a one for which the student with the smallest grade d_1 is matched with the student with the largest grade d_n. To maximize the smallest of the remaining grades, we need to match the smallest of the remaining grades d_2 with the largest of the remaining grades d_{n-1}, etc. The proposition is proven.

Proof of Proposition 7. According to Proposition 4, we select between subdivisions in which in each group, one element is taken from the lower set L_0 and one element is taken from upper-level sets $L_1, \ldots, L_{g-1}$. Similarly to the proof of Proposition 2, among such subdivisions, we select a one for which the student with the smallest grade d_1 is matched with the students with the largest grades from the sets $L_1, \ldots, L_{g-1}$. To maximize the smallest of the remaining grades, we need to match the smallest of the remaining grades d_2 with the largest of the remaining grades, etc. The proposition is proven.

Proof of Propositions 8–13. We have already mentioned, in the proof of Proposition 1, that optimizing the average grade is equivalent to optimizing the overall gain.

In model described by Definition 4, the gain coming from interaction between the i-th and the j-the students with $d_i < d_j$ is equal to

$$\alpha \cdot (d_j - d_i) + \beta \cdot (d_j - d_i) = \alpha' \cdot (d_j - d_i),$$

where $\alpha' \stackrel{\text{def}}{=} \alpha + \beta$. Thus, in the new model, the overall gain is described by the same formula as in the old model, but with a new coefficient α' instead of the original coefficient α. Since the formula for the overall gain is the same, the optimal subdivisions are also the same.

With respect to optimizing worst grades, the relation with the original model is even easier: the new formula does not change the grades of the worst-performing students. The propositions are thus proven.

Proof of Propositions 14–15 is similar to the proof of Proposition 2.

REFERENCES

[1] D. Bhattacharya, "Inferring Optimal Peer Assignment from Experimental Data," *J. of the Amer. Stat. Association*, 2009, Vol. 104, pp. 486–500.

[2] S. E. Carrell, R. L. Fullerton, and J. E. West, "Does Your Cohort Matter? Estimating Peer Effects in College Achievement," *J. of Labor Economics*, 2009, Vol. 27, No. 3, pp. 439–464.

[3] S. E. Carrell, F. V. Malmstrom, and J. E. West, "Peer Effects in Academic Cheating," *J. of Human Resources*, 2008, Vol. 43, No. 1, pp. 173–207.

[4] S. E. Carrell, B. I. Sacerdote, and J. E. West, *From Natural Variation to Optimal Policy? An Unsuccessful Experiment in Using Peer Effects Estimates to Improve Student Outcomes*, UC Davis Working Paper, 2012, http://www.econ.ucdavis.edu/faculty/scarrell/sortexp.pdf

[5] G. Foster, "It's not your peers, and it's not your friends: Some progress toward understanding the educational peer effect mechanism," *J. of Public Economics*, 2006, Vol. 90, No. 8-9, pp. 1455–1475.

[6] B. S. Graham, G. W. Imbens, and G. Ridder, *Complementarity and Aggregate Implications of Assortative Matching: A Nonparametric Analysis*, Working Paper 14860, National Bureau of Economic Research, 2009.

[7] O. Kosheleva and V. Kreinovich, "Towards Optimal Effort Distribution in Process Design under Uncertainty, with Application to Education", *Int'l J. of Reliability and Safety*, 2012, Vol. 6, No. 1–3, pp. 148–166.

[8] D. S. Lyle, "Estimating and Interpreting Peer and Role Model Effects from Randomly Assigned Social Groups at West Point," *Review of Economics and Statistics*, 2007, Vol. 89, No. 2, pp. 289–299.

[9] B. I. Sacerdote, "Peer Effects with Random Assignment: Results for Dartmouth Roommates," *Quarterly J.of Economics*, 2001, Vol. 116, No. 2, pp. 681–704.

[10] R. Stinebrickner and T. R. Stinebrickner, "What can be learned about peer effects using college reoomates? Evidence from new survey data and students form disadvantaged backgrounds," *J. Public Economics*, 2006, Vol. 90, No. 8–9, pp. 1435–1454.

[11] D. J. Zimmerman, "Peer Effects in Academic Outcomes: Evidence From a Natural Experiment," *The Review of Economics and Statistics*, 2003, Vol. 85, No. 1, pp. 9–23.