Are writing questions in math fair?

Lex Konnelly
University of Toronto
a.konnelly@mail.utoronto.ca

Nathan Sanders
University of Toronto
nathan.sanders@utoronto.ca

Jason Siefken
University of Toronto
siefkenj@math.toronto.edu

Pocholo Umbal
University of Toronto
p.umbal@mail.utoronto.ca

Abstract: In this paper, we examine whether a student's language background and other demographic factors have any relationship to their performance on prose questions in math, which we define as questions with open-ended answers containing one or more complete sentences of English. Prose questions stand in contrast to non-prose questions, which are more traditional questions in math courses, requiring an objective answer, such as a number, an equation, a diagram, etc. Performing an exploratory analysis on exam scores for 463 students in a first-year linear algebra course, we use step-down regression to identify significant factors contributing to a student's non-prose tilt: how much better a student performs on non-prose versus prose questions. We find that gender is the only significant factor contributing to a student's non-prose tilt. In particular, no linguistic factors we considered, including whether or not a student was a native English speaker, emerged as significant.

Keywords: equitable assessment, language bias, gender gap

1 Introduction

In a math course, is asking a writing question—one that must be answered in full sentences—fair? What if a large portion of your students are not native speakers of the language of instruction? What if you are asking students to provide an informal explanation, rather than a more technical proof with conventionalized structure and language?

Written communication plays an important role in mathematics (NCTM, 2008), but it may be difficult to assess students' mathematical communication skills fairly. At the authors' institution, a large public research university in Canada, more than a quarter of all students come from other countries (where English is often not a dominant language), and even many of our domestic students use languages other than English in the home. This raises questions of whether asking students to write and be assessed on their English prose in a math course might be unfair or inequitable, given that native English speakers would seem to have certain advantages: being able to write more quickly under time pressure, avoiding grammatical errors, using more colloquial and fluid verbiage, etc.

In this paper, we examine whether a student's language background and other demographic factors have any relationship to their performance on *prose questions*, which we define here as questions with open-ended answers containing one or more complete sentences of English. Prose questions stand in contrast to *non-prose questions*, which are more traditional questions in math courses, requiring an objective answer, such as a number, an equation, a diagram, a formally defined mathematical object, a selected choice from a list of options, etc.

We study this issue in the context of a first-year linear algebra course, and we find that prose questions do *not* appear to give an advantage (or disadvantage) to native English speakers or domestic students; that is, they are fair. A student's performance on prose questions is consistent with their performance on non-prose questions, regardless of which languages they use in the home or whether they are domestic or international. That said, we do see signs of (dis)advantage in the social dimension of gender, which requires further study.

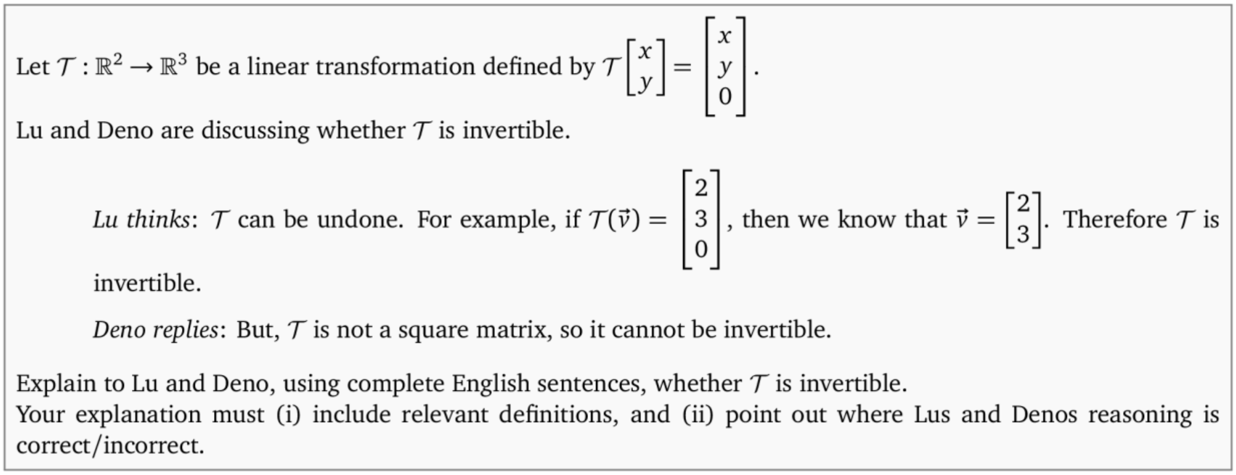
2 Background

With increased globalization and the greater access to information it brings, science communication is an important skill for our students to practice (Kahan et al., 2012). In addition, a focus on written communication specifically is known to have broad educational and cognitive benefits (National Commission on Writing in America's Schools & Colleges, 2003; McArthur, Graham, & Fitzgerald, 2006, Menary, 2007; National Institute for Literacy, 2007), including in mathematics education (Pugalee, 2005).

However, mathematics instructors usually have little training in teaching or assessing writing themselves, so they can struggle with evaluating the *quality* of a student's argument or explanation as different from its *correctness* (Moore, 2016). Further, since mathematics questions can often be asked and answered in ways that minimize the use of ordinary natural language, mathematics teachers may strive to assess their students only in a language-agnostic way, with non-prose questions. For example, questions could be asked that require answers with only equations, diagrams, or true/false responses. In the extreme, questions could even be phrased using only symbolic logic.

Despite the challenges of using prose questions, we believe that they should be used in the mathematics classroom. Not only do they provide the benefits mentioned above, it can also be useful for the instructor to examine a student's writing to provide insights into their (mis)conceptions about the course material and to provide a method of evaluation that focuses on process and understanding rather than just getting the "right" answer (Seto & Meel, 2006).

To that end, we partnered with the Writing-Integrated Teaching program at our institution to bring mathematical writing and communication tasks to our first-year math courses for non-majors. These tasks require students to write multiple sentences or paragraphs of mathematical prose which are evaluated on both correctness and clarity of communication. An example of a prose question for linear algebra is given in Figure 1.



Let $\mathcal{T} : \mathbb{R}^2 \to \mathbb{R}^3$ be a linear transformation defined by $\mathcal{T}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ y \\ 0 \end{bmatrix}$.

Lu and Deno are discussing whether $\mathcal{T}$ is invertible.

Lu thinks: $\mathcal{T}$ can be undone. For example, if $\mathcal{T}(\vec{v}) = \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}$, then we know that $\vec{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$. Therefore $\mathcal{T}$ is invertible.

Deno replies: But, $\mathcal{T}$ is not a square matrix, so it cannot be invertible.

Explain to Lu and Deno, using complete English sentences, whether $\mathcal{T}$ is invertible.
Your explanation must (i) include relevant definitions, and (ii) point out where Lus and Denos reasoning is correct/incorrect.

Figure 1: Example prose question.

International students make up 56% of students in our first-year linear algebra course, and many of these students, as well as many of our domestic students, grew up in homes using languages other than English. Students' language backgrounds have been shown to affect academic achievement (Grayson, 2009), with English as an Additional Language (EAL) students often put in a disadvantaged position relative to their non-EAL peers. In addition, there is variation across countries (Hunt & Wittmann, 2008; Becker, Coyle et al., 2022) and between genders (Turner & Bowen, 1999; Munir & Winter-Ebmer, 2018) in relative performance in math and science versus reading and writing, a so-called "ability tilt" (Coyle et al., 2015), where a *tilt* is an individual's difference across two dimensions of ability (for example, math versus verbal ability). We might expect these observed ability tilts to affect performance on prose versus non-prose questions.

Given the potential for such differences, we worry that questions that rely on English (or any particular language), such as prose questions, might incorrectly assess a student's math knowledge due to their language background, country of origin, gender, or other factors. For example, we might naively expect that students with the same underlying level of mathematical ability would score differently on prose questions if their English skills are different. If there are differences between students, we can better make early identification of which students need the most support and provide appropriate interventions.

3 Methods

We explored this issue by examining data from two midterm exams in a large, multi-section introductory linear algebra course in the fall of 2021 (the final exam for this course was canceled due to the COVID-19 pandemic, so it could not be included). Data was collected and analyzed with respect to the protocol approved by the university's Research Ethics Board.

3.1 Exam Format

Students in all sections took the same two 110-minute in-person midterm exams. The graded portion of the exams consisted of eight questions, each with multiple subparts. Both exams shared the following format:

- one question asking for students to state definitions

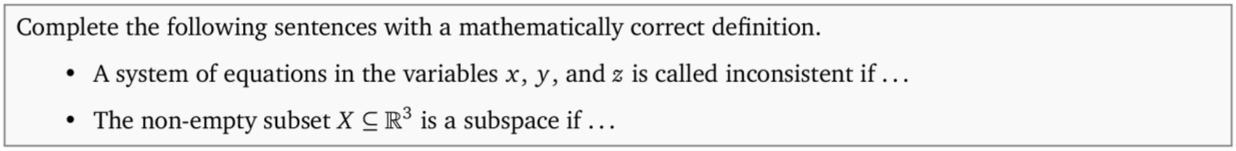
- one *long-form writing question*, where students explain a linear algebra concept and are instructed to do so in full paragraphs
- one question where a student's solution consists of pictures/graphs, without explanations
- one question where students are asked to provide mathematical examples satisfying specific properties or explain why such an example is impossible
- 2–3 multiple choice questions

For reference, the exact questions for Midterm 1 can be found in Appendix 1.

3.2 Two Types of Questions

We coded the questions from these exams as either prose questions or non-prose questions, as defined in Section

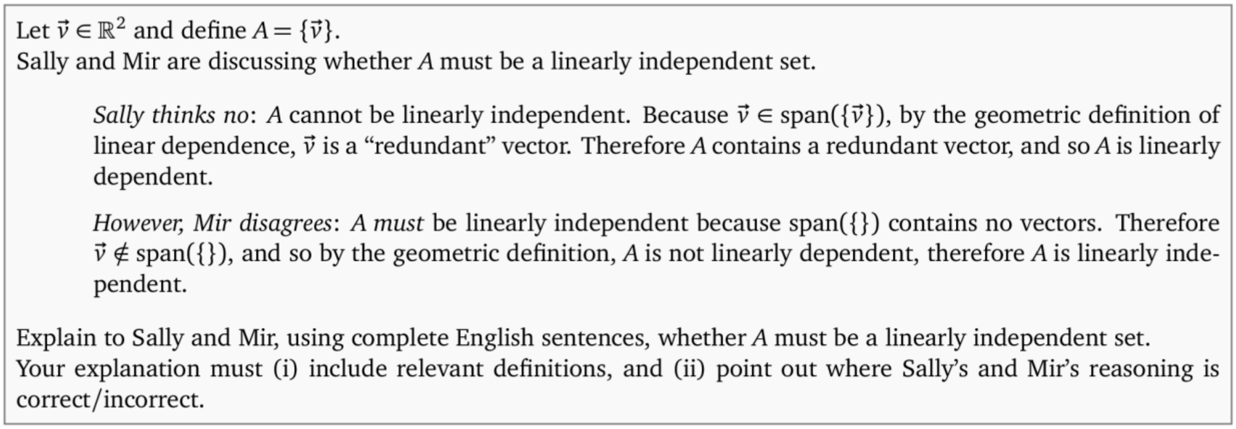
1. Some questions we coded as mixed between the two styles and were excluded from the analysis. There were only two types of questions that qualified as prose questions: (i) definition questions that could not be completed using only a formula and (ii) the long-form writing questions, where students were instructed to write in complete paragraphs. Examples of each of these two types of prose question are given in Figures 2 and 3. Across both exams, seven questions totaling 22 points were categorized as prose questions (approximately 20% of the exams' points).



Complete the following sentences with a mathematically correct definition.

- A system of equations in the variables x, y, and z is called inconsistent if . . .
- The non-empty subset $X \subseteq \mathbb{R}^3$ is a subspace if . . .

Figure 2: Example definition questions. First type of prose question, requiring students to state definitions.



Let $\vec{v} \in \mathbb{R}^2$ and define $A = \{\vec{v}\}$.
Sally and Mir are discussing whether A must be a linearly independent set.

> *Sally thinks no*: A cannot be linearly independent. Because $\vec{v} \in \text{span}(\{\vec{v}\})$, by the geometric definition of linear dependence, $\vec{v}$ is a "redundant" vector. Therefore A contains a redundant vector, and so A is linearly dependent.
>
> *However, Mir disagrees*: *A must* be linearly independent because span($\{\}$) contains no vectors. Therefore $\vec{v} \notin \text{span}(\{\})$, and so by the geometric definition, A is not linearly dependent, therefore A is linearly independent.

Explain to Sally and Mir, using complete English sentences, whether A must be a linearly independent set.
Your explanation must (i) include relevant definitions, and (ii) point out where Sally's and Mir's reasoning is correct/incorrect.

Figure 3: Example long-form writing question. Second type of prose question, requiring students to provide an explanation in complete English sentences.

Most other questions, including definitions that could be stated via a formula, multiple choice questions, and drawing questions were categorized as non-prose questions. Overall, 49 questions totaling 95 points were classified as non-prose questions.

3.3 Differences Between Question Types.

Overall, student performance on non-prose questions was higher than on prose questions by about 3 percentage points (see discussion in Section 4). We define the *non-prose tilt* for a student as the difference between the student's average scores on non-prose questions minus their average score on prose question. In this case, the average non-prose tilt for all students is positive. This is expected, given that non-prose questions are more common in mathematics education, so they are more familiar. Indeed, anecdotally via course evaluations and informal discussions, the authors have found that students are often surprised to find prose questions in math courses.

If prose questions are fair, then we would expect non-prose tilt to vary randomly across students rather than being biased (positively or negatively) for certain demographics. In particular, if prose questions are *linguistically* fair, students' non-prose tilt should not be sensitive to their language background or related demographics like international status, which might correlate with language background. That is, we would expect every language-related demographic group to have roughly the same non-prose tilt.

By focusing on non-prose tilt, we attempt to disentangle students' underlying math ability from the impact of having to express themselves in English. In essence, we expect that students with the same mathematical ability, but different linguistic ability, should perform the same on non-prose questions, but we could potentially see differences in their performance on prose questions, if those questions are not fair.

3.4 Grading Details

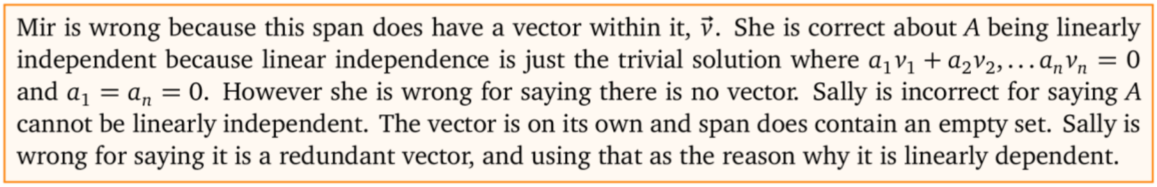
The grading of exams was done online using the *Gradescope* distributed marking platform. Teaching assistants (TAs) were provided with a rubric (see Appendix 2 for the full rubric for Midterm 1), and exams were anonymized, so that TAs were only presented with a student's response to questions with no other identifying information present. For most questions, TAs were asked to grade 20 papers and then wait for feedback from their marking coordinator (an experienced TA or instructor who was assigned to supervise the marking process) before continuing. Most questions were graded strictly and few partial marks were awarded.

The exception to the above process was for the long-form writing question (Figure 3). TAs assigned to mark this question met virtually for a *benchmarking session*. At this benchmarking session, the marking coordinator reviewed several sample student responses with the TAs, and they discussed what points should be awarded to which answer. After the benchmarking session, TAs graded independently and were spot-checked by the marking coordinator. TAs were instructed to mark the question out of 6 points, with half the points assigned for mathematical correctness and half for the quality of presentation. The rubric for this question is as follows:

- Mathematics (3 points, minimum of 0) To get these points, a student must include relevant definitions and have a correct explanation. Deduct points as follows:
  -1pt for not including the definition of linear independence.
  -1pt for not correctly showing when A is linearly independent.
  $-1/2$pt for not explaining what part of Sally's reasoning is incorrect.
  $-1/2$pt for not explaining what part of Mir's reasoning is incorrect.
  $-1/2$pt for not explaining what part of Sally's reasoning is correct.
  $-1/2$pt for not explaining what part of Mir's reasoning is correct.
- Presentation (3 points, minimum of 0) To get these points, a student must provide a well written response with a logical flow. Deduct points as follows:
  -1pt for an answer that is difficult to follow.

−1pt for an answer that is incorrect but is well written.
−2pt for an answer that is very difficult to understand or is not written in complete sentences.
−2pt for an answer that did not include a sufficient amount of detail to answer the question.

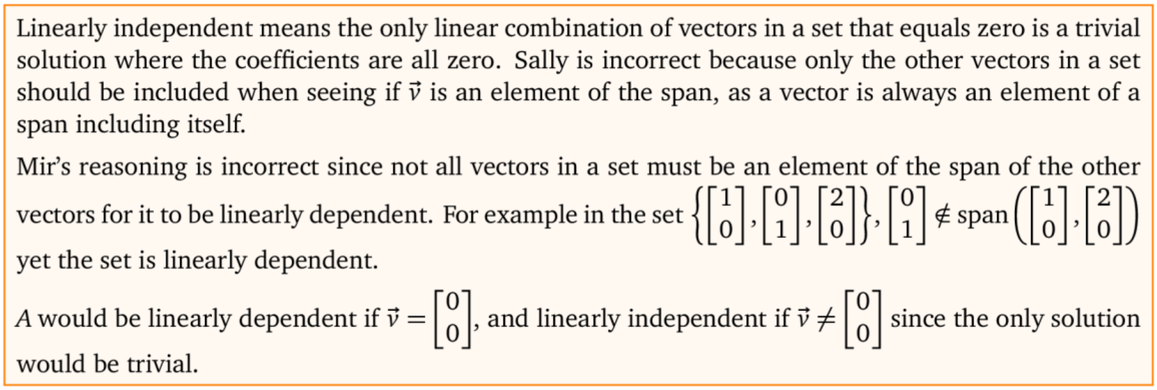
Three example responses to the long-form writing question from Midterm 1 are given in Figures 4–6, along with their scores according to the rubric. The samples were originally hand-written, but they have been typed here for clarity, with each student's original formatting replicated.



Mir is wrong because this span does have a vector within it, $\vec{v}$. She is correct about A being linearly independent because linear independence is just the trivial solution where $a_1v_1 + a_2v_2, \ldots a_nv_n = 0$ and $a_1 = a_n = 0$. However she is wrong for saying there is no vector. Sally is incorrect for saying A cannot be linearly independent. The vector is on its own and span does contain an empty set. Sally is wrong for saying it is a redundant vector, and using that as the reason why it is linearly dependent.

Figure 4: First example response. This response to the long-form writing question on Midterm 1 was scored 1 out of 6 points.

The first example response to the long-form writing question (Figure 4) received 1/6 points. The student lost 3 points for mathematics: 1 point for not stating definitions correctly, 1 point for not correctly showing when A is linearly independent, 0.5 points for not pointing out where Sally's reasoning is correct, and 0.5 points for not showing where Mir's reasoning was correct. This student also lost 2 points for presentation, receiving the rubric feedback item "You have not correctly answered a sufficient amount of the question", with additional clarifying comments provided on the student's paper.



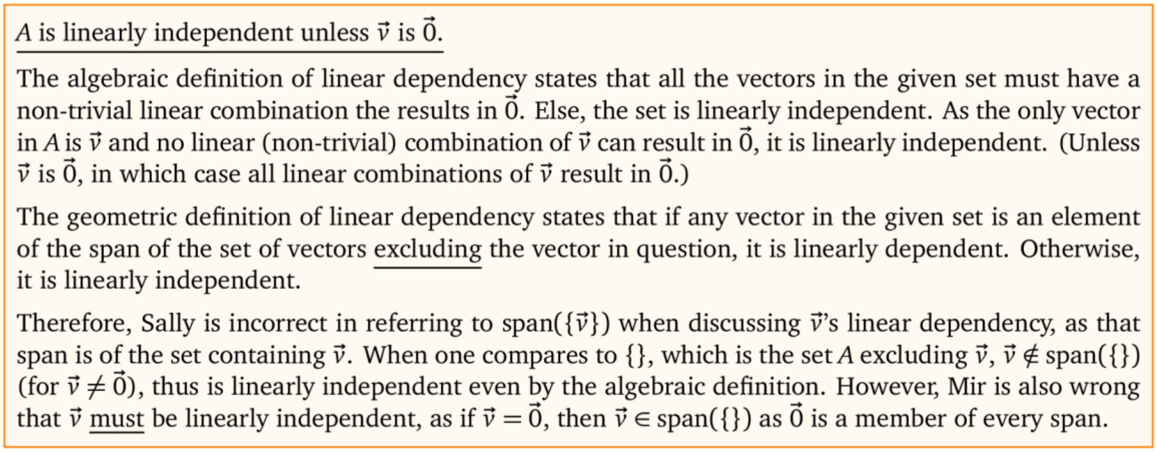
Linearly independent means the only linear combination of vectors in a set that equals zero is a trivial solution where the coefficients are all zero. Sally is incorrect because only the other vectors in a set should be included when seeing if $\vec{v}$ is an element of the span, as a vector is always an element of a span including itself.

Mir's reasoning is incorrect since not all vectors in a set must be an element of the span of the other vectors for it to be linearly dependent. For example in the set $\left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}2\\0\end{bmatrix}\right\}$, $\begin{bmatrix}0\\1\end{bmatrix} \notin \operatorname{span}\left(\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}2\\0\end{bmatrix}\right)$ yet the set is linearly dependent.

A would be linearly dependent if $\vec{v} = \begin{bmatrix}0\\0\end{bmatrix}$, and linearly independent if $\vec{v} \neq \begin{bmatrix}0\\0\end{bmatrix}$ since the only solution would be trivial.

Figure 5: Second example response. This response to the long-form writing question on Midterm 1 was scored 3.5 out of 6 points.

The second example response to the long-form writing question (Figure 5) received 3.5/6 points. The student lost 1.5 points for mathematics: 0.5 points each for not explaining where Sally's reasoning is correct, where Mir's reasoning is correct, and where Mir's reasoning is incorrect. The student also lost 1 point for presentation, receiving the comment "Although your answer is not correct or is incomplete, it is very well-written".



A is linearly independent unless $\vec{v}$ is $\vec{0}$.

The algebraic definition of linear dependency states that all the vectors in the given set must have a non-trivial linear combination the results in $\vec{0}$. Else, the set is linearly independent. As the only vector in A is $\vec{v}$ and no linear (non-trivial) combination of $\vec{v}$ can result in $\vec{0}$, it is linearly independent. (Unless $\vec{v}$ is $\vec{0}$, in which case all linear combinations of $\vec{v}$ result in $\vec{0}$.)

The geometric definition of linear dependency states that if any vector in the given set is an element of the span of the set of vectors excluding the vector in question, it is linearly dependent. Otherwise, it is linearly independent.

Therefore, Sally is incorrect in referring to span($\{\vec{v}\}$) when discussing $\vec{v}$'s linear dependency, as that span is of the set containing $\vec{v}$. When one compares to $\{\}$, which is the set A excluding $\vec{v}$, $\vec{v} \notin$ span($\{\}$) (for $\vec{v} \neq \vec{0}$), thus is linearly independent even by the algebraic definition. However, Mir is also wrong that $\vec{v}$ must be linearly independent, as if $\vec{v} = \vec{0}$, then $\vec{v} \in$ span($\{\}$) as $\vec{0}$ is a member of every span.

Figure 6: Third example response. This response to the long-form writing question on Midterm 1 was scored 5.5 out of 6 points.

Finally, the third example response to the long-form writing question (Figure 6) received 5.5/6 points. The student lost 0.5 points for mathematics, for not explaining what part of Mir's reasoning is correct. The student received full marks for presentation.

3.5 Demographics

Students in the course were primarily in their first year of university. In a typical year, approximately 80% of students who enroll in this course are studying a Science, Technology, Engineering, and Mathematics-related (STEM-related) field, while 20% are studying business, economics, or a liberal arts/social science (statistics for course of study for students in the specific academic session for this study were not collected, but there is no reason to believe that they differed from those of a typical year). Approximately 56% of students in this study were international students (that is, they did not qualify for domestic tuition), with the majority of these international students coming from mainland China. Near the end of the semester, a survey was sent to all students asking about:

1. their use of English as a home language;
2. which languages they are fluent in;
3. their self-assessed proficiency in academic English writing; and
4. their living situation.

We supplemented this survey data with their registration status as an international or domestic student, their gender, their overall exam scores on the two midterm exams, their scores on individual prose and non-prose questions from the midterm exams, and their non-prose tilt (calculated as described above).

3.6 Identifying Significant Factors

This research is exploratory: we want to know what demographic factors might impact a student's non-prose tilt and, in particular, whether their language background is relevant. For this reason, we used a step-down regression procedure, which starts with a model with many predictors and iteratively removes non-significant predictors to find the subset of predictors that builds the best model, as measured by the Akaike information criterion (Akaike, 1974). The step-down regression models were built using the step() function from the stats package in R (R Core Team, 2020). The significance level was set to $\alpha = 0.05$, and p-values were calculated using the lmerTest package (Kuznetsova et al., 2017).

The factors we considered were:

- *native English speaker* (binary, based on whether English was or was not used as a home language)
- *multilingualism* (binary, based on whether the student was fluent in one or more than one language)
- *self-assessed writing proficiency* (linear scale, 1–4, with 4 the highest)
- *living situation* (ternary: on campus, off campus alone or with roommates, and off campus with family)
- *gender* (binary, based on university records; we excluded the 15 students who did not have a recorded gender)
- *international student status* (binary, based on whether their home country was the same or different from the location of the university).

We used these factors to model their (i) midterm average (the average of the scores of the two midterm exams), (ii) average score on all prose questions, (iii) average score on all non-prose questions, and (iv) non-prose tilt (that is, (iii) minus (ii)).

4 Results

In total, there were n = 463 students who took both midterms and filled out the survey with interpretable results. The results for the class as a whole are given in Table **1**. We find a non-prose tilt of 2.83 Note that the overall midterm average is closer to the non-prose average, because non-prose questions make up the majority of the midterm questions. We also carried out preliminary analyses for the two midterm exams separately, but due to their high correlation (R^2 = 0.68), the results were similar enough that we instead report here aggregated results from the two midterm exams together.

Table 1: Combined midterm scores (*n* = 463)

Score Type	Score (%)	SD
Non-prose Questions (Mean)	67.4	18.1
Prose Questions (Mean)	64.6	21.0
Non-prose Tilt (Non-prose – Prose)	2.83	15.0
Overall Midterm (Mean)	66.9	17.7

For midterm averages and the average on non-prose questions, we find three significant factors out of all those tested: living situation, gender, and self-assessed writing proficiency. For the average on prose questions, we find only two significant factors: living situation and self-assessed writing proficiency. The averages for each of these groups for non-prose and prose questions are given in Tables 2–4 and graphed in Figures 7–9.

Table 2: Average scores by question type and living situation

	Campus (*n* = 200)		Self or roommate (*n* = 195)		Family (*n* = 68)	
Question Type	Score (%)	SD	Score (%)	SD	Score (%)	SD
Non-prose	72.3	2.2	65.1	2.6	59.8	4.6
Prose	68.7	2.7	61.8	2.9	60.7	5.6

Table 3: Average scores by question type and gender

	Female (*n* = 195)		Male (*n* = 268)	
Question Type	Score (%)	SD	Score (%)	SD
Non-prose	65.0	2.4	69.2	2.2
Prose	64.6	2.8	64.6	2.6

Table 4: Average scores by question type and self-assessed writing proficiency

	1 = weaker (*n* = 14)		2 (*n* = 122)		3 (*n* = 207)		4 = stronger (*n* = 120)	
Question Type	Score (%)	SD	Score (%)	SD	Score (%)	SD	Score (%)	SD
Non-prose	64.7	12.2	64.3	3.2	67.8	2.4	70.2	3.4
Prose	59.3	16.5	62.0	3.7	64.2	2.8	68.5	3.8

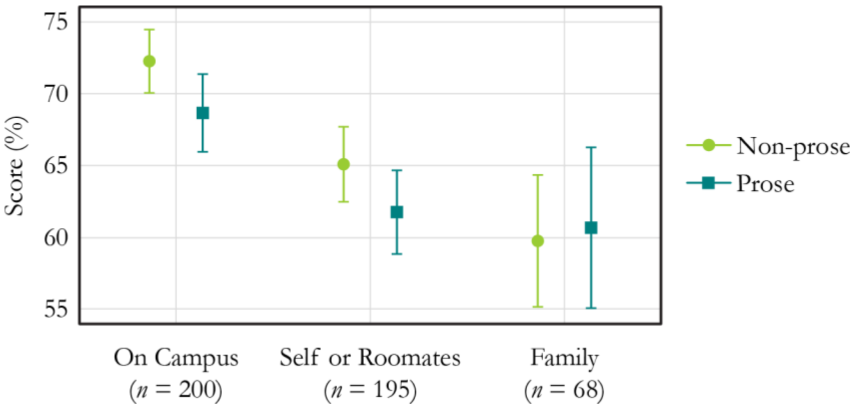


Figure 7: Average scores by question type and living situation. Living situation was a significant factor in non-prose and prose questions.

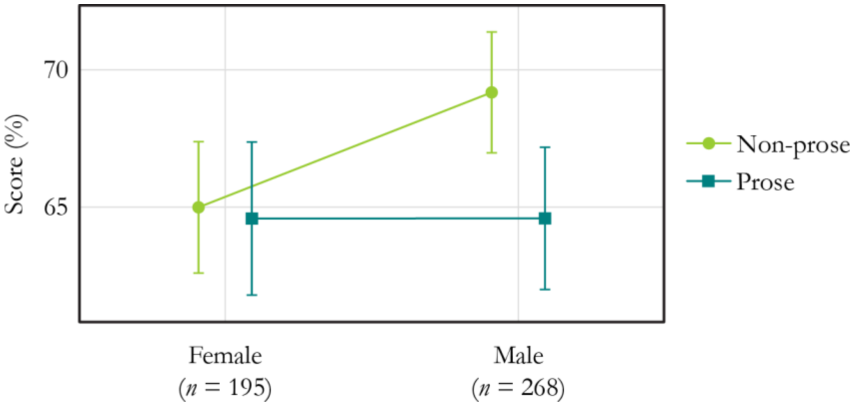


Figure 8: Average scores by question type and gender. Gender was a significant factor only for non-prose question scores. Prose scores are included here for completeness.

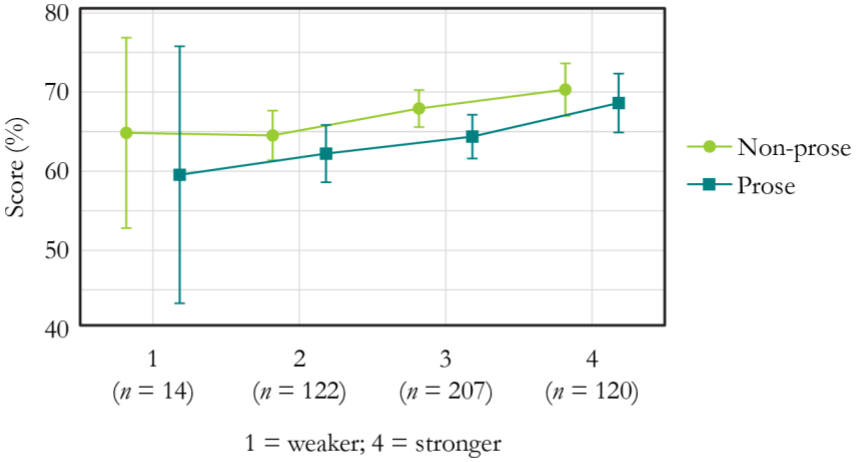


Figure 9: Average scores by question type and self-assessed writing proficiency. Self-assessed writing proficiency was a significant factor in non-prose and prose questions.

The non-prose tilt for each group can be seen in the graphs in Figures 7–9 as the difference in height between the non-prose and prose questions: the higher the non-prose score (the lime green dot) is above the prose score (the dark green square), the greater the non-prose tilt. For example, for living situation (Figure 7), although students living on campus performed the highest overall on both question types, the non-prose tilt is about the same for those students as those living off campus alone or with roommates, with a similar difference between the two question types, while students living with family have a slightly negative non-prose tilt due to scoring higher on prose questions.

For both living situation and self-assessed writing proficiency, the differences between groups is fairly consistent across both question types, so there is no effect on non-prose tilt. However, gender turns out to be a significant factor for non-prose tilt, which is readily apparent in Figure 8, with female students performing about the same on both types of questions, while male students perform about the same as female students on prose questions, but much higher on non-prose questions. The actual non-prose tilt for these two groups is shown in Figure 10.

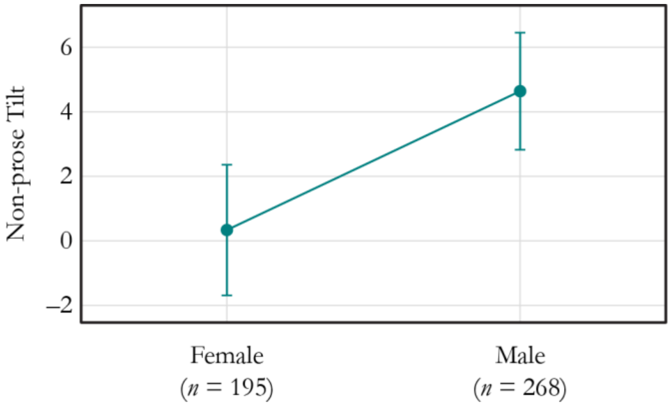


Figure 10: Non-prose tilt by gender.

5 Discussion

The original motivation for the study was primarily to see if language background might correlate to disproportionately lower scores on prose questions, indicating some sort of bias. For example, we expected to find a bias against non-native English speakers, with them having a higher non-prose tilt than native English speakers. We included other demographic variables to explore other possible factors that may also contribute to a bias in performance on prose questions.

Surprisingly, none of the language-related factors (native English proficiency, multilingualism, international status, and self-reported writing proficiency) are significant factors for non-prose tilt, and most linguistic factors (native English proficiency, multilingualism, and international status) are not significant factors for overall midterm scores or for prose or non-prose questions. The only language-related factor that is significant for any aspect of the analysis is self-assessed writing proficiency: students who self-report a higher writing proficiency perform better on both prose and non-prose questions.

The lack of linguistic bias between prose and non-prose questions is an important and valuable result, because it suggests that *writing questions are indeed fair*. They can be asked and evaluated in a math course in ways that do not disproportionately disadvantage students in the linguistic minority. In the rest of this section, we discuss why this superficially counter-intuitive result may arise, and we also consider some possible explanations for why the non-linguistic factors may contribute to a difference in non-prose tilt between certain groups.

5.1 Why Isn't Language Relevant to Non-prose Tilt?

We propose three possible explanations for why language background has no impact on non-prose tilt. First, the *nature of the grading* may have helped minimize biases from the graders. Student exams were anonymized, so graders would not be influenced by student names. The answers were also graded according to a rubric which had graders focus on content rather than gramamr. Furthermore, many of the graders were not native English speakers themselves, so they may have been less attuned to linguistic errors and/or more likely to overlook them.

Second, to be admitted to the university, incoming students who do not speak English as a native language must still demonstrate a *minimum level of fluency* by passing a standardized English test (TOEFL, IELTS, etc.). It may be the case that, when appropriate marking rubrics are used, the university's entry requirements put non-native speakers on a level playing field with native speakers.

These two explanations suggest that careful rubrics and existing university frameworks serve students of diverse language backgrounds equitably. However, it is also possible that students are instead impacted by the *use of English throughout the course*. All course instruction was conducted in English, and all midterm questions were written in English. Thus, students with less proficiency in English could end up with lower performance on all questions regardless of type, due to less effective learning from lectures and/or more difficulty in understanding the midterm questions. If these overall effects are strong enough, they could overshadow any effect of language ability that might be specific only to performance on prose questions.

Regardless, whatever barriers student may face to their performance due to their language background, it seems that prose questions are viable supplements and alternatives to the non-prose questions traditionally used in math courses.

5.2 Factors Affecting Absolute Scores Equally

Two factors, living situation and self-assessed writing proficiency, affect the absolute scores for both non-prose and prose questions, but they do so equally, so there is no difference in non-prose tilt. This means that the differences between these groups is not exacerbated or ameliorated by introducing prose questions.

Perhaps the more surprising of these two factors is living situation, but there are some reasonable explanations for why students living on campus perform might better than other students on both question types. For example, students living off campus have to commute, sometimes as much as 2–3 hours each way, so students living on campus may be able to spend more time studying and may be more likely to attend lectures. They may also make greater use of campus resources (office hours, writing centers, study groups, etc.). Given the high cost of campus residence, they may also be more likely to come from families of higher socioeconomic status, which is known to affect academic performance. More research is needed to untangle why living situation appears to have a significant impact on overall scores.

Explaining the correlation between self-assessed writing proficiency on absolute scores is more straightforward: writing skills and math skills co-vary. That is, students who think they can write better are also better at answering math questions. This may simply correlate to academic ability in a broader sense, because students who rate themselves highly on writing may be the students who just do well in all their school subjects. As with living situation, this affects the scores on the two question types roughly equally, and this effect essentially cancels out when looking at non-prose tilt. This is apparent in Figure 9, with the gap between the two question types remaining fairly constant across the four groups, resulting in mostly parallel trend lines.

Although living situation and self-assessed writing proficiency do not affect non-prose tilt, they do relate to performance on both question types, so it is still important that they be addressed. For example, there appears to be sufficient academic support and learning opportunities for students living on campus, so institutions and instructors should explore how they can better support other students, especially those living off campus with family, who have the lowest performance.

Self-assessed writing proficiency should also be taken more seriously. We see evidence that students are actually quite good at assessing their own ability, at least as it correlates to performance. Interestingly, we see that self-assessment of *writing* ability correlates to performance on *mathematical* questions. Perhaps the use of early self-assessments can be used to help identify students who need more or better support for their learning across the board.

5.3 Gender

Gender differences in STEM courses is an extensively studied subject (Eddy & Brownell, 2016), which is why we included it as a variable of interest. We find that overall midterm scores are statistically significantly higher for male students (68.3%) than for female students (64.9%), which follows the expected gender gap in STEM.

However, when we analyze the midterm scores separately by question type, we find that male and female students perform *equally* on prose questions (64.6% each). This means that gender differences in overall performance on the midterm exams are due to differences only in the non-prose questions, with male students scoring an average of 69.2% compared to female students with 65.0%. Consequently, this shows up as a difference in non-prose tilt, and indeed, gender was the only factor we looked at that had a statistically significant effect on non-prose tilt.

Research suggests that the gender gap in math is likely not due to innate differences between genders (Friedman, 1989; Hyde, 2014). If true, that could mean that prose questions may be a more equitable way to measure a student's math ability. Further theorizing, it is possible that students' lack of experience with prose questions in comparison to more traditional non-prose questions may put all students on more level ground and reduce the impact of existing biases and attitudes about math.

5.4 What do Prose Questions Measure?

A core issue in this analysis is whether (i) prose questions and non-prose questions both measure the same underlying mathematical skills, or whether (ii) these two question types measure distinct skills. Maybe both are true. More research is needed to draw strong conclusions, but we argue that our results support hypothesis (i).

If hypothesis (ii) were correct, we would expect students with a stronger background in English to have a smaller non-prose tilt than those with a weaker background. However, we see no such relationship between non-prose tilt and any linguistic traits. Meanwhile, self-assessed writing ability positively correlates to student performance on *both* prose and non-prose questions, suggesting that self-assessed writing ability may actually correspond to a student's general academic ability rather than writing ability specifically.

Further evidence for hypothesis (i) comes from comparing the two midterm exams to each other. If the skills for each question type were different, we might expect to see differences in improvement over time. For example, students who underperform on prose questions on the first midterm exam may adapt to the question type by the time of the second midterm exam and reduce their non-prose tilt. However, as noted at the outset of this work, our results hold for each midterm exam separately, in the same way. The same factors are significant (or not), to the same extent. Student performance did indeed improve between the midterm exams, but it did so *uniformly* for all groups and for both question types.

5.5 Limitations

As with most educational research, there are many possible confounding variables that may offer alternative interpretations of our data:

- Data on language status and ability is self-reported, and we have evidence of at least some confusion in interpreting the survey questions based on student responses. For example, at least one student marked that they could not speak English, which is a university requirement for all courses.
- There is a selection effect related to who took the survey. About 65% of all students in the course took the survey, and those who did not take the survey scored 8% lower on their midterm exams overall ($p < 0.001$), indicating they are a distinct population.
- Only data from students who took both midterm exams *and* completed the course was analyzed; students who dropped the course may show different results.
- Our university has somewhat unusual demographics, with a large proportion of international students as well as domestic students raised without English as a home language, so these results may not generalize to other situations.

The last point is worth expanding upon. In a pilot phase of data exploration, we had initially assumed that domestic students would overwhelmingly be native speakers of English. However, while most international students in the course were not native English speakers (Mandarin was the most common home language), domestic students were split equally among those with English as a home language and those without. We had hoped to use international student status as a proxy for whether a student was a native English speaker, in order to analyze data from other courses that did not use our demographic survey. However, upon analyzing the survey results, we found that international student status is a poor proxy for whether a student is a native speaker.

Finally, because gender is such a key factor in the analysis, we must note that we are using a binary gender categorization that ignores the nuances of gender identity. The data in our study comes from the gender information provided by the university, and future studies of this type should instead survey gender directly from students.

6 Conclusion

We sought to determine whether asking prose and non-prose questions in a math course would lead to some groups, but not others, doing better or worse on one of the two types of questions. We were especially interested in whether linguistics factors might play a role, given that prose questions require writing prose answers using full sentences of English.

While we find that some populations do perform better or worse on all questions overall, only gender seems to correlate to a difference in performance between the two question types (that is, significantly affecting the non-prose tilt). Crucially, linguistic factors have no significant effect on scores at all, either within question types or for non-prose tilt. Thus, the kind of prose questions developed in this course do not seem to disadvantage linguistically minoritized students, or any student, based on their language background. Furthermore, prose questions may offer another possible tool for helping to narrow the gender gap in math. Mathematics instructors interested in providing more equitable assessment should consider adding more prose questions in their courses.

Prose questions must be used cautiously, of course. The questions and rubrics designed for this course turned out to be fair, but this may not always be the case with prose questions. It is important to make sure that the questions are still accessible regardless of a student's language background and that the rubric for grading focuses on content rather than linguistic form.

To answer the question in the title, writing questions are indeed fair. They allow us to test student knowledge in different ways. Where there are disadvantages, they do not vary by question type (prose versus non-prose), but instead may reflect larger social and structural issues within the education system.

Acknowledgements

We thank the audience at the 6th Northeastern Conference on Research in Undergraduate Mathematics Education for their feedback on this work. We thank the students in MAT223 for their participation in this project. We thank the Writing-Integrated Teaching Program at the University of Toronto for their partnership.

Appendix 1: Midterm 1.

A complete copy of Midterm 1, consisting of eight questions. Questions 1c, 1d, and 4 were coded as prose. Questions 1a, 1b, 1e, 2a, 3a–c, 3e, 5a–d, 6a–e, 7a–e, and 8b were coded as non-prose. Questions 2b, 2c, 3d, and 8a were coded as mixed.

1. Complete the following sentences with a mathematically correct definition. No marks will be awarded for a "close" but incorrect definition.

   (a) (2 points) The *span* of the vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3$ is

   (b) (2 points) The vector $\vec{w}$ is a *linear combination* of the vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3$ if

   (c) (2 points) The vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3$ are *linearly dependent* if

   (d) (2 points) A system of equations in the variables x, y, and z is called *inconsistent* if

   (e) (2 points) The vector $\vec{u} \in \mathbb{R}^3$ is a *unit vector* if

2. Consider the system (A) $\begin{cases} 2x_1 + 4x_2 \phantom{+ \frac{1}{2}x_3} = -2 \\ x_1 + 2x_2 + \frac{1}{2}x_3 = -\frac{1}{2} \\ -2x_1 - 4x_2 + 3x_3 = 5 \end{cases}$.

In this problem you may use the fact that

$$\operatorname{rref}\left(\begin{bmatrix} 2 & 4 & 0 & -2 \\ 1 & 2 & \frac{1}{2} & -\frac{1}{2} \\ -2 & -4 & 3 & 5 \end{bmatrix}\right) = \begin{bmatrix} 1 & 2 & 0 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

(a) (2 points) Write the complete solution to system (A). Express your answer in vector form.

(b) (2 points) If possible, express the complete solution to system (A) using only a span. Otherwise, explain why it cannot be done. (Please mark **Possible** or **Not Possible** in addition to your answer/explanation.)

○ Possible ○ Not Possible

(c) (2 points) If possible, write down a system of **exactly two** linear equations in the variables x_1, x_2, x_3 whose complete solution is the same as the complete solution to system (A). Otherwise, explain why it cannot be done. (Please mark **Possible** or **Not Possible** in addition to your answer/explanation.)

○ Possible ○ Not Possible

3. For each of the following, mark **Possible** if the described object exists, otherwise mark **Not Possible**. If you marked **Possible**, provide an example. If you marked **Not Possible**, explain why it is not possible.

   (a) (2 points) A *unit* vector $\vec{v}$ such that $\vec{v}$ is orthogonal to $\begin{bmatrix}1\\1\end{bmatrix}$, $\begin{bmatrix}2\\2\end{bmatrix}$, and $\begin{bmatrix}0\\0\end{bmatrix}$.

   ○ Possible ○ Not Possible

   (b) (2 points) A *consistent* system of three linear equations in two variables.

   ○ Possible ○ Not Possible

   (c) (2 points) Distinct vectors $\vec{a}, \vec{b} \in \mathbb{R}^2$ such that $\operatorname{span}(\{\vec{a}, \vec{b}\})$ is a *line*.

   ○ Possible ○ Not Possible

   (d) (2 points) Distinct vectors $\vec{a}, \vec{b}, \vec{c} \in \mathbb{R}^2$ such that $\{\vec{a}, \vec{b}, \vec{c}\}$ is linearly *independent*.

   ○ Possible ○ Not Possible

   (e) (2 points) A non-zero vector $\vec{v}$ such that $\vec{v}$ is orthogonal to *every* vector in the set $Z = \{\vec{p} \in \mathbb{R}^3 : \text{the coordinates of } \vec{p} \text{ sum to zero}\}$.

   ○ Possible ○ Not Possible

4. (6 points) Let $\vec{v} \in \mathbb{R}^2$ and define $A = \{\vec{v}\}$. Sally and Mir are discussing whether A must be a linearly independent set.

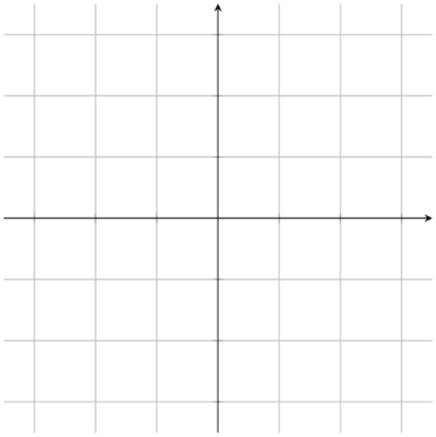
> *Sally thinks no*: A cannot be linearly independent. Because $\vec{v} \in \operatorname{span}(\{\vec{v}\})$, by the geometric definition of linear dependence, $\vec{v}$ is a "redundant" vector. Therefore A contains a redundant vector, and so A is linearly dependent.
>
> *However, Mir disagrees*: A *must* be linearly independent because $\operatorname{span}(\{\})$ contains no vectors. Therefore $\vec{v} \notin \operatorname{span}(\{\})$, and so by the geometric definition, A is not linearly dependent, therefore A is linearly independent.

Explain to Sally and Mir, using complete English sentences, whether A must be a linearly independent set. Your explanation must (i) include relevant definitions, and (ii) point out where Sally's and Mir's reasoning is correct/incorrect.

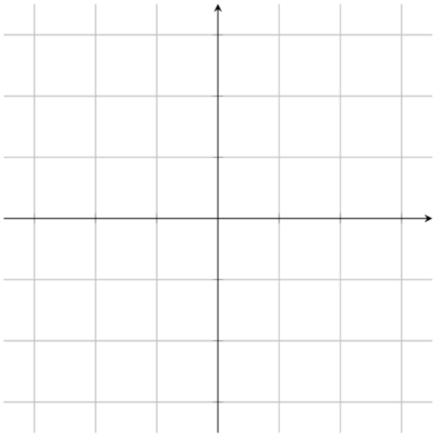
5. Let $B = \left\{ \begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}1\\1\end{bmatrix} \right\}$ and let $\vec{\boldsymbol{k}} = \vec{e}_1 + \vec{e}_2$, where $\vec{e}_1, \vec{e}_2$ are the standard basis vectors for $\mathbb{R}^2$. Further, let $U = \left\{ \vec{u} \in \mathbb{R}^2 : \|\vec{u}\| = 2 \text{ and } \{\vec{u}, \vec{\boldsymbol{k}}\} \text{ is linearly } \boldsymbol{\mathit{independent}} \right\}$.

Draw the following subsets of $\mathbb{R}^2$.

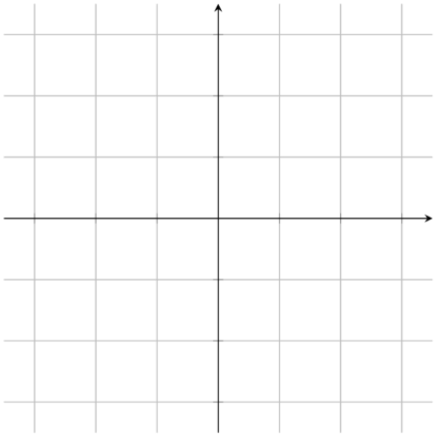
(a) (2 points) A line $\ell \subseteq \mathbb{R}^2$ that ***cannot*** be expressed as a span.



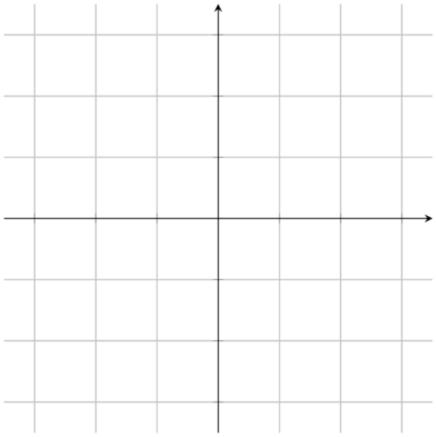
(b) (2 points) A linearly ***dependent*** set $Y = \{\vec{a}, \vec{b}, \vec{c}\}$, consisting of three vectors, such that $\text{span}(Y)$ is a line.



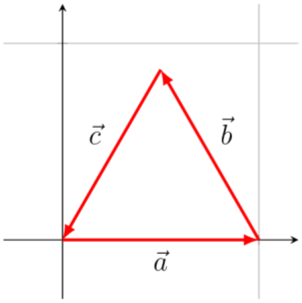
(c) (2 points) U



(d) (2 points) A set X such that $B + X$ consists of *exactly* six points. (Draw X, not $B + X$.)



6. Let $\vec{a}, \vec{b}, \vec{c} \in \mathbb{R}^2$ be ***unit*** vectors drawn below.



(a) (2 points) For $\vec{x} \in \{\vec{a}, \vec{b}, \vec{c}\}$, which $\vec{x}$ make $\vec{a} \cdot \vec{x}$ **positive**? Mark all that apply.

○ $\vec{a}$ ○ $\vec{b}$ ○ $\vec{c}$ ○ None of these

(b) (2 points) For $\vec{x} \in \{\vec{a}, \vec{b}, \vec{c}\}$, which $\vec{x}$ make $\vec{a} \cdot \vec{x}$ **negative**? Mark all that apply.

○ $\vec{a}$ ○ $\vec{b}$ ○ $\vec{c}$ ○ None of these

(c) (2 points) For $\vec{x} \in \{\vec{a}, \vec{b}, \vec{c}\}$, which $\vec{x}$ make $\vec{a} \cdot \vec{x}$ **zero**? Mark all that apply.

○ $\vec{a}$ ○ $\vec{b}$ ○ $\vec{c}$ ○ None of these

(d) (2 points) Which of the following vector equations are **consistent**? Mark all that apply.

○ $t\vec{a} + s\vec{b} = \vec{c}$ ○ $t\vec{a} + s\vec{b} = \vec{0}$ ○ $t\vec{a} = -\vec{b}$ ○ None of these

(e) (3 points) Let $\ell \subseteq \mathbb{R}^2$ be the line segment from $\vec{0}$ to $2\vec{a}$ (including its endpoints). Express ℓ in set-builder notation.

$$\ell = \left\{ \qquad\qquad\qquad\qquad \right\}$$

Scratch work:

7. Let $\vec{v}_1 = \begin{bmatrix}0\\0\\1\end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix}1\\a\\0\end{bmatrix}$, and $\vec{v}_3 = \begin{bmatrix}b\\b\\b\end{bmatrix}$ for unknown constants $a, b \in \mathbb{R}$ which satisfy $a \geq 1$ and $b \leq 1$.

For each question below mark

- **ALWAYS TRUE** if the statement is always true,
- **ALWAYS FALSE** if the statement is always false **or** if the statement does not make mathematical sense, and
- **DEPENDS ON *a*/*b*** if the statement could be true or could be false, depending on the values of a and/or b.

No justification is needed.

(a) (2 points) $\begin{bmatrix}2\\0\\0\end{bmatrix} \in \mathcal{P}$ where $\mathcal{P}$ is given in vector form by $\vec{x} = t\vec{v}_1 + s\vec{v}_2$.

○ **ALWAYS TRUE** ○ **ALWAYS FALSE** ○ **DEPENDS ON *a*/*b***

(b) (2 points) The line $\ell = \text{span}\{\vec{v}_1\}$ can be written in vector form as "$\vec{x} = t\vec{v}_1$ for some $t \in \mathbb{R}$".

○ **ALWAYS TRUE** ○ **ALWAYS FALSE** ○ **DEPENDS ON *a*/*b***

(c) (2 points) $\text{span}\{\vec{v}_1, \vec{v}_2, \vec{v}_3\} = \mathbb{R}^3$.

○ **ALWAYS TRUE** ○ **ALWAYS FALSE** ○ **DEPENDS ON *a*/*b***

(d) (2 points) $\text{span}\{\vec{v}_1, \vec{v}_2, \vec{v}_3\} = \mathbb{R}^2$.

○ **ALWAYS TRUE** ○ **ALWAYS FALSE** ○ **DEPENDS ON *a*/*b***

(e) (2 points) The set $\{\vec{v}_1, \vec{v}_2\}$ is linearly **independent**.

○ **ALWAYS TRUE** ○ **ALWAYS FALSE** ○ **DEPENDS ON *a*/*b***

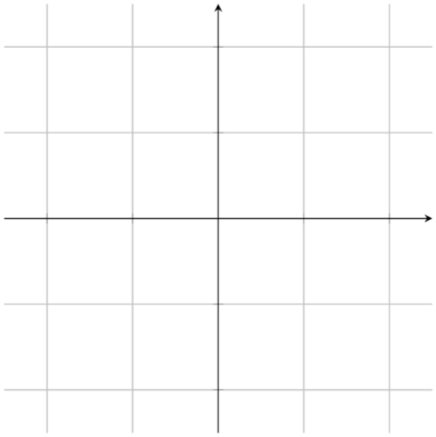
8. In this question, you will work with a new definition.

Let $K \subseteq \mathbb{R}^2$. The set K is called $(1,1)$-*independent* if the vector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ can be written *uniquely* as a linear combination of vectors in K.

(a) (3 points) Is $\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right\}$ a $(1,1)$-independent set? Mark **Yes** or **No** in addition to justifying your answer.

○ Yes ○ No

(b) (2 points) Let $X = \left\{ \vec{v} \in \mathbb{R}^2 : \text{ the set } \{\vec{v}\} \text{ is } (1,1)\text{-independent} \right\}$. Draw X.



Appendix 2: Midterm 1 Rubric.

Reproduced below is the rubric given to TAs, including instructions. In addition to the provided rubric, TAs were spot-checked in their marking by the marking coordinators and asked to redo any marking that was inconsistent or did not fit the rubric.

- Please use the rubric items provided. Do **not** change a rubric item without the permission of your marking coordinator (changing an item will change it for all tests simultaneously).
- If you feel like the rubric items don't fit well with the answers your seeing, talk to your marking coordinator about creating new/modifying the rubric items.
- If you need to add a one-off comment to a booklet, use the comment box at the bottom of the page. Do not attach points to your comment (all points must come from rubric items).
- Use the keyboard shortcuts! You will save a lot of time. See the cheatsheet on the top right corner from any grading page.

Grading scheme

1. (2 points each) Please read the definitions carefully. **We will not give any points for a "close" but incorrect correct definition.** For each part give:

- 2 points for a correct definition
- 0 points otherwise.
- (a) Saying "all linear combinations" is worth 0. Saying "the **set** of all linear combinations of $\vec{v}_1$, $\vec{v}_2$, $\vec{v}_3$" is worth full points.
- (b) 0 points if they omitted a quantifier.
- (d) "there is a solution to the system" is worth full points.
- (e) 0 points if they wrote "$\sqrt{u_1^2 + u_2^2 + u_3^2}$" if they didn't define what u_1, u_2, and u_3 were.

2.

(a) (2 points)

- 2 points for a correct answer written in vector form. If they write $\{\vec{x} : \vec{x} = t\vec{d} + \vec{p}$ for some $t\}$ (as long as the vectors are correct and the variable is quantified correctly) give full points.
- Give them 1 point if they write a correct solution to the system in a different form.
- 0 points otherwise.
- 0 points if they added a quantifier "for some" or "for all" to their vector-form equation.

(b) (2 points)

- 1 point for saying "No".
- 1 point for correct reasoning.

(c) (2 points)

- 1 point for saying "Yes".
- 1 point for correct example.

3. (2 points each)

Parts (d)

- 1 point for saying that it is impossible.
- 1 point for a correct explanation.

Parts (a),(b),(c),(e).

- 2 points for a correct example.
- 0 otherwise.

4. (6 points)

- Mathematics (3 points, minimum of 0) To get these points, a student must include relevant definitions and have a correct explanation. Deduct points as follows:
  -1pt for not including the definition of linear independence.
  -1pt for not correctly showing when A is linearly independent.
  $-1/2$pt for not explaining what part of Sally's reasoning is incorrect.
  $-1/2$pt for not explaining what part of Mir's reasoning is incorrect.
  $-1/2$pt for not explaining what part of Sally's reasoning is correct.
  $-1/2$pt for not explaining what part of Mir's reasoning is correct.
- Presentation (3 points, minimum of 0) To get these points, a student must provide a well written response with a logical flow. Deduct points as follows:
  -1pt for an answer that is difficult to follow.
  -1pt for an answer that is incorrect but is well written.
  -2pt for an answer that is very difficult to understand or is not written in complete sentences.
  -2pt for an answer that did not include a sufficient amount of detail to answer the question.

5. (2 points each)

- 2 points for a correct answer.
- 0 points otherwise.

Note: if their drawing is unclear and e.g., you can't tell what's a line segment and what's a vector, give them 0 with a comment explaining. If they indicated clearly but it is *close* to being unclear, give them full points, but add a comment.

6.

(a)–(d) (2 points)

- 2 points for circling **all** correct answers.
- 0 points otherwise.

(e) (3 points)

- 3 points for a correct set.
- 0 points for any errors.

Note: if they add their own set brackets, and have inadvertently written $\{\{\vec{v} : \dots\}\}$ do not take off any points.

7. (2 points each)

- 2 points for a correct answer.
- 0 points otherwise.

8.

(a) (2 points)

- 1 points for specifying "Yes"
- 2 points for a good explanation that is well written.
- 0 points otherwise.

(b) (2 points)

- 2 points for a correct drawing.
- 0 points otherwise.

Note: if $\vec{0}$ is included in their drawing, it is incorrect. They may have used words to further explain what is happening in their drawing. Please read the supporting words if there are any.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716-723. doi: 10.1109/TAC.1974.1100705

Becker, D., Coyle, T. R., Minnigh, T. L., & Rindermann, H. (2022). International differences in math and science tilts: The stability, geography, and predictive power of tilt for economic criteria. *Intelligence*, *92*(May–June), 101646.

Coyle, T. R., Snyder, A. C., & Richmond, M. C. (2015). Sex differences in ability tilt: Support for investment theory. *Intelligence*, *50*, 209–220.

Eddy, S. L., & Brownell, S. E. (2016). Beneath the numbers: A review of gender disparities in undergraduate education across science, technology, engineering, and math disciplines. *Physical Review Physics Education Research*, *12*(2), 020106.

Friedman, L. (1989). Mathematics and the gender gap: A meta-analysis of recent studies on sex differences in mathematical tasks. *Review of Educational Research*, *59*(2), 185–213. doi: 10.2307/1170414

Grayson, J. P. (2009). Language background, ethno-racial origin, and academic achievement of students at a Canadian university. *International Migration*, *47* (2), 33–67.

Hunt, E., & Wittmann, W. (2008). National intelligence and national prosperity. *Intelligence*, *36*(1), 1–9.

Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology*, *65*, 373-398. doi: 10.1146/annurev-psych-010213-115057

Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change*, *2*(10), 732–735. doi: 10.1038/nclimate1547

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. doi: 10.18637/jss.v082.i13

McArthur, C., Graham, S., & Fitzgerald, J. (2006). *The handbook of writing research*. New York, NY: Guilford Press.

Menary, R. A. (2007). Writing as thinking. *Language Sciences*, *29*(5), 621–632.

Moore, R. C. (2016). Mathematics professors' evaluation of students' proofs: A complex teaching practice. *International Journal of Research in Undergraduate Mathematics Education*, *2*(2), 246–278.

Munir, F. M., & Winter-Ebmer, R. (2018). Decomposing international gender test score differences. *Journal for Labour Market Research*, *52*, 12.

National Commission on Writing in America's Schools & Colleges. (2003). *The neglected "R": The need for a writing revolution*. New York, NY: The College Board.

National Institute for Literacy. (2007). *What content-area teachers should know about adolescent literacy*. Washington, DC: National Institute for Literacy.

NCTM. (2008). *Principles and standards for school mathematics* (5th ed.). Reston, VA: National Council of Teachers of Mathematics.

Pugalee, D. K. (2005). *Writing to develop mathematical understanding*. Norwood, MA: Christopher-Gordon.

R Core Team. (2020). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org.

Seto, B., & Meel, D. E. (2006). Writing in mathematics: Making it work. *Problems, Resources, and Issues in Mathematics Undergraduate Studies*, *16*(3), 204–232.

Turner, S. E., & Bowen, W. G. (1999). Choice of major: The changing (unchanging) gender gap. *Industrial and Labor Relations Review*, *52*(2), 289–313.