

# Verbal, Visual, Valuable: Making Effect Sizes Understandable for Teachers

Florian Kühlwein<sup>1</sup>, Samuel Merk<sup>1</sup>, Jürgen Schneider<sup>2</sup>, Kirstin Schmidt<sup>1</sup>

✉ florian.kuehlwein@ph-karlsruhe.de ✉ samuel.merk@ph-karlsruhe.de ✉ ju.schneider@dipf.de ✉ kirstin.schmidt@ph-karlsruhe.de

<sup>1</sup>Karlsruhe University of Education, <sup>2</sup>DIPF | Leibniz Institute for Research and Information in Education



## Relevance & Rationale

- Teachers should consider and implement findings from empirical educational research (EU, 2007)
- For effective instruction, teachers need to have a rough estimate of the effectiveness of educational interventions
- Trying to communicate these effect sizes often leads to inaccurate estimations or misconceptions (Hanel & Mehler, 2019; Kim et al., 2022; Lortie-Forgues et al., 2021; Schmidt et al., 2023; Zhang et al., 2023)
- The effects of communicating effect sizes verbally vs. visually have hardly been investigated in direct comparison
- RQ1:** How accurate, informative, difficult, and relevant are verbally and visually designed effect sizes perceived by teachers?
- RQ2:** Are there any differences between the presentation modes regarding DV1 - DV4?

## Design

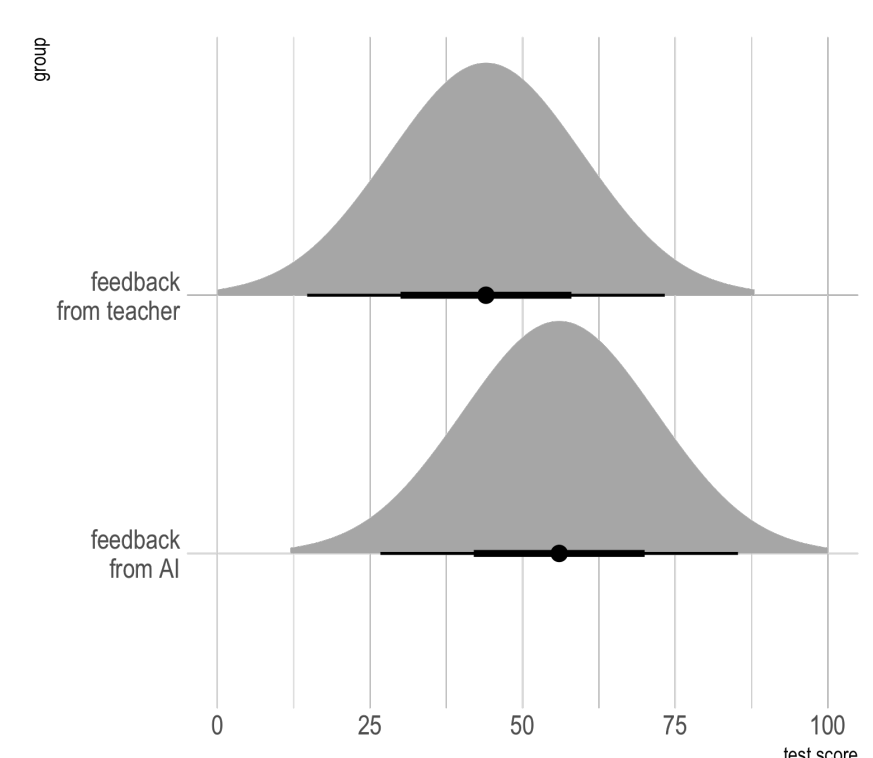
- 2 x 6 within-person design
- Independent variables:** Presentation Mode (**Verbal** vs. **Visual**) & Effect Size
- Dependent variables:**
- Probability of superiority (PoS): *»How many times out of 100 do you estimate that a randomly selected member of the AI tutor group would have a higher score in the reading test than a randomly selected person from the teacher feedback group?«*
- Perceived informativeness: *»How informative do you perceive the way the information is presented?«*
- Perceived difficulty: *»How difficult was it for you to understand the graph/text?«*
- Relevance: *»How much are you willing to spend on an AI reading tutor license for a class of 30 students?«*

## Independent variable 1: Presentation Mode

**Verbal:**  
Cohen's  $U_3$  expressions

*»78.8% of the students who practiced reading with the AI tutor scored higher on the reading test than the average score of the group who practiced reading with the help of their teacher.«*

**Visual:**  
Half-eye plot



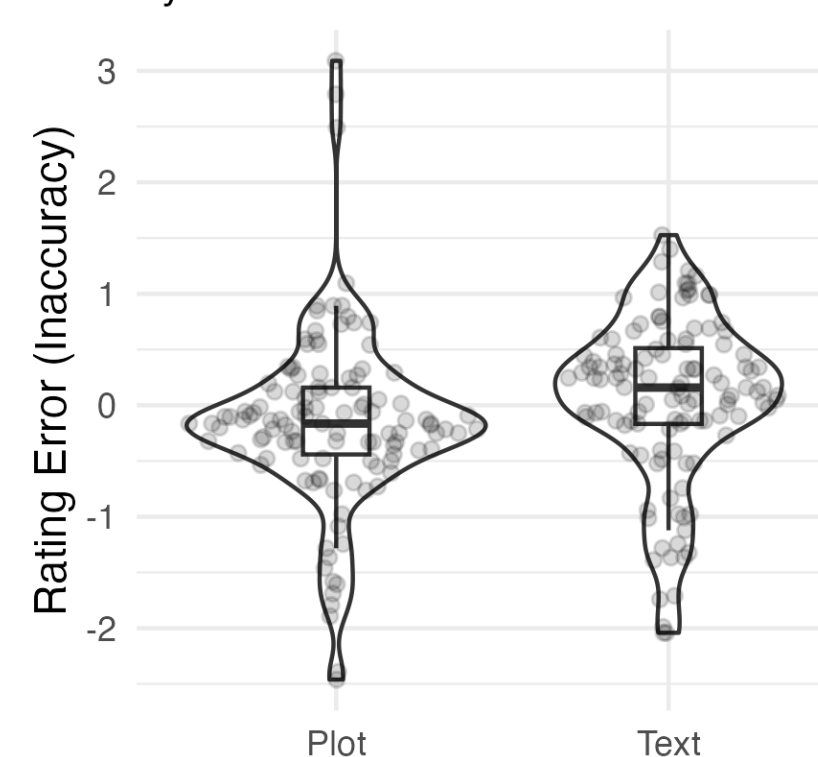
## Independent variable 2: Effect Size

Six different effect sizes: -0.8, -0.5, -0.2, 0.2, 0.5, 0.8

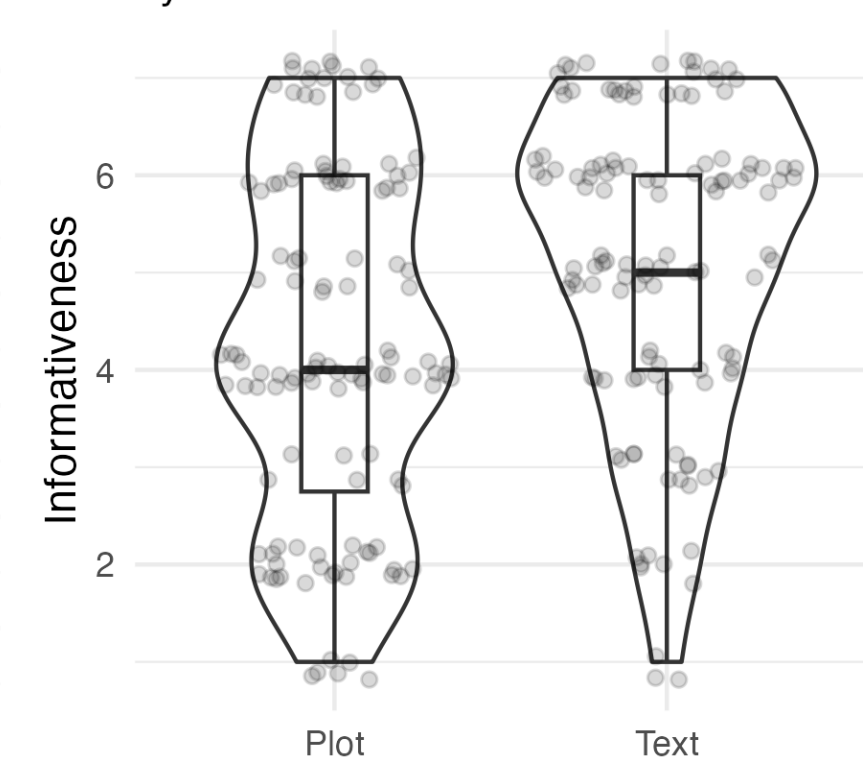
## Pilot Study

- $N = 20$  English-speaking teachers from the USA & UK
- Result a):** Teachers perceive Cohen's  $U_3$  expressions as more informative (Cliff's delta = -0.27, 95% CI [-0.40, -0.13]) and less difficult to understand than half-eye plots (Cliff's delta = -0.31, 95% CI [-0.44, -0.17])
- Result b):** The PoS/accuracy scores for the verbal mode are lower than for the visual mode (Cohen's  $d = -0.26$ , 95% CI [-0.51, 0.00])
- Bayesian estimation provides evidence for substantial effects for DV1-DV3 (Kruschke, 2018)

Accuracy by Presentation Mode



Informativeness by Presentation Mode



References



Data & Stimulus

