**Analysis**

**Check which problems got both answers wrong, why, what’s the mutual? What’s the topic?**

**I wanted to use Dynamic batch processing(read what is it again), but I felt it unnecessary as the problems are not that complicated.**

1. - Both answers are **correct**.

- Difficulty level: **basic**

- runtime: **modified more than doubled**

1. - **The initial prompt** came up with 10 examples(one was sufficient) and all of them were **inaccurate** and did not satisfy the condition.

**The modified prompt** came up with only 2 examples and they were both **correct**, it also had a clear and accurate explanation of the steps to the solution.

- Difficulty level: **basic**

- runtime: **more than tripled**

1. - The **initial prompt** showed 8 answers, it should’ve been only 3, but important to mention is that the first answer was **correct**, the others were wrong, but the **modified** prompt got one answer and it was **wrong**.

- Difficulty level: **basic**

- runtime: **modified more than 4x**

1. **Initial prompt was wrong**, even though the question was multiple choice of only 2 choices, so there was a 50% chance to get it right.

**The modified prompt got it right.**

- Difficulty level: **intermediate**

- runtime: **modified more than doubled**

1. Surprisingly the **initial** prompt gave the **correct** answer, while the **modified** prompt **couldn’t**, even after making so many steps to get the answer.

- Difficulty level: **intermediate**

- runtime: **modified more than 5x (and still got it wrong)**

The initial prompt only needed 0.7 seconds to get the right answer, while the modified got a wrong answer after 5.5 seconds.

1. Both are wrong. (different answers)

- Difficulty level: **intermediate**

- runtime: **modified more than 5x (and still got it wrong)**

1. Both are wrong. (different answers)

- Difficulty level: **intermediate**

- runtime: **modified more than 5x (and still got it wrong)**

1. Both are wrong. (different answers)

- Difficulty level: **intermediate**

- runtime: **modified more than 5x (and still got it wrong)**

1. Both are correct, with different answers

- Difficulty level: **intermediate**

- runtime: **modified more than 4x**

1. Question is basic and so easy but both failed to answer, the initial prompt answer is worse though.

- Difficulty level: **basic**

- runtime: **modified more than 3x**

1. The initial prompt got some answers correct, but not all. The modified answer only made one answer(there should be more) and it was correct, but at least it didn’t hallucinate and give wrong answers.

- Difficulty level: **basic**

- runtime: **modified more than 3x**

1. Both are wrong, initial took less than a second, while the modified took longer than 7 seconds(too much for a wrong answer in the end)

- Difficulty level: **basic**

- runtime: **modified more than 7x**

1. Both are wrong.

- Difficulty level: **intermediate**

- runtime: **modified more than 2x**

1. Both are wrong (different answers)

- Difficulty level: **intermediate**

- runtime: **modified more than 5x**

1. The initial is stupidly wrong, the question says it should be a 5-stellige number and it is only 3. The modified is mostly correct, it just has a small mistake in the end.

- Difficulty level: **advanced**

- runtime: **modified more than 3x**

1. Modified is more accurate, the initial is not wrong but is missing, it doesn’t have the complete answer.

- Difficulty level: **intermediate**

- runtime: **modified more than 4x**

1. Modified is correct, initial is wrong.

- Difficulty level: **advanced**

- runtime: **modified more than 4x**

1. Initial is horrible and nonsense. Modified has some valid steps but it has made wrong assumptions.

- Difficulty level: **advanced**

- runtime: **not a very big difference, initial also took a lot of time (4.1 seconds).**

1. Initial wrong. Modified correct.

* Difficulty level: **advanced**
* runtime: **modified more than 4x**

1. Both are wrong but the question was not very clear to answer.

* Difficulty level: **intermediate**
* runtime: **modified more than 4x**

1. Initial is wrong, modified is 100% accurate.

* Difficulty level: **basic**
* runtime: **modified more than 3x**

1. Initial is wrong, modified is 100% accurate.

* Difficulty level: **intermediate**
* runtime: **modified more than 2x**

1. Both are wrong.

* Difficulty level: **intermediate**
* runtime: **modified more than 6x(still wrong after all this computing)**

1. Initial is not even close, while modified is correct, 100% accurate.

* Difficulty level: **advanced**
* runtime: **modified more than 7x**

1. The question is tricky and it should’ve been more clear because the answer of both prompts are actually correct, but this is not exactly what the question wanted.

* Difficulty level: **advanced**
* runtime: **modified more than 2x**

1. Initial is wrong, but modified is accurate and correct.

* Difficulty level: **advanced**
* runtime: **modified more than 4x**

1. Both are wrong. Initial was very quick, this type of algebraic equation questions usually cannot be solved this fast, this answer is hallucination and not anywhere near correct.

* Difficulty level: **advanced**
* runtime: **modified more than 4x**

1. Initial got 2 out of 3 answers correct and the modified got it all correct with easy steps and the runtime was good.

* Difficulty level: **intermediate**
* runtime: **modified more than 3x**

1. Both are correct but the modified one is more clear and more accurate.

* Difficulty level: **intermediate**
* runtime: **modified more than 5x (initial took 0.6 seconds)**

1. Both are wrong.

* Difficulty level: **basic**
* runtime: **modified more than 4x**

1. The initial is correct, while the modified is wrong. **(This has never happened before, this is the first such case)**

* Difficulty level: **basic**
* runtime: **initial 3.7 - modified 4.5**

1. Both are wrong.

* Difficulty level: **intermediate**
* runtime: **modified more than 3x**

1. Both are wrong.

* Difficulty level: **intermediate**
* runtime: **modified more than 3x**

1. Initial is almost correct, it is just missing a small detail. Modified is 100% accurate.

* Difficulty level: **intermediate**
* runtime: **modified more than 5x**

1. Both have the same answer and they are correct.

* Difficulty level: **intermediate**
* runtime: **modified more almost 3x**

1. Both are wrong.

* Difficulty level: **advanced**
* runtime: **modified more almost 10x (initial only took 0.5)**

1. Both are wrong.

* Difficulty level: **intermediate**
* runtime: **modified more almost 13x and still wrong.**

1. Both are wrong.

* Difficulty level: **intermediate**
* runtime: **modified more than 10x (initial only took 0.9)**

1. Both are correct.

* Difficulty level: **advanced (for some advanced problems, even the initial prompt solves it correctly, that proves that the LLM is capable but the problem usually lies in phrasing the prompt)**
* runtime: **modified more than 5x (initial only took 0.9)**

1. Both are completely wrong.

* Difficulty level: **advanced**
* runtime: **modified more than 5x**

1. Both are correct.

* Difficulty level: **advanced**
* runtime: **modified more than 4x**

**Logic(edit numbers later):**

1. Both are wrong, but modified was going okay till it made wrong calculations.

* Difficulty level: **intermediate**
* runtime: **modified more than 3x**

1. Both are accurate.

* Difficulty level: **basic**
* runtime: **modified more than 2x**

1. Both are correct.

* Difficulty level: **basic**
* runtime: **modified more than 3x (initial was very quick, took less than a second)**

1. Initial was wrong, while modified was correct.

* Difficulty level: **basic**
* runtime: **difference is only 1 second**

1. Both are wrong.

* Difficulty level: **intermediate**
* runtime: **modified more than 10x**

1. Initial was wrong, modified was correct

* Difficulty level: **basic**
* runtime: **modified more than 4x**

1. Both are correct.

* Difficulty level: **basic**
* runtime: **modified more than 6x, initial took less than a second**

1. Both are correct. (The solution in the book is arguably wrong.)

* Difficulty level: **basic**
* runtime: **modified more than 5x, initial took less than a second**

1. Both are wrong, but modified is more confused than the initial

* Difficulty level: **intermediate**
* runtime: **modified more than 4x**

1. Initial is wrong. Modified got 2 out of 3 answers correct, better than initial.

* Difficulty level: **advanced**
* runtime: **modified more than 2.5x**

1. Initial wrong, modified correct.

* Difficulty level: **intermediate**
* runtime: **modified more than 3x**

1. Initial very wrong, modified correct.

* Difficulty level: **intermediate**
* runtime: **modified more than 3x**

1. Initial is wrong and modified is 100% accurate.

* Difficulty level: **basic**
* runtime: **almost the same, 0.4s difference**

1. Initial wasn’t entirely correct, but not entirely wrong either. Modified was completely wrong.

* Difficulty level: **intermediate**
* runtime: **modified more than 5x**

1. This question doesn’t exactly have one solution, but the initial’s answer doesn’t make sense.

* Difficulty level: **intermediate**
* runtime: **modified more than 3x**

1. Both are correct.

* Difficulty level: **basic**
* Runtime: **modified more than 2x**

1. Both are correct.

* Difficulty level: **basic**
* Runtime: **modified more than 3x**

1. Both answers are plausible because the question is more of a creativity question.

* Difficulty level: **intermediate**
* Runtime: **modified more than 4x**

1. Initial is not wrong but may not be sufficient, modified is correct.

* Difficulty level: **basic**
* Runtime: **modified more than 5x**

1. Both are wrong.

* Difficulty level: **basic**
* Runtime: **modified more than 2x**

1. Initial is wrong. Modified is correct.

* Difficulty level: **basic**
* Runtime: **modified more than 5x(initial only took 0.5s)**

1. Both are wrong.

* Difficulty level: **intermediate**
* Runtime: **modified more than 2x**

1. Initial is not even close but modified is getting close, it just didn’t cover all the points, but it didn’t make any mistakes.

* Difficulty level: **basic**
* Runtime: **modified more than 3x**

1. Both failed to answer.

* Difficulty level: **advanced**
* Runtime: **modified more than 2x**

1. Initial is wrong. Modified is correct.

* Difficulty level: **intermediate**
* Runtime: **modified more than 6x(initial only took 0.5s)**

1. Initial is wrong. Modified was correct but it stopped before reaching the final solution.(the way was correct)

* Difficulty level: **advanced**
* Runtime: **modified more than 4x**

1. Initial was bluntly wrong. Modified is perfectly accurate.

* Difficulty level: **intermediate**
* Runtime: **initial 0.6, modified 1.3**

1. Initial is completely wrong, modified got one out of two correct.

* Difficulty level: **advanced**
* Runtime: **modified more than 4x**

1. Both are wrong.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.9, modified is 1.7**

**Book2 (edit numbers later):**

1. (2.7) Both are correct.

* Difficulty level: **intermediate**
* Runtime: **initial is 1, modified is 3.3**

1. (2.23) Initial is completely wrong, while the modified is very accurate.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.69, modified is 4.1**

1. (2.24) Both are correct.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.85, modified is 2.2**

1. (3.4) Both are correct.

* Difficulty level: **basic**
* Runtime: **initial is 0.54, modified is 2.3**

1. (3.14) Initial is wrong, while the modified is totally accurate.

* Difficulty level: **basic**
* Runtime: **initial is 0.59, modified is 4.9**

1. (3.16) Both are wrong.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.88, modified is 5.49**

1. (3.20) Both are wrong.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.79, modified is 4**

1. (3.31) Initial is wrong, while the modified is accurate.

* Difficulty level: **basic**
* Runtime: **initial is 1.14, modified is 2.5**

1. (3.32) Initial is very wrong, while the modified is totally accurate.

* Difficulty level: **basic**
* Runtime: **initial is 0.98, modified is 3.6**

1. (3.36) Initial is wrong, while the modified is totally accurate.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.54, modified is 4.59**

1. (5.4) Initial got the 3 easiest questions right and the last and more difficult ones wrong, while the modified was totally accurate in all 6 of them.

* Difficulty level: **basic**
* Runtime: **initial is 2.45, modified is 11.7**

1. (7.13) Both are correct.

* Difficulty level: **basic**
* Runtime: **initial is 0.82, modified is 3.9**

1. (9.5) initial is very wrong, it uses undefined logarithms. Modified is completely accurate.

* Difficulty level: **basic**
* Runtime: **initial is 0.81, modified is 4.41**

1. (9.15) Both are wrong but the modified is almost correct.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.65, modified is 6.79**

1. (10.19) initial is very wrong and very far away from the answer but the modified is accurate and correct.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.79, modified is 5.7**

1. (14.1) initial is correct, while modified is wrong.(this was a yes or no question, that might have been the reason the initial got it right, could be luck since it’s a 50% chance).

* Difficulty level: **basic**
* Runtime: **initial is 0.56, modified is 4.89**

1. (14.2) Both are wrong.

* Difficulty level: **basic**
* Runtime: **initial is 1.13, modified is 5.87**

1. (14.3) Both are correct.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.549, modified is 2.57**

1. (14.4) Both are correct.

* Difficulty level: **basic**
* Runtime: **initial is 0.5, modified is 3**

1. (14.14) Both are wrong.

* Difficulty level: **basic**
* Runtime: **initial is 0.67, modified is 4.65**

1. (15.11(HINWEIS)) Both are correct.

* Difficulty level: **basic**
* Runtime: **initial is 0.67, modified is 4.65**

1. (16.1) Both are correct.

* Difficulty level: **basic**
* Runtime: **initial is 0.8, modified is 5**

1. (16.5) initial is wrong but the modified is correct.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.49, modified is 3.27**

1. (18.1) Both are correct.

* Difficulty level: **basic**
* Runtime: **initial is 0.97, modified is 5.9**

1. (36.3) Both are correct.

* Difficulty level: **advanced**
* Runtime: **initial is 0.98, modified is 7.5**

1. (36.8) Both are correct.

* Difficulty level: **basic**
* Runtime: **initial is 1.2, modified is 4.1**

1. (36.9) Both are correct. (yes or no question, so that’s why runtime is very short)

* Difficulty level: **basic**
* Runtime: **initial is 0.57, modified is 3**

1. (37.10) initial is wrong, it calculated a number that it shouldn’t have, it was capable of solving the answer. Modified is correct.

* Difficulty level: **basic**
* Runtime: **initial is 0.55, modified is 3.58**

1. (37.11) Both are correct.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.53, modified is 3.16**

1. (37.11) Both are correct.

* Difficulty level: **advanced!!!**
* Runtime: **initial is 0.54, modified is 4**

1. (37.15) Initial is a wrong number, but the modified is the exact correct number.

* Difficulty level: **basic**
* Runtime: **initial is 0.64, modified is 5.3**

1. (37.16) initial is wrong, modified is accurate.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.82, modified is 3.2**

1. (37.18) Both are correct.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.81, modified is 2**

1. (37.21) Both answers are wrong.

* Difficulty level: **advanced**
* Runtime: **initial is 0.68, modified is 4.48**

1. (37.22) Initial is wrong, modified is correct.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.71, modified is 7.96**

1. (37.21) Both answers are wrong.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.79, modified is 8.89**

1. (39.6) Initial is wrong, even though it’s answering with a yes or no, modified is correct.

* Difficulty level: **basic**
* Runtime: **initial is 0.47, modified is 4**

1. (39.7) Initial is wrong, even though it’s answering with a yes or no, modified is mostly correct but could be clearer.

* Difficulty level: **basic**
* Runtime: **initial is 0.68, modified is 6.2**

1. (39.10) Initial’s answers are mostly wrong, but the modified answers are mostly correct except for one answer.

* Difficulty level: **basic**
* Runtime: **initial is 1.2, modified is 16.9(took so long)**

1. (39.11) initial is wrong. Modified has a possibly correct approach but it doesn’t finish calculating the answer till the end.

* Difficulty level: **basic**
* Runtime: **initial is 0.6, modified is 5.1**

1. (39.12) initial is wrong. Modified has a correct approach but it doesn’t show the final answer.

* Difficulty level: **basic**
* Runtime: **initial is 0.5, modified is 7.2**

1. (39.16) The initial is way off and inaccurate, while the modified one is almost correct, but not exactly, it made a minor error in one of the calculations.

* Difficulty level: **basic**
* Runtime: **initial is 0.5, modified is 7.2**

1. (39.22) The two answers are wrong.

* Difficulty level: **basic**
* Runtime: **initial is 0.6, modified is 5.68**

1. (39.23) The two answers are wrong. The modified has a valid approach but made some calculation errors.

* Difficulty level: **basic**
* Runtime: **initial is 0.93, modified is 5**

1. (39.24) The two answers are wrong.

* Difficulty level: **basic**
* Runtime: **initial is 0.7, modified is 11**

1. (39.25) The two answers are correct.

* Difficulty level: **basic**
* Runtime: **initial is 1.4, modified is 6**

1. (5.6) The two answers are wrong.

* Difficulty level: **basic**
* Runtime: **initial is 1.2, modified is 11.7**

1. (5.7.1) The two answers are wrong.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.8, modified is 5.1**

1. (5.7.2) The two answers are wrong, but the modified was almost correct, it got the real numbers correct but got the imaginary part wrong.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.69, modified is 7.5**

1. (5.7.3) The two answers are wrong. The modified has the correct structure but makes some errors on the way.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.76, modified is 4.2**

1. (5.8) The two answers are wrong.

* Difficulty level: **advanced**
* Runtime: **initial is 0.8, modified is 6.7**

1. (5.9) The two answers are wrong.

* Difficulty level: **advanced**
* Runtime: **initial is 0.56, modified is 2.2**

1. (6.12) The two answers are wrong but the modified answer was almost correct.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.79, modified is 6.2**

1. (7.6.1) The two answers are wrong.

* Difficulty level: **basic**
* Runtime: **initial is 0.59, modified is 4.3**

1. (7.6.2) The two answers are correct.

* Difficulty level: **basic**
* Runtime: **initial is 0.59, modified is 1.6**

1. (7.6.3) The two answers are correct.

* Difficulty level: **basic**
* Runtime: **initial is 0.62, modified is 2.79**

1. (7.6.4)The two answers are wrong.

* Difficulty level: **basic**
* Runtime: **initial is 0.56, modified is 1.8**

1. (9.3.1) The two answers are correct.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.63, modified is 4.29**

1. (9.3.2) The initial is wrong, but the modified is correct.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.56, modified is 6.76**

1. (9.6.1) The initial is wrong. The modified is also wrong but it was almost correct, it was on the right track before making a simplifying mistake.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.76, modified is 4.58**

1. (9.6.2) The initial is wrong. The modified is also wrong but it was almost correct.

* Difficulty level: **intermediate**
* Runtime: **initial is 1.16, modified is 4.2**

1. (9.6.3) The two answers are wrong.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.8, modified is 11.19**

1. (9.6.4) The two answers are wrong.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.7, modified is 9.5**

1. (9.8) The two answers are wrong. The modified was almost correct but it made a problem in simplification.

* Difficulty level: **advanced**
* Runtime: **initial is 0.7, modified is 9.5**

1. (10.4) The two answers are wrong.

* Difficulty level: **basic**
* Runtime: **initial is 0.7, modified is 8.2**

1. (10.6) The two answers are wrong.

* Difficulty level: **basic**
* Runtime: **initial is 1.5, modified is 7.9**

1. (11.13) The initial is wrong but the modified is correct.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.69, modified is 2.3**

1. (11.16) The initial is wrong but the modified is correct.

* Difficulty level: **basic**
* Runtime: **initial is 0.68, modified is 6.5**

1. (11.17) The two answers are wrong.

* Difficulty level: **intermediate**
* Runtime: **initial is 1, modified is 4**

1. (12.5) The two answers are correct.

* Difficulty level: **basic**
* Runtime: **initial is 0.7, modified is 2.3**

1. (12.6) The two answers are wrong.

* Difficulty level: **intermediate**
* Runtime: **initial is 1, modified is 4**

1. (12.7) The two answers are wrong.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.88, modified is 2.89**

1. (12.8) The initial is wrong but the modified is correct.

* Difficulty level: **intermediate**
* Runtime: **initial is 1, modified is 4**

1. (12.9) The initial answer got the first one right but the second wrong, the modified did the opposite.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.7, modified is 6.7**

1. (12.10) The two answers are wrong.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.75, modified is 1.78**

1. (13.6.1) The two answers are correct.

* Difficulty level: **basic**
* Runtime: **initial is 1.45, modified is 4.25**

1. (13.6.2) Both are wrong but the modified was almost correct, it just made a tiny mistake that led to the change of the negative sign to a positive.

* Difficulty level: **basic**
* Runtime: **initial is 0.78, modified is 6**

1. (13.6.3) The initial was wrong but the modified was correct and accurate.

* Difficulty level: **basic**
* Runtime: **initial is 1.08, modified is 8.26**

1. (13.7.1) The initial is wrong but the modified is correct.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.73, modified is 7.7**

1. (13.7.2) The two answers are wrong.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.6, modified is 9.1**

1. (13.9) Both are wrong.

* Difficulty level: **basic**
* Runtime: **initial is 1.4, modified is 6.5**

1. (13.10) Both are wrong.

* Difficulty level: **intermediate**
* Runtime: **initial is 1.24, modified is 6.5**

1. (13.11) Both are wrong.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.62, modified is 5.19**

1. (16.9) Both are wrong, but the modified got half of the answers right, while the initial got them all wrong.

* Difficulty level: **basic**
* Runtime: **initial is 5.09, modified is 14.49**

1. (16.11) The initial is wrong, while the modified is correct in theory, but does not show the end results.

* Difficulty level: **intermediate**
* Runtime: **initial is 3.3, modified is 11.99**

1. (16.13) Both are wrong, but the modified has correct calculation steps for the matrices but the final answer was calculated wrong in the end.

* Difficulty level: **intermediate**
* Runtime: **initial is 3.17, modified is 16.04**

1. (16.14) Both are wrong.

* Difficulty level: **intermediate**
* Runtime: **initial is 1.85, modified is 8.5**

1. (16.15.1) Both are wrong.

* Difficulty level: **basic**
* Runtime: **initial is 0.8, modified is 7.2**

1. (16.15.2) Both are wrong.

* Difficulty level: **basic**
* Runtime: **initial is 0.84, modified is 9.73**

1. (18.9.1) The initial is wrong, while the modified is correct.

* Difficulty level: **basic**
* Runtime: **initial is 2.2, modified is 10.1**

1. (18.9.2) Both are correct.

* Difficulty level: **basic**
* Runtime: **initial is 1.89, modified is 14.7**

1. (18.9.3) The initial is wrong, while the modified is correct.

* Difficulty level: **basic**
* Runtime: **initial is 4.35, modified is 5.83**

1. (18.11) Both are wrong.

* Difficulty level: **intermediate**
* Runtime: **initial is 1.54, modified is 14.66**

1. (21.14) Both are wrong.

* Difficulty level: **intermediate**
* Runtime: **initial is 1.86, modified is 9.47**

1. (21.15) The initial is wrong, while the modified only showed the (correct) method and way to solve the question but did not provide the final answer.

* Difficulty level: **advanced**
* Runtime: **initial is 3.49, modified is 6.17**

1. (21.6) The initial is completely wrong, while the modified is half correct and the other half is wrong.

* Difficulty level: **basic**
* Runtime: **initial is 0.8, modified is 10.54**

1. (24.15) The two answers are mostly correct.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.87, modified is 13.7**

1. (24.17) The two answers are wrong.

* Difficulty level: **basic**
* Runtime: **initial is 1.29, modified is 7.55**

1. (24.18) The two answers are wrong.

* Difficulty level: **intermediate**
* Runtime: **initial is 2, modified is 8.28**

1. (24.19) The two answers are wrong.

* Difficulty level: **advanced**
* Runtime: **initial is 1.07, modified is 6.94**

1. (25.8.1&2) The two answers are wrong.

* Difficulty level: **basic**
* Runtime: **initial is 1.07, modified is 6.94**

1. (27.13) The initial is wrong, while the modified only showed the steps to the solution, without the final answer.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.9, modified is 7.16**

1. (28.7.1&2) The initial is only partially correct, while the modified is completely correct.

* Difficulty level: **intermediate**
* Runtime: **initial is 1.2, modified is 6.45**

1. (28.8) Initial was wrong, while the modified only showed the essential steps and then suggested the use of a computer algebra system to calculate the final answer.

* Difficulty level: **basic**
* Runtime: **initial is 1.67, modified is 5.27**

1. (28.9) The two answers are wrong.

* Difficulty level: **intermediate**
* Runtime: **initial is 1.23, modified is 9.14**

1. (32.18) The two answers are wrong.

* Difficulty level: **advanced**
* Runtime: **initial is 0.78, modified is 6.02**

1. (32.25. 1,2,3) The two answers are wrong.

* Difficulty level: **basic**
* Runtime: **initial is 1.14, modified is 11.5**

1. (33.8) The two answers are wrong.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.75, modified is 9.8**

1. (33.9) The two answers are wrong.

* Difficulty level: **intermediate**
* Runtime: **initial is 2.38, modified is 9.46**

1. (33.1) The two answers are correct.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.67, modified is 5.66**

1. (33.4. 1,2,3,4) The initial got 2 out of 4 answers correct, while the modified got 3 out of 4 correct.

* Difficulty level: **basic**
* Runtime: **initial is 1.9, modified is 13.8**

1. (35.9) The initial is completely wrong. The modified showed the correct steps but didn’t complete till the end to find the final solution.

* Difficulty level: **basic**
* Runtime: **initial is 1.12, modified is 8.39**

1. (36.15) The initial is completely wrong. The modified showed the correct steps but didn’t complete till the end to find the final solution.

* Difficulty level: **basic**
* Runtime: **initial is 1, modified is 15.46**

**The problems i gathered**

1. The initial was wrong, even though the question is extremely easy and basic, while the modified was accurate.

* Difficulty level: **basic**
* Runtime: **initial is 0.54, modified is 1.87**

1. The initial was wrong, while the modified was accurate.(the question is very easy)

* Difficulty level: **basic**
* Runtime: **initial is 0.62, modified is 1.47**

1. Both answers are correct.

* Difficulty level: **basic**
* Runtime: **initial is 0.66, modified is 1.26**

1. Both answers are correct.

* Difficulty level: **basic**
* Runtime: **initial is 0.67, modified is 1.25**

1. Both answers are correct.

* Difficulty level: **basic**
* Runtime: **initial is 0.66, modified is 2.26**

1. The initial was wrong, while the modified was accurate.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.68, modified is 1.16**

1. Both answers are correct.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.66, modified is 2.72**

1. The initial is wrong, while the modified is correct.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.52, modified is 2.18**

1. The initial is wrong, while the modified is correct.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.59, modified is 3.93**

1. The initial is wrong, while the modified is correct

* Difficulty level: **intermediate**
* Runtime: **initial is 1.03, modified is 3.2**

1. Both are wrong, the modified one didn’t complete the answer as it requires a lot of calculation.

* Difficulty level: **advanced**
* Runtime: **initial is 0.6, modified is 9.88**

1. The initial is wrong, while the modified is correct

* Difficulty level: **advanced**
* Runtime: **initial is 0.66, modified is 6.65**

1. The initial is wrong, while the modified is correct

* Difficulty level: **advanced**
* Runtime: **initial is 0.49, modified is 5.95**

1. Both are wrong answers.

* Difficulty level: **advanced**
* Runtime: **initial is 0.91, modified is 5.6**

1. The initial is wrong, while the modified is correct

* Difficulty level: **advanced**
* Runtime: **initial is 0.73, modified is 8.5**

1. The initial is wrong, while the modified is correct

* Difficulty level: **advanced**
* Runtime: **initial is 0.71, modified is 2.7**

1. Both answers are wrong.

* Difficulty level: **basic**
* Runtime: **initial is 0.49, modified is 1.27**

1. Both answers are correct.

* Difficulty level: **basic**
* Runtime: **initial is 0.61, modified is 1.88**

1. Both answers are correct.

* Difficulty level: **basic**
* Runtime: **initial is 0.94, modified is 2.04**

1. Both answers are correct.

* Difficulty level: **basic**
* Runtime: **initial is 0.55, modified is 1.96**

1. Both answers are correct.

* Difficulty level: **basic**
* Runtime: **initial is 0.6, modified is 1.59**

1. Both have the same answer and both are wrong.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.68, modified is 3.6**

1. The initial is wrong but the modified is correct.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.8, modified is 2.14**

1. The initial is wrong but the modified is correct.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.43, modified is 4.64**

1. Both answers are correct.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.49, modified is 6.56**

1. Both answers are wrong, even though it is not an advanced logic deduction.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.65, modified is 1.8**

1. The initial is wrong but the modified is correct.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.52, modified is 1.87**

1. Both answers are wrong, even though it is not an advanced logic deduction.

* Difficulty level: **intermediate**
* Runtime: **initial is 0.65, modified is 1.74**

1. Both answers are wrong.

* Difficulty level: **advanced**
* Runtime: **initial is 0.5, modified is 3.4**

1. The initial is wrong but the modified is correct.

* Difficulty level: **advanced**
* Runtime: **initial is 0.75, modified is 5.3**

1. Both answers are wrong.

* Difficulty level: **advanced**
* Runtime: **initial is 0.67, modified is 2.59**

1. Both answers are wrong.

* Difficulty level: **advanced**
* Runtime: **initial is 0.66, modified is 3.5**