

---

---

# Есть ли жизнь в математике? Различия между заданиями PISA, TIMSS и ОГЭ

---

---

Капуза Анастасия - научный сотрудник  
Бучко Павел - стажер-исследователь  
Международная лаборатория оценки практик и инноваций в образовании,  
Институт образования НИУ ВШЭ

# Международные сравнительные исследования

## PISA

Programme for International Student  
Assessment

Международная программа по оценке  
образовательных достижений  
учащихся, OECD

Каждые три года начиная с 2000

Прим. 75 стран

Математическая, естественнонаучная и  
читательская грамотности

15-летние учащиеся

## TIMSS

Trends in Mathematics and Science Study

Тренды в математическом и  
естественнонаучном образовании, IEA

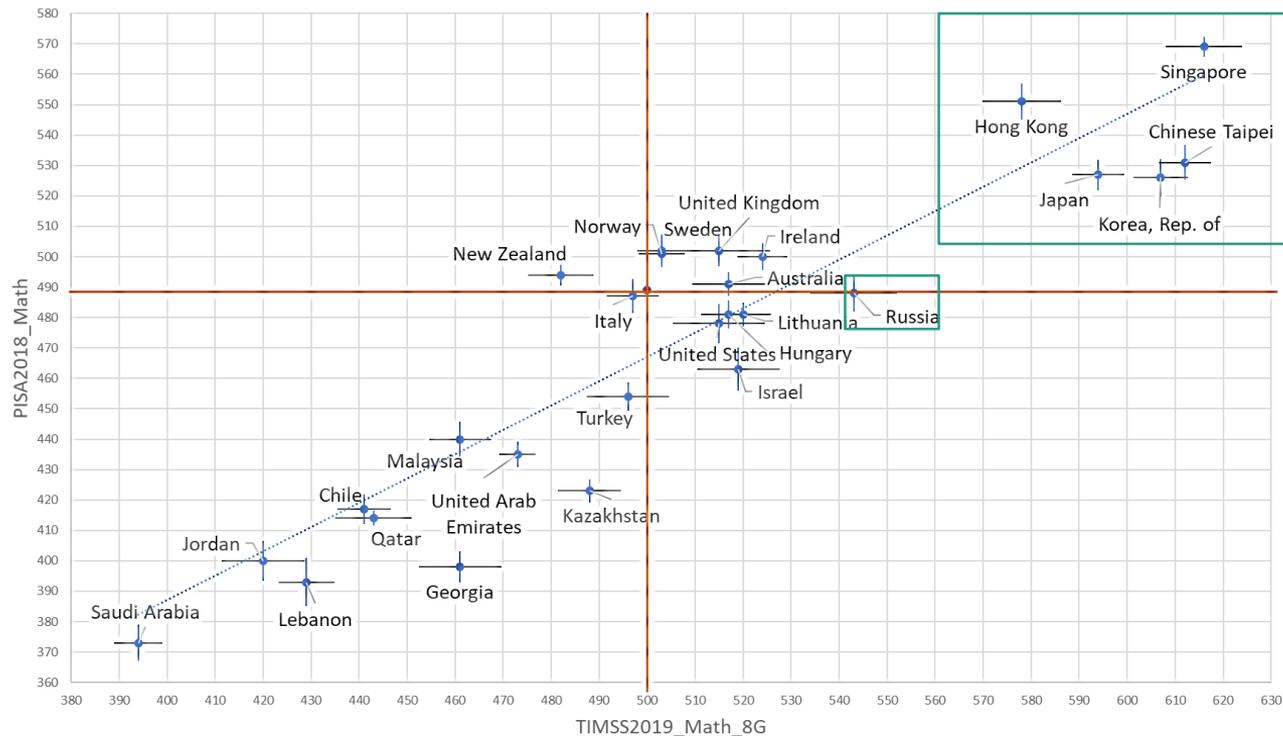
Каждые четыре года начиная с 1995

Прим. 70 стран

Тестирование по математике и  
естественнонаучным предметам

Учащиеся 4 и 8 классов

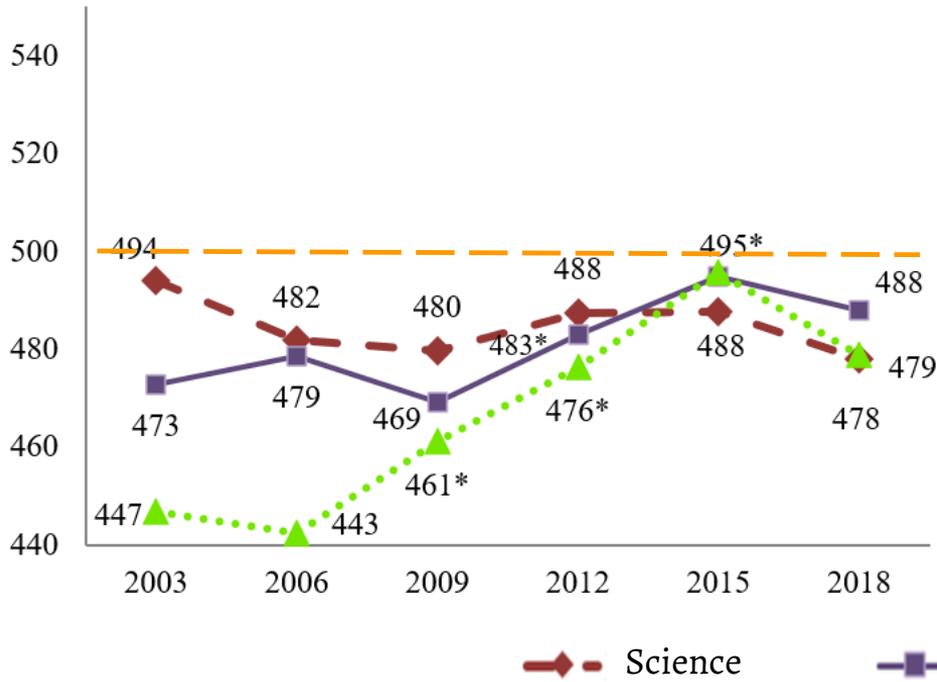
# Большинство стран показывают более высокие результаты TIMSS



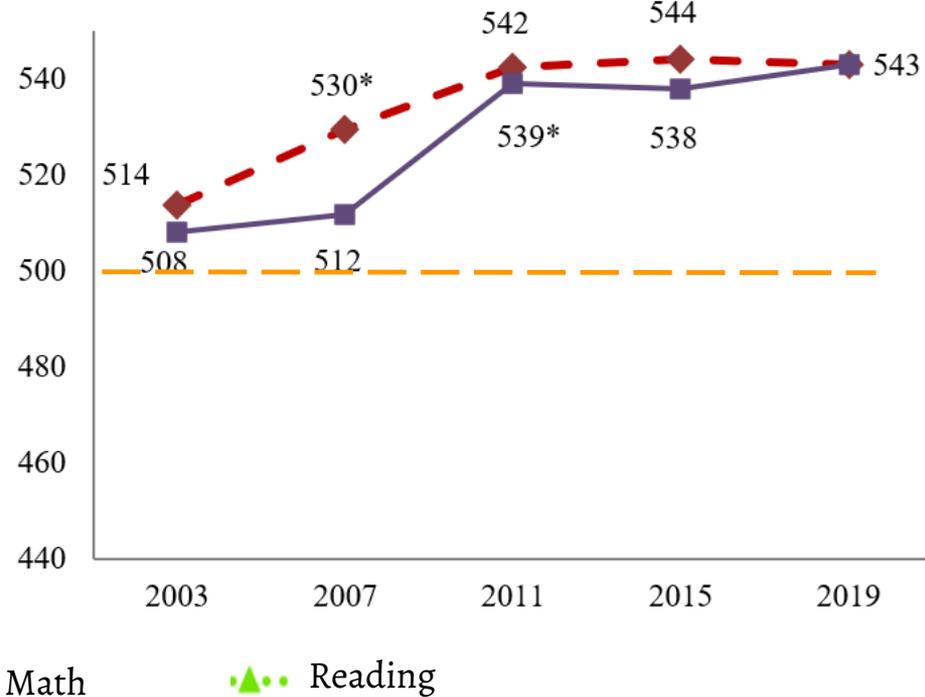
“Азиатские тигры”

# Россия – яркий пример этой тенденции

## PISA



## TIMSS



**Национальный проект “Образование”  
особенное внимание обращает на результаты  
PISA как на индикатор качество образования**

**Нужно понимать, какими путями можно  
достичь повышения балла PISA (сейчас -  
развития тех навыков, которые измеряет PISA)**

# ОГЭ – самый важный экзамен для основного общего образования

ОГЭ сдают все после 9 класса

ОГЭ по математике и русскому языку обязателен

**!** PISA, ОГЭ и TIMSS предназначены для учащихся примерно одного возраста: TIMSS для 8 класса, PISA для пятнадцатилетних, ОГЭ для 9 класса, что соответствует возрасту 15-16 лет

# Почему PISA сложнее, чем TIMSS? Предположения

## 1. Literacy (PISA) vs achievements (TIMSS)

В PISA используются задачи, заявленные как ориентированные на бытовой контекст (real life problems), требующие в большей степени размышлений, чем применения знаний как в TIMSS (Rindermann, & Baumeister, 2015). Считается, что такой формат непривычен для российских школьников (Тюменева, 2015).

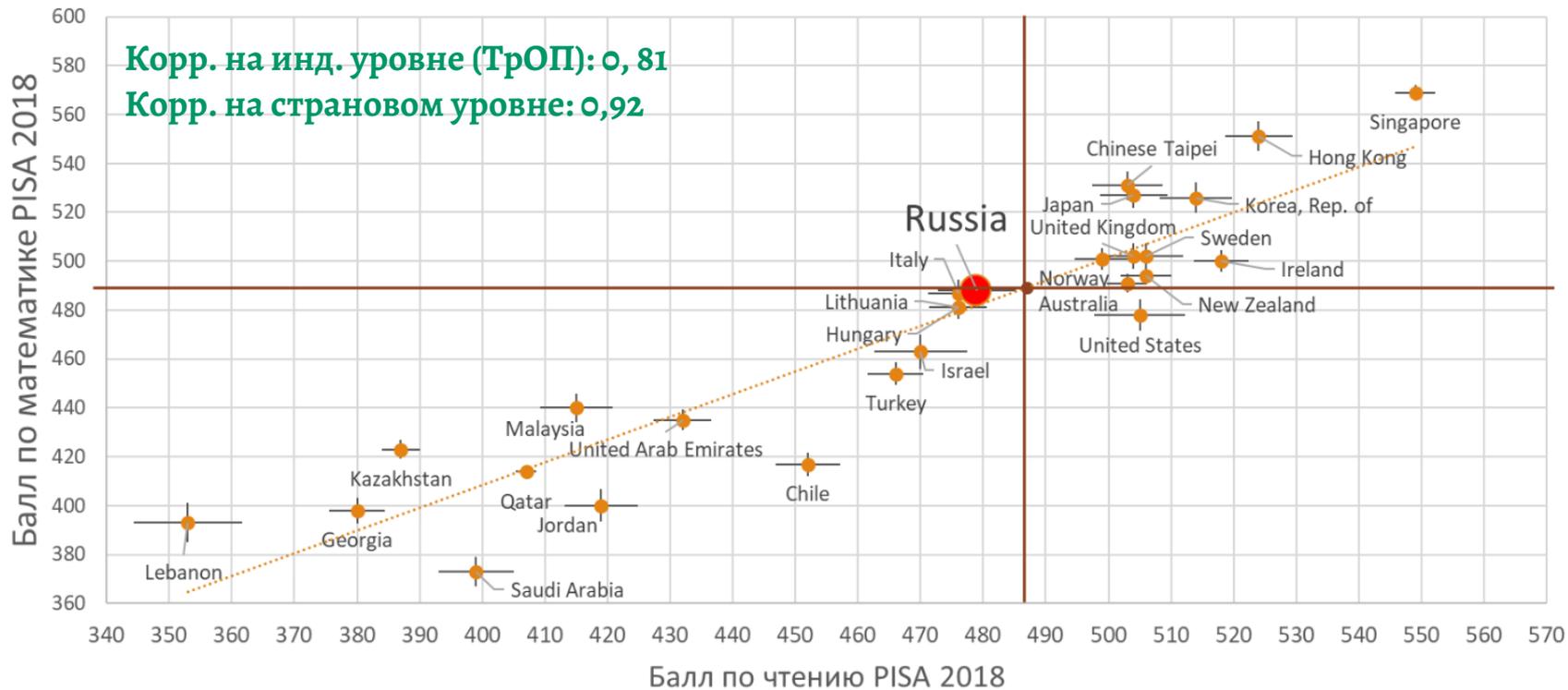
Т.е. российских учеников просто не готовят к такому типу тестов (Тюменева и др., 2015)

## 2. Для решения математических задач PISA нужны сильные навыки чтения

Нужно подавить нерелевантную информацию, семантические искажения, заученные процедуры и поверхностные реакции, вызванные формулировкой (напр. Duque et al, 2021)

“Calculate first, think later” (Siegler & Jenkins, 2014)

# Есть сильная связь между результатами PISA по математике и чтению



# PISA. ПРИМЕРЫ ЗАДАНИЙ

Пешеходная тропа Готемба на гору Фудзи имеет длину около 9 километров.

Пешеходам нужно вернуться после 18 км прогулки к 20 часам.

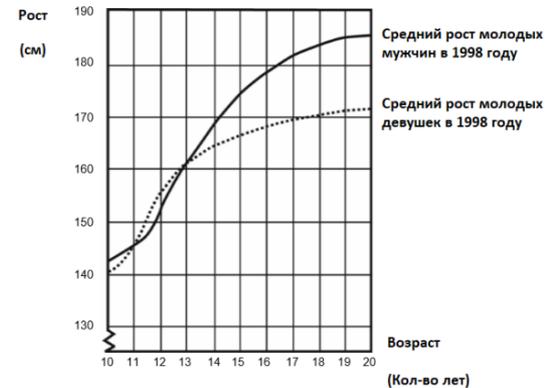
Тоши прикинул, что он может подняться на гору со средней скоростью 1,5 км/ч и спуститься со скоростью в два раза больше этой. При движении с этими скоростями остаётся время на то, чтобы поесть и отдохнуть.

Используя скорости, установленные Тоши, определите самое позднее время, когда Тоши может начать свой подъём, чтобы он мог вернуться к 20 ч.

## РОСТ

### МОЛОДЕЖЬ СТАНОВИТСЯ ВЫШЕ

Средний рост молодых мужчин и молодых девушек в Нидерландах 1998 года показан на графике.



### Вопрос 1: РОСТ

С 1980 года средний рост 20-летних девушек увеличился на 2,3 см (достиг 170,6 см). Какой был рост 20-летних девушек в 1980 году?

# TIMSS. ПРИМЕРЫ ЗАДАНИЙ

Какая пара чисел  $(x; y)$  удовлетворяет уравнению  $3x + 4y = 24$ ?

- A  $(0; 8)$
- B  $(3; 4)$
- C  $(4; 3)$
- D  $(6; 0)$

**Верными** или **неверными** являются следующие утверждения относительно любого натурального числа  $n$ ?  
Отметьте один кружок для каждого утверждения.

	Верное	Неверное
$n + 4 = 4 + n$	<input type="checkbox"/>	<input type="checkbox"/>
$n - 5 = 5 - n$	<input type="checkbox"/>	<input type="checkbox"/>
$n \cdot 6 = 6 \cdot n$	<input type="checkbox"/>	<input type="checkbox"/>
$n : 7 = 7 : n$	<input type="checkbox"/>	<input type="checkbox"/>

Класс отправляется на экскурсию в музей. Обед для всего класса стоит  $a$  рублей.

Входной билет для каждого ученика стоит 50 рублей. В классе  $x$  учащихся. Общая стоимость экскурсии  $k$  рублей.

Запишите формулу для вычисления значения  $k$ .

$k =$

У вас есть рецепт для выпечки торта, для которого требуются 2 яйца и 0,3 литра молока. Вы хотите испечь самый большой торт, какой только возможно, и у вас для этого есть 5 яиц.

Сколько литров молока вам потребуется, чтобы испечь самый большой торт, который только возможно?

Ответ:  л



# Для определения степени real life problem мы используем рамку практической задачи (Ларина, 2016)

1. Отсутствие прямого указания на последовательность действий (**новизна формулировок**)

*Нет:* в первом ряду 3 плитки, во втором 5, сколько плиток в 7? / Сложите 2 числа

*Да:* за сколько дней окупится...

1. Перевод с обыденного языка в термины математики (**обыденность семантики**)

*Нет:* реши уравнение: ...

*Да:* за сколько дней окупится новый парус, если судно станет перевозить грузы в 2 раза быстрее

1. Относительная **жесткость структуры**

*Нет:* найди значение ..

*Да:* подумай как ...

# Есть еще два показателя степени real-life, но их сложно оценить

## 4. Ситуационная значимость - от решения что-то может измениться

*Нет:* сколькими способами можно засеять три поля...

*Да:* сколько пузырьков лекарства нужно купить, если в первые 7 дней нужно увеличивать дозу на 10 капель и затем пить его еще неделю?

## 5. Обстоятельства решения

*Нет:* строгое ограничение по времени, без вспомогательных источников

*Да:* возможность пользоваться вспомогательными источниками и обсуждать решение

# Примеры анализа текстов: гендерная репрезентация в учебниках

	Малайзия	Индонезия	Пакистан	Бангладеш
% женских имен в текстах	50,4	42,9	13,6	36,6
% женщин в домашних ролях (готовит и т.д.)	85	63	100	73
% женщин в профессиональных ролях	44	44	8	29

Islam, K. M. M., & Asadullah, M. N. (2018). Gender stereotypes and education: A comparative content analysis of Malaysian, Indonesian, Pakistani and Bangladeshi school textbooks. *PLoS one*, 13(1), e0190807.

# Примеры анализа текстов: учебники русского языка

ЛЮБИТЬ		
РКР	РКН	РКИ
Русский как родной	Русский как неродной	Русский как иностранный
Любить + объект (N)		
березу русскую <sup>10</sup> , Родину, животных, коровушку, курочек, чистоту, труд, песенки, грозу, <b>осень</b>	пирожки, спорт, сказки, <b>зиму</b>	кошку, маму блины, мороженое, сок, черный хлеб, мед, мясо, зеленый чай русскую литературу, старое кино, хорошую погоду, хорошую музыку, цирк, <b>зиму</b>

Лебедева М.Ю. и др. (2020). Лексический состав учебников по русскому языку для носителей и инофонов  
Компьютерная лингвистика и интеллектуальные технологии: по материалам международной конференции «Диалог 2020»

## Цель этого исследования

- оценить различия между тестами PISA, TIMSS и OGE,  
используя рамку определения степени real life

с помощью методов анализа текстов

# Методология

# Материалы

В качестве единицы анализа использовалась одна задача (задание с независимым содержанием и конкретным вопросом).

**PISA** – 79 демо заданий с 2009 по 2021;

**TIMSS** 8 класс – 107 демо заданий с 2018 по 2021;

**OGE** – 270 демо заданий с 2009 по 2021.

Все эти тексты были очищены от стоп-слов (местоимений, предлогов, междометий) и были лемматизированы с помощью пакета MyStem в R Studio ([Segalovich, 2003](#)).

# Используемые метрики

## Новизна формулировок

## Обыденность семантики

## Жесткость структуры

% математических терминов (чем ↑ тем менее практическая)

Топ 20 слов (найди, реши, вычисли - менее практическая)

Количество слов в задании (чем ↓ тем менее практическая)

Процент императивов в задании (чем ↑ тем менее практическая)

Процент прилагательных в задании (чем ↓ тем менее практическая)

Индикаторы сложности текста (чем сложнее, тем менее практическая)

Все показатели сравнивались с помощью ANOVA, а затем применялся метод множественных сравнений Тьюки для проведения попарных сравнений и подсчета уровня значимости.

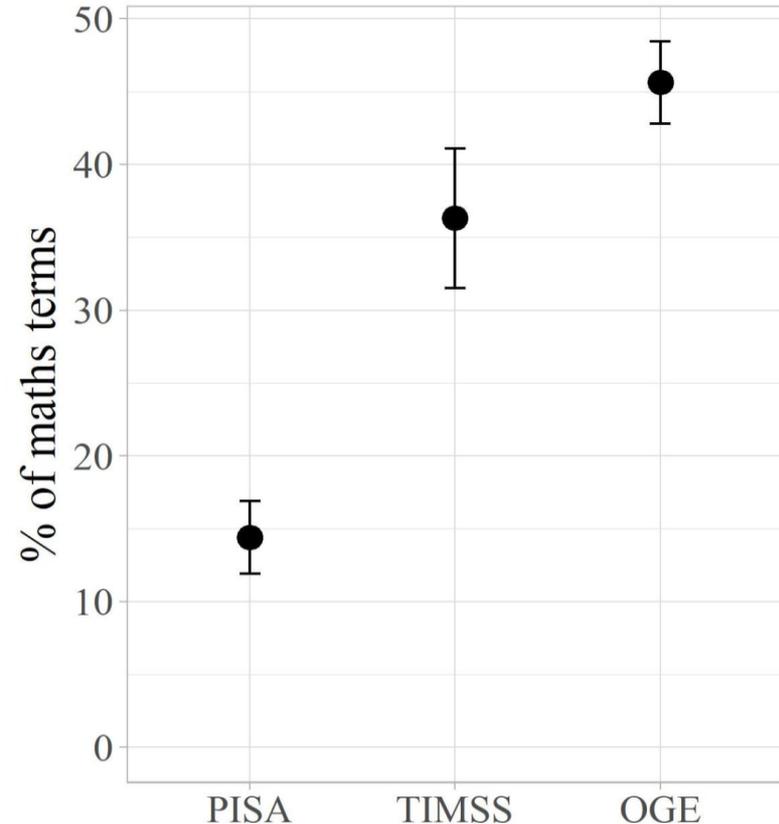
# Результаты

# 1. Новизна формулировок: % терминов

Значимо ниже в PISA, чем в TIMSS и OGE ( $p < 0,01$ )

В среднем задание PISA содержало более 80% повседневного языка

TIMSS и PISA состояли примерно из 40% математических терминов (36,3 и 45,6 соответственно,  $p < 0,1$ )



# 1. Новизна формулировок: топ 20 слов

## PISA

1. количество
2. время
3. год
4. скорость
5. мочь
6. ответ
7. приводить
8. показывать
9. ниже
10. машина
11. средний
12. зед
13. км
14. представлять
15. минута

## TIMSS

1. фигура
2. плитка
3. равный
4. количество
5. число
6. учащийся
7. каков
8. значение
9. красный
10. длина
11. рисунок
12. составлять
13. см
14. выражение
15. показывать

## ОГЭ

1. находить
2. равный
3. треугольник
4. число
5. точка
6. значение
7. график
8. угол
9. выражение
10. изображать
11. отметка
12. рисунок
13. круг
14. учащийся
15. получать

## 2. Обыденность семантики: топ 20 слов

### PISA

1. количество
2. время
3. год
4. скорость
5. мочь
6. ответ
7. приводить
8. показывать
9. ниже
10. машина
11. средний
12. зед
13. км
14. представлять
15. минута

### TIMSS

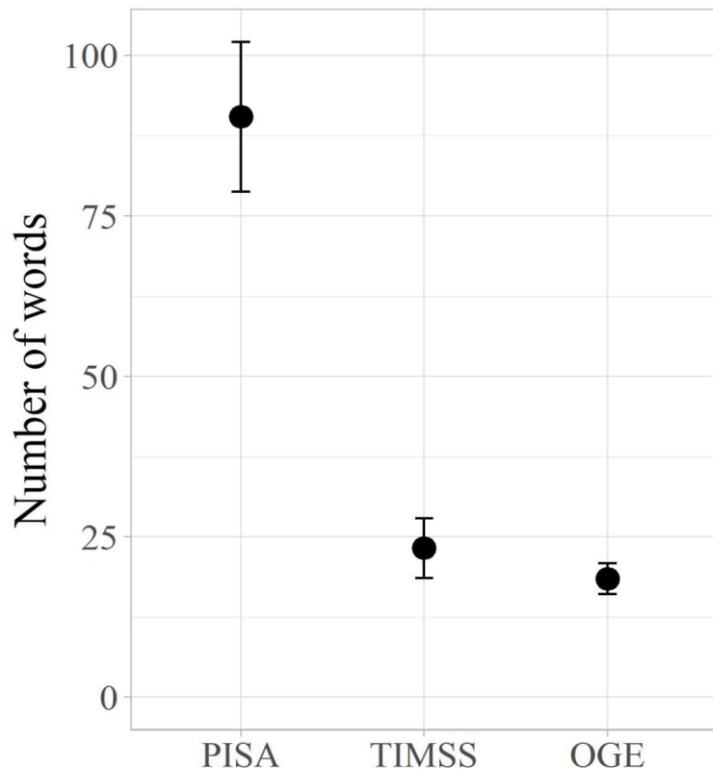
1. фигура
2. плитка
3. равный
4. количество
5. число
6. учащийся
7. каков
8. значение
9. красный
10. длина
11. рисунок
12. составлять
13. см
14. выражение
15. показывать

### ОГЭ

1. находить
2. равный
3. треугольник
4. число
5. точка
6. значение
7. график
8. угол
9. выражение
10. изображать
11. отметка
12. рисунок
13. круг
14. учащийся
15. получать

С точки зрения  
обыденности  
семантики, в  
текстах PISA  
были различные  
слова и глаголы,  
обозначающие  
действие

## 2. Обыденность семантики: количество слов



Среднее количество слов в задании PISA было в четыре раза выше (90,5), чем в других тестах (TIMSS 23,2, OGE 18,5,  $p < 0,01$ )

В то же время TIMSS и ОГЭ значимо не различались ( $p = 0,72$ )

## 2. Обыденность семантики: сложность текста

	PISA	TIMSS	ОГЭ
Flash Index	45.9 [44.0; 47.8]	41.4 [38.7; 44.1]	27.9 [26.2; 29.6]
Flash-Kincaid Index	17.3 [16.9; 17.7]	17.9 [17.4; 18.4]	20.5 [20.2; 20.8]
Gunning Index	6.8 [6.5; 7.1]	6.3 [5.9; 6.7]	8.2 [7.9; 8.5]

PISA и TIMSS было легче читать и понимать, чем ОГЭ ( $p < 0,01$  по всем индексам)

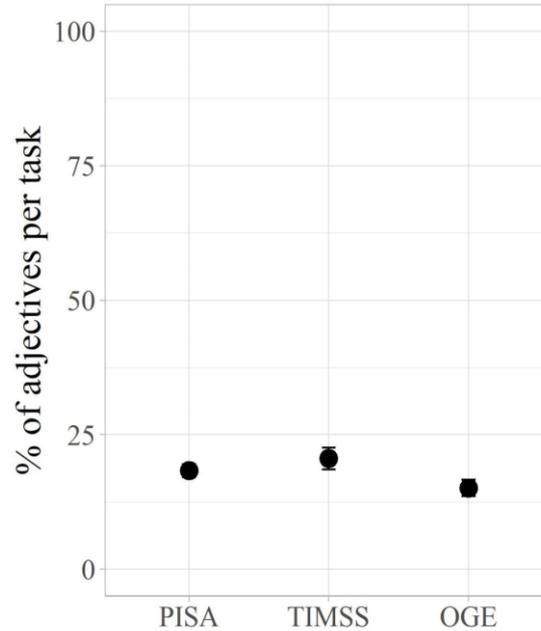
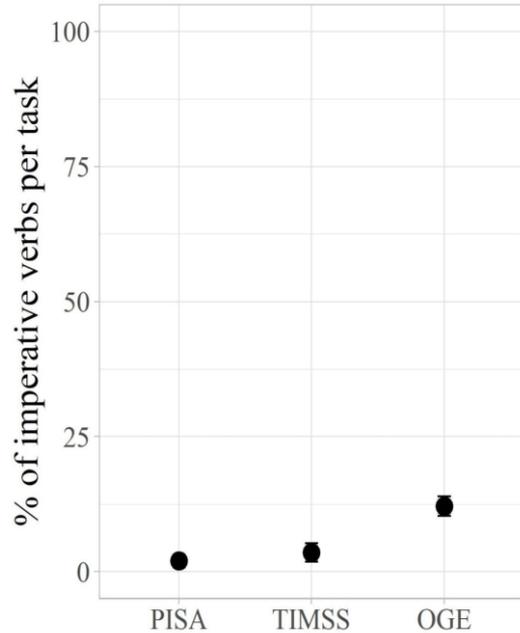
Согласно Индексу Флеша-Кинкейда, все эти тексты считаются сложными научными текстами

Индекс Флэша показал, что все тексты (особенно ОГЭ) были сложными для учащихся 5-11 классов

Индекс Ганнинга определил PISA и TIMSS как слишком легкие для этих студентов

\* Средние и 95% ДИ

## 2. Обыденность семантики: % императивов и прилагательных



% прилагательных показывает описательный характер текста

PISA и TIMSS имели одинаково высокие проценты (18,3% и 20,6%,  $p=0,5$ )

PISA и ОГЭ были схожи (15,1% в ОГЭ,  $p=0,36$ ), что отчасти может быть объяснено тем фактом, что задания ОГЭ содержат гораздо меньше слов и даже одно прилагательное «весит» много

% повелительных глаголов показывает, сколько прямых инструкций представлено в тексте («найди», «реши», «нарисуй» и т.д.)

В ОГЭ был самый высокий процент таких глаголов - 12,1%

### 3. Жесткость структуры

Этот параметр основан на уже описанных показателях: количестве слов, процентном соотношении повелительных глаголов и прилагательных

Каждое десятое слово в ОГЭ призывало к действию, в то время как в других тестах это было одно из 100

Это дает ОГЭ больше возможностей повторить структуру «основная часть + призыв к действию» (топ 20 слов даже показывает, какой – «найди»)

Основываясь на этих выводах, можно предположить, что структура заданий ОГЭ более жесткая и, наоборот, задания PISA имеют менее жесткую структуру

# Выводы и дискуссия

## Кратко результаты

- Разработанная модель измерения и показатели определяют PISA как тест, ориентированный на реальную жизнь, ее можно использовать для анализа других учебных материалов
- Согласно разработанной модели ОГЭ является наименее ориентированы на применение в реальной жизни
- В то же время TIMSS по некоторым параметрам близок к ОГЭ, по другим к PISA

# О чем нам это может говорить

ВАЖНО создавать базу для более сложных навыков через обучение более простым – без тренировок на похожих, повторяющихся задачах, не происходит перехода к следующему уровню

ОДНАКО и другим типам задач с некоторой степенью неопределенности тоже необходимо тренировать, знакомить учащихся с различными форматами

ОГЭ является основополагающим инструментом формирования учебных планов и ориентиром для учителей



# Ограничения исследования

- Для анализа использованы только демо-варианты, возможно «внутри» произошли изменения
- Анализ текста автоматизирован, невозможно учесть все нюансы
- Возможны и другие интерпретации результатов + другие рамки

# Вопросы для обсуждения

Стоит ли в принципе ориентироваться на real-life модель PISA?

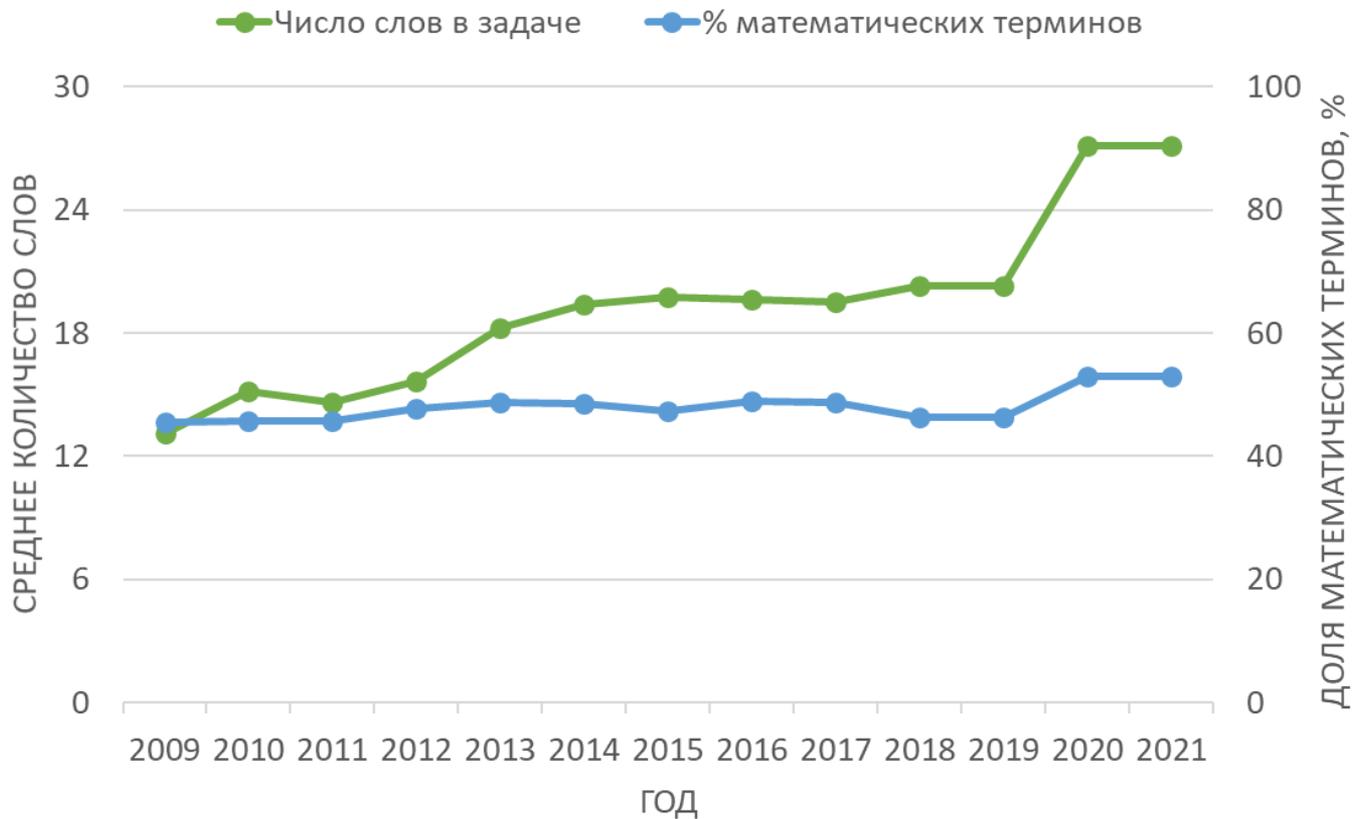
У нее есть и защитники, и критики

Если да, как школьная программа может это сделать?

Планируется ли (или уже) при создании заданий ОГЭ ориентироваться на PISA?

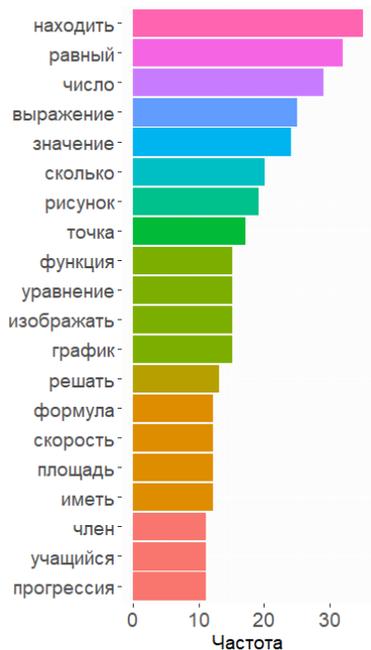
**Спасибо!**

# Динамика по годам ГИА - кажется, становится более real life (но в демо-вариантах много пересекающихся заданий)



# Топ-20 слов ГИА по годам - появляются повседневные объекты

2009-2013



2014-2017



2018-2021

