



# ChatGPT is not a pocket calculator

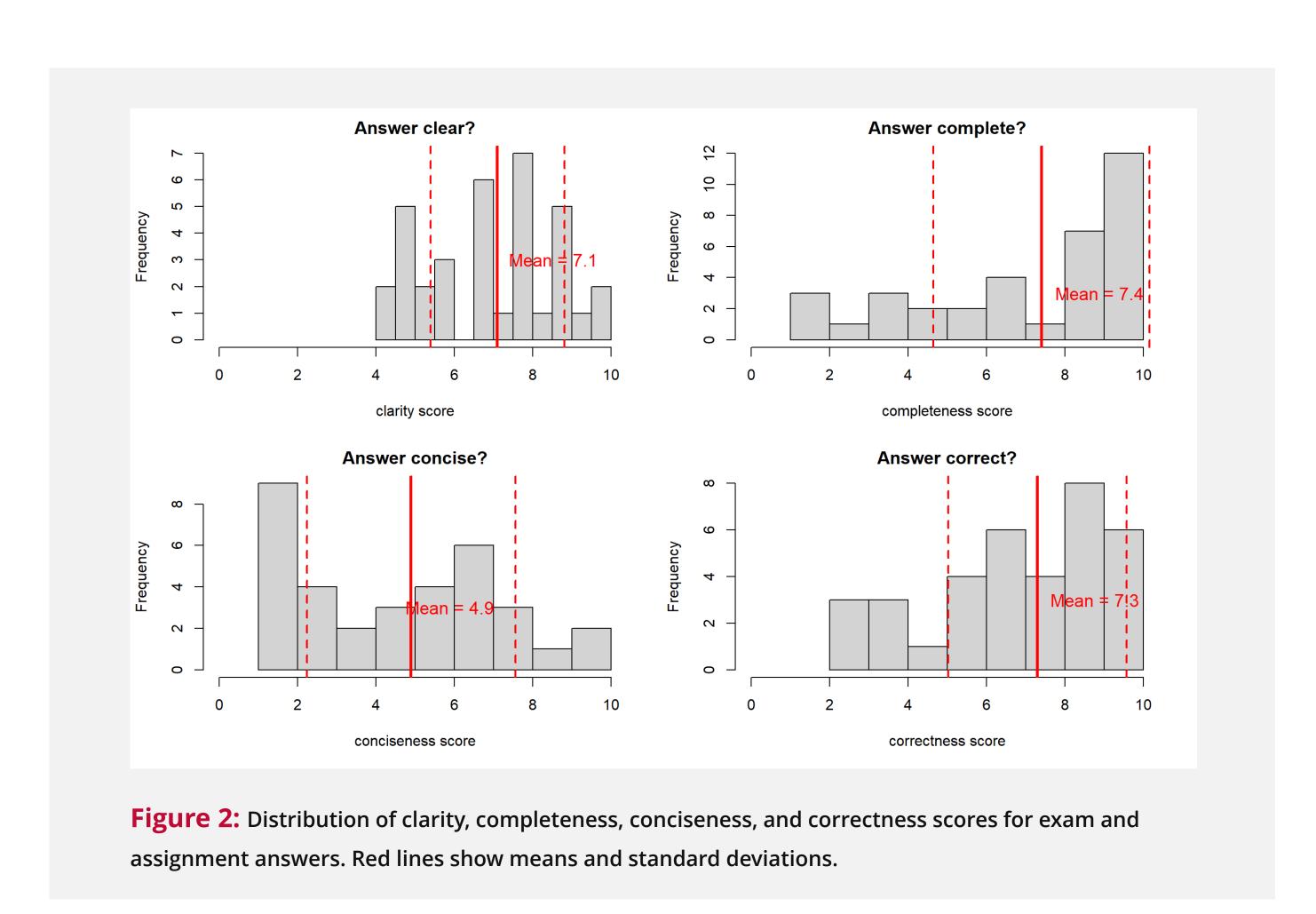
Problems of AI-chatbots for teaching Geography

Simon Scheider (s.scheider@uu.nl), Harm Bartholomeus, Judith A. Verstegen

### Introduction

The recent success of large language models and AI chatbots such as ChatGPT has a severe impact on teaching and learning Geography and GIScience. The underlying revolution is often compared to the introduction of **pocket calculators**, suggesting that **lower-level learning goals are substitutable by AI**, and supervision and assessments can be **refocused on higher-level goals** (Figure 1, right).

However, the success of such a strategy rests on the assumption that ChatGPT does not interfere with the higher-level learning goals; if it does, the use of ChatGPT is **fraudulent** because it threatens the validity of assessments (Figure 1, left).



# "Fraud and plagiarism are defined as an action or omission on the part of students which produces an incorrect representation of their own performance as regards their knowledge, skills and understanding, which may result in the examiner no longer being able to assess the knowledge or ability of the students in a proper and fair manner." Article 'Fraud and Plagiarism', Paragraph 5 of the Education and Examination Regulations. Bloom's Taxonomy Produce new or official work Longitude new or official work

### **Methods & Results**

We tested this assumption by asking university teachers in Geography and GIScience about ChatGPT's quality in answering their exam and assignment questions. Hereto, we ran an online survey among teachers at Utrecht University and Wageningen University (NL) and got 41 valid survey responses, by 15 females, 26 males.

Our preliminary results indicate that **ChatGPT will likely pass Geography and GIScience exams** and assignments in its present form (Figure 2); this is **independent of the learning-level** (Bloom's taxonomy) (Figure 3). As such, even assessing only (true) academic skills does not help, as ChatGPT masters these too.

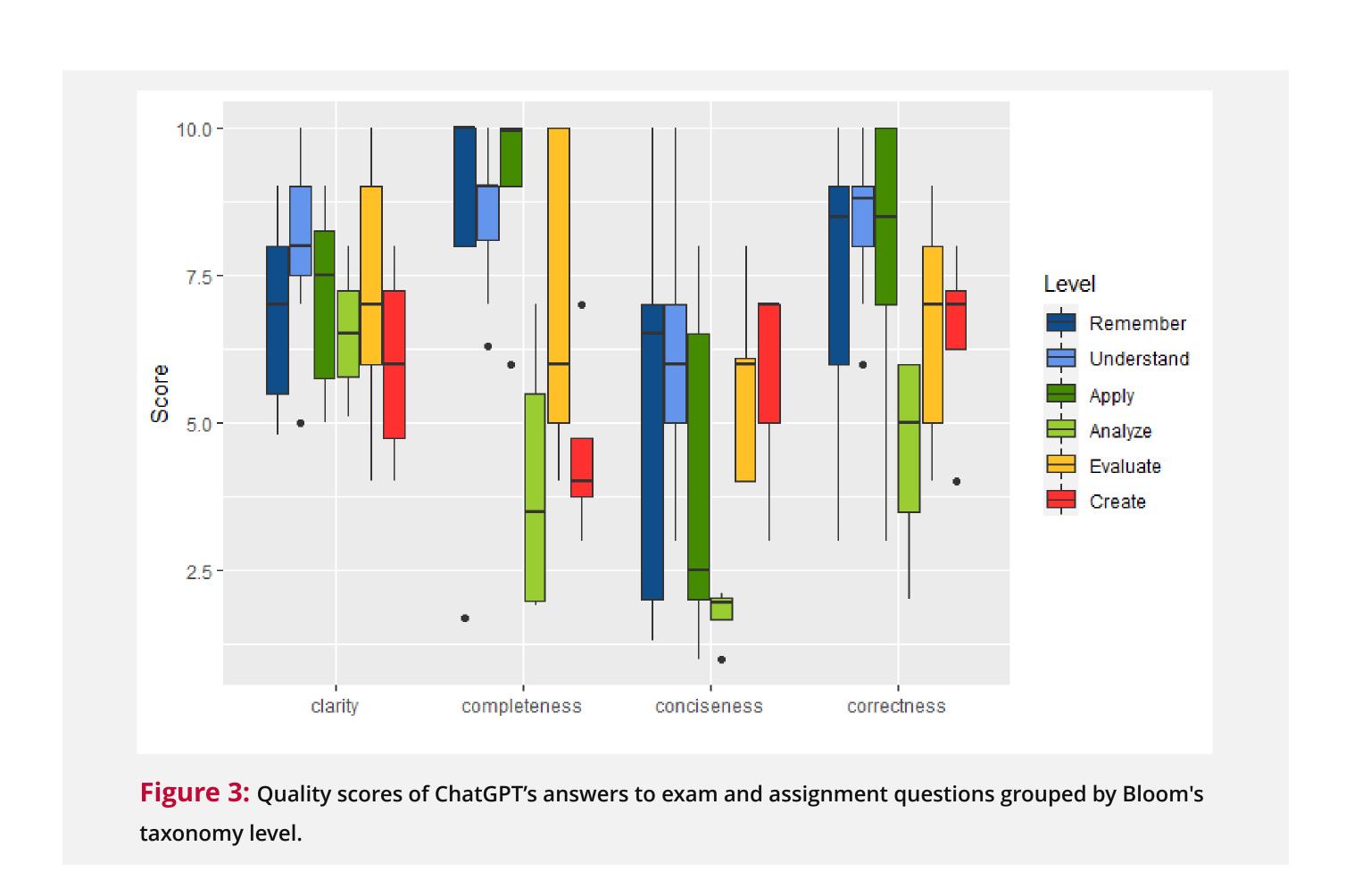
Since the skills underlying these higher levels are not substitutable when learning and thus require supervision, the introduction of ChatGPT is not comparable to the introduction of the pocket calculator. This means that we are forced to change Geography and GIScience assessments in one way or another.

# **Conclusions & Recommendations**

What to do about potentially invalid assessments?

- 1. Adapt the **learning goals**;
- 2. Control the assessment **environment**;
- 3. Control and assess the **learning process** instead of the obtained skills; or
- 4. Forbid the use of chatbots and check fraud **a-posteriori** and punish in case of non-compliance

Our analysis does not imply that chatbots cannot be beneficial for teaching at all. ChatGPT offers great assistance in, e.g., scripting tasks and may replace browsing through help-functions or personal assistance. Yet, including it in any product that is meant to assess whether a student has achieved a learning goal introduces **a major risk for the validity of assessment**.



# Reference

Scheider, S., Bartholomeus, H., Verstegen J.A. (2023). ChatGPT is not a pocket calculator - Problems of Al-chatbots for teaching Geography. arXiv:2307.03196. DOI: 10.48550/arXiv.2307.03196.