Comparing Assessments of Graph Comprehension
Hannah S. Lloyd\textsuperscript{a}, Holly Huey\textsuperscript{a}, Erik Brockbank\textsuperscript{a}, Lace Padilla\textsuperscript{b} & Judith E. Fan\textsuperscript{a,c}
\textsuperscript{a}University of California San Diego, \textsuperscript{b}Northeastern University, \textsuperscript{c}Stanford University

Overview

Current approaches to graph comprehension:

- **“GGR”**
  - level 1: read the data
  - level 2: read between the data
  - level 3: read beyond the data

- **“VLAT”**
  - retrieve value
  - find trends & correlations
  - find clusters

Graph comprehension is important.

How reliable are existing assessments for measuring it?

<table>
<thead>
<tr>
<th>number of items</th>
<th>13</th>
<th>53</th>
</tr>
</thead>
</table>

How well does performance on one test predict performance on the other?

GGR and VLAT scores are moderately correlated

Scores for individual question types are more strongly correlated within test than between test

What grouping of test items best predict observed error patterns?

Small number of latent factors better explains the error patterns than test, question, or graph type

What is the relationship between formal math training and graph comprehension?

Positive relationship between formal math/stats training and performance in both samples

Takeaways

Our findings suggest that graph comprehension encompasses a suite of capabilities that do not cleanly correspond to graph or task.

More work is needed to develop reliable and valid assessments of graph literacy that predict response patterns.

---

Factors

- Factor 1
  - high reliability between U.S. university and U.S. representative samples

- Factor 2
  - high reliability between U.S. university and U.S. representative samples

- Factor 3
  - high reliability between U.S. university and U.S. representative samples

Study Design

- 2 assessments
  - "GGR" & "VLAT"

- 11 question types
  - i.e., retrieve value, find clusters

- 13 graph types
  - i.e., bar chart, line chart, histogram

- 66 items
  - i.e., retrieve value & line chart

n = 1,140

University students and U.S. demographically representative

---

**How well does performance on one question type predict performance on the others?**

- Average within-test corr = 0.452 95% CI (0.40, 0.443)

- Average between-test corr = 0.315 95% CI (0.314, 0.316)

---

**Graph comprehension toolslab/gcb_public2023**

data & materials will soon be available:

https://github.com/cogtoolslab/gcb_public2023