You are based in the United Kingdom, and supervise a Master thesis investigating syntactic ambiguity resolution in native speakers vs. intermediate (i.e. a proficiency level of B1) L2 learners of English. The self-paced reading experiment is based on a comparison between temporarily-ambiguous sentences such as (1a) and otherwise very similar disambiguated counterparts such as (1b) (based on Sturt, 2007):

(1a) The explorers found the south pole was impossible to reach.

(1b) The explorers found that the south pole was impossible to reach.

You ask your student to compare reading times for the disambiguating segment (in the above example, the segment "*was impossible"*). The student initially finds it quite challenging to recruit a large-enough sample of such relatively low-proficiency L2 learners in the UK, but eventually shows you an analysis for a sample consisting of 32 native English speakers and 32 L2 learners of B1 proficiency. Mean reading times for the disambiguating segment are shown in Table 1.

Table 1. Mean reading times (and standard deviations) for the disambiguating segment by condition and participant group in the self-paced reading task.

	L1 group (N=32, mean age=23.2; SD=2.1)	L2 group (N=32, mean age=25.2; SD=8.1)
Ambiguous	530(94)	505(49)
Unambiguous	636(102)	635(45)

As the student is not familiar with linear-mixed effects models, you allow them to analyse the data with traditional frequentist statistics. The student provides you with Table 2, which contains the results from a by-participants repeated-measures ANOVA[1]:

Table 2. By-participants repeated-measures ANOVA.

Ambiguity	F(1,62)=45.305, **p<.001**
Group	F(1,62)=0.987, p>.1
Ambiguity:Group	F(1,62)=1.013, p>.1

In addition, the student also provides by-participants t-tests comparing the effect of ambiguity within each of the two participant groups:

Table 3. By-participants t-tests for the effects of ambiguity within each participant group.

Effect of ambiguity in L1 group	t(31)=7.31, **p<.001**
Effect of ambiguity in L2 group	t(16)=7.82, **p<.001**

When asked for an interpretation of the data pattern, the student suggests that both groups show a similar difference between the ambiguous and unambiguous conditions.

1) Do you notice anything strange or unusual in the analysis presented above? Is there anything in there that would make you suspicious? Make a list of your observations.
2) Based on the sum of all the unusual things you have noticed, can you come up with an idea of what exactly may have gone wrong here?
3) The issues discussed above aside, do you think the presentation of the results could be improved in any way to make it easier to understand the analyses (particularly for readers who are less experienced with experimental research?
4) Finally, if we assume that the issues discussed in 1) and 2) are corrected and the changes to the presentation discussed in 3) are implemented, do you still see any

aspects of the analysis which are not in principle incorrect, but could nonetheless still be improved upon?

[1] As a side note, a proper analysis would obviously also have to contain the corresponding by-items (F_2) ANOVAs in addition to the by-participants (F_1) ANOVAs, to check whether the effect also generalizes from the sample of experimental sentences to the population of all sentences with the structure of interest. I somewhat simplified things here and have left these F_2-ANOVAs out, just because this is not what the exercise is all about.