IES Institute of Education Sciences

Examining the Differences in Item Revisit Patterns across Disability Groups

Susu Zhang, University of Illinois at Urbana-Champaign
Xin Wei, Stanford Research Institute
Jihong Zhang, University of Iowa

Annual IES Principal Investigators Meeting
January 25-27, 2022
Advancing Equity and Inclusion in the Education Sciences

I am an Assistant Professor of Quantitative Psychology and Statistics at UIUC.

My research pertains to quantitative methods for solving practical problems in testing.

In particular, I am interested in the use of problem-solving **process data** to inform measurement and learning.

Introduction

- NAEP math assessment data show that **learners with disabilities continuously exhibit much poorer math performance** than their peers without disabilities.
- Our study uses the restricted-use process data from the NAEP 2017 Math Assessment to understand the difference in **review and revision patterns across disability groups.**

NAEP 2017 Mathematics Process Data

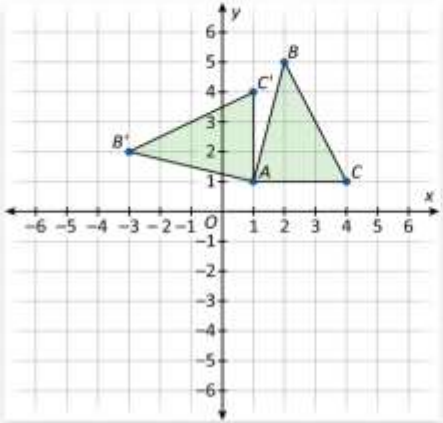
- Largest nationally administered low-stakes test that represents student achievement in the United States.
- The 2017 NAEP grade 8 mathematics assessment included two 30-minute blocks of cognitive items in each **digital test form**.
- Process data refers to the **ordered sequence of observed events** that an examinee executes in pursuit of solving a problem.
- E.g., keystrokes, clicks, use of UDL tools
- Here, we restrict the discussion to **revision and review process**.

Example Item

VH287980
1717MA2N03CUD30EX

Time Left: 30 minutes

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 Review



Which of the following transformations in the *xy*-plane will map triangle *ABC* onto triangle *AB′C′* ?

- A ○ A 90° clockwise rotation about point *A*
- B ○ A 90° clockwise rotation about the origin
- C ○ A 90° counterclockwise rotation about the origin
- D ○ A 90° counterclockwise rotation about point *A*
- E ○ A 20° counterclockwise rotation about point *A*

Clear Answer

eNAEP Version: N/A
Device: Unknown
Specification: N/A
Browser Version: k-webkit55
eNAEP: Database:
Chrome Extension: 3.1.3
Date Recorded: 2017-01-26

Example of Revision/Review Process

Action	Timestamp
Enter_Item	0.0
Choose_right	21.1
Clear_Answer	35.5
Choose_wrong	43.6
Exit_Item	52.0
Enter_Item	62.2
Exit_Item	70.2

Final Outcome

Final score: 0

Total response time: 60 seconds

Additional Information in Revision Process

Hesitation

Those who are unsure about the correct response may be more likely to exhibit answer change.

Test navigation style

E.g., Does the student respond to the question on the first visit? Does the student go back to earlier questions, e.g., to review?

Research Questions

- For NAEP 2017 Math multiple choice questions:
  - RQ1: Are there different types of review and revision behavior on the same item?
  - RQ2: Do learners with learning disability (LD), autism spectrum disorder (ASD), and their typically developing (TD) peers differ in revision and review behavior?

Structure of Process Data

Variable length

Different examinees can differ in sequence length

E.g.,

- Examinee 1:

Enter_Item, Choose_right, Choose_wrong, Choose_right, Exit_Item

- Examinee 2:

Enter_Item, Choose_right, Exit_Item

Categorical

Each action is sampled from the set of all possible actions

E.g.,

{Enter_Item, Exit_Item, Choose_right, Choose_wrong, Clear_Answer}

{Enter_Item, Exit_Item, Part1_right, Part1_wrong, Part2_right, Part2_wrong, Clear_Answer}

Ordered

Temporal ordering of actions is substantively meaningful.

E.g.,

- Preview:

Enter_Item, Exit_Item, Enter_Item, Choose_right, Exit_Item

- Review:

Enter_Item, Choose_right, Exit_Item, **Enter_Item, Exit_Item**

Clustering of Revision/Review Log Data

- To answer RQ1 (types of review/revision behavior), we apply **hierarchical agglomerative clustering** (Ward, 1963) to the revision log, i.e., the sequence of revision and review-related actions:
  - Click choice: indicating answer selection; recoded to right/wrong answers for each question/subquestion.
  - Clear Answer: removing previous answers.
  - Enter/Exit: indicating entering and leaving an item. Visits shorter than 3 seconds were removed.
- Individuals in the same cluster have lower **dissimilarity** in their revision log than those in different clusters.

Sequence Dissimilarity Measure

Order-based Sequence Dissimilarity (Gómez-Alonso & Valls, 2008; Tang et al., 2020):

- For two examinees *i* and *j*, the pairwise dissimilarity of their revision log (s_i, s_j) is

$$d_{ij} = \frac{f(\mathbf{s}_i, \mathbf{s}_j) + g(\mathbf{s}_i, \mathbf{s}_j)}{L_i + L_j}$$

where

$$f(\mathbf{s}_i, \mathbf{s}_j) = \frac{\sum_{a \in C_{ij}} \sum_{m=1}^{K_{ij}^a} |s_i^a(m) - s_j^a(m)|}{\max\{L_i, L_j\}},$$ ordering difference in common actions

$$g(\mathbf{s}_i, \mathbf{s}_j) = \sum_{a \in U_{ij}} L_i^a + \sum_{a \in U_{ji}} L_j^a.$$ number of unique actions to each sequence

Data

- 5640 individuals who completed the 30-minute version of NAEP Grade 8 Math restricted-use partial form block.
  - 5000 typically developing (TD) learners (sampled from 20000+)
  - 570 learners with learning disability (LD)
  - 70 learners in Autism Spectrum Disorder (ASD)
- The block contains 15 items in total. Only five multiple choice items were included.
  - 2 w/ single part and correct option
  - 2 w/ multiple (3 or 4) parts and single correct option
  - 1 w/ multiple (4 out of 6) correct responses

Number of Clusters Identified

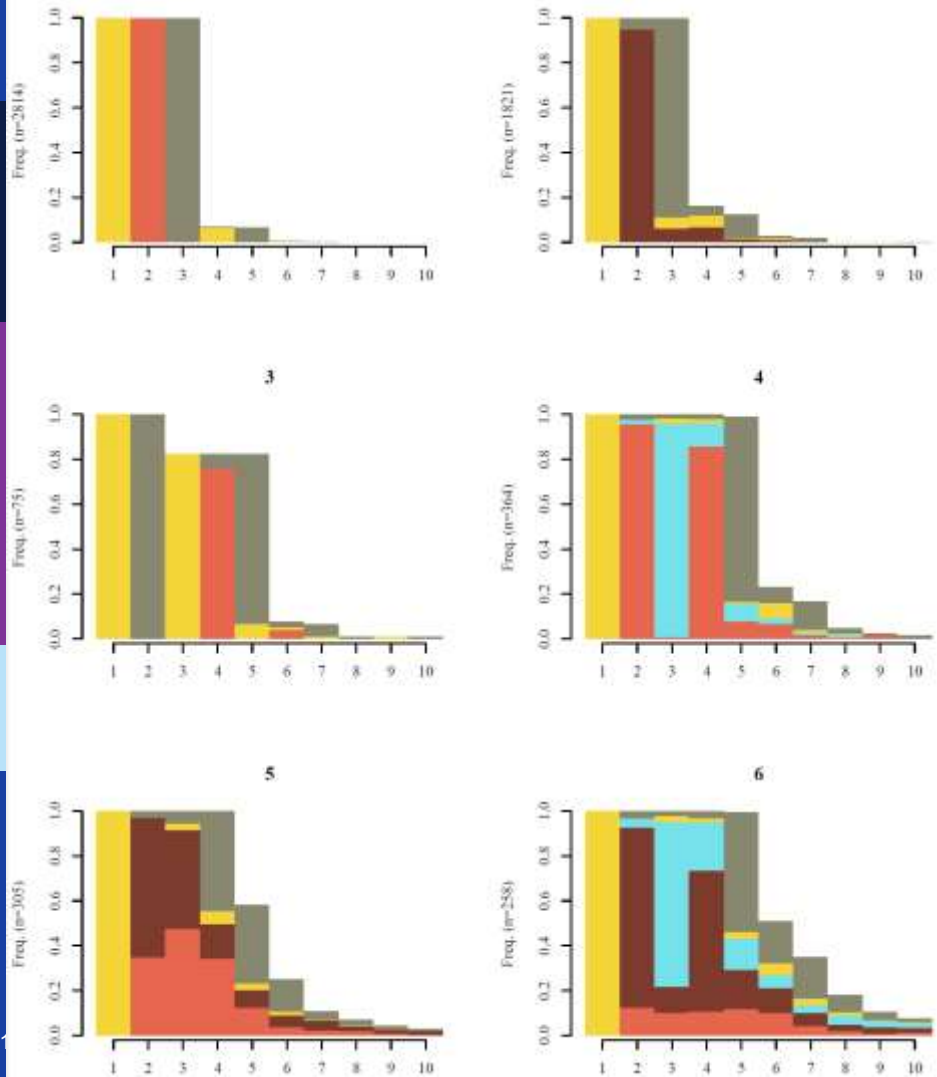
- Number of clusters were chosen based on a combination of the average silhouette index (Rousseeuw, 1987), the Dunn index (Dunn, 1974), the C-index (Hubert and Levin, 1976), and clarity of interpretations.

Item	VH266695	VH263651	VH287980	VH302900	VH271388
Position	1	5	8	10	12
# Clusters	6	8	7	8	15

Institute of Education Sciences

1

Single-part, single-select items



Institute of Education Sciences

Sequence distribution plots for each of the 6 clusters on VH266695:

chs_right	Clear_Answer	Exit_Item
chs_wrong	Enter_Item	NA

Clusters 1 and 2:

Firm correct and incorrect responses.

Cluster 3:

Omitting and Previewing before correct response

Cluster 4:

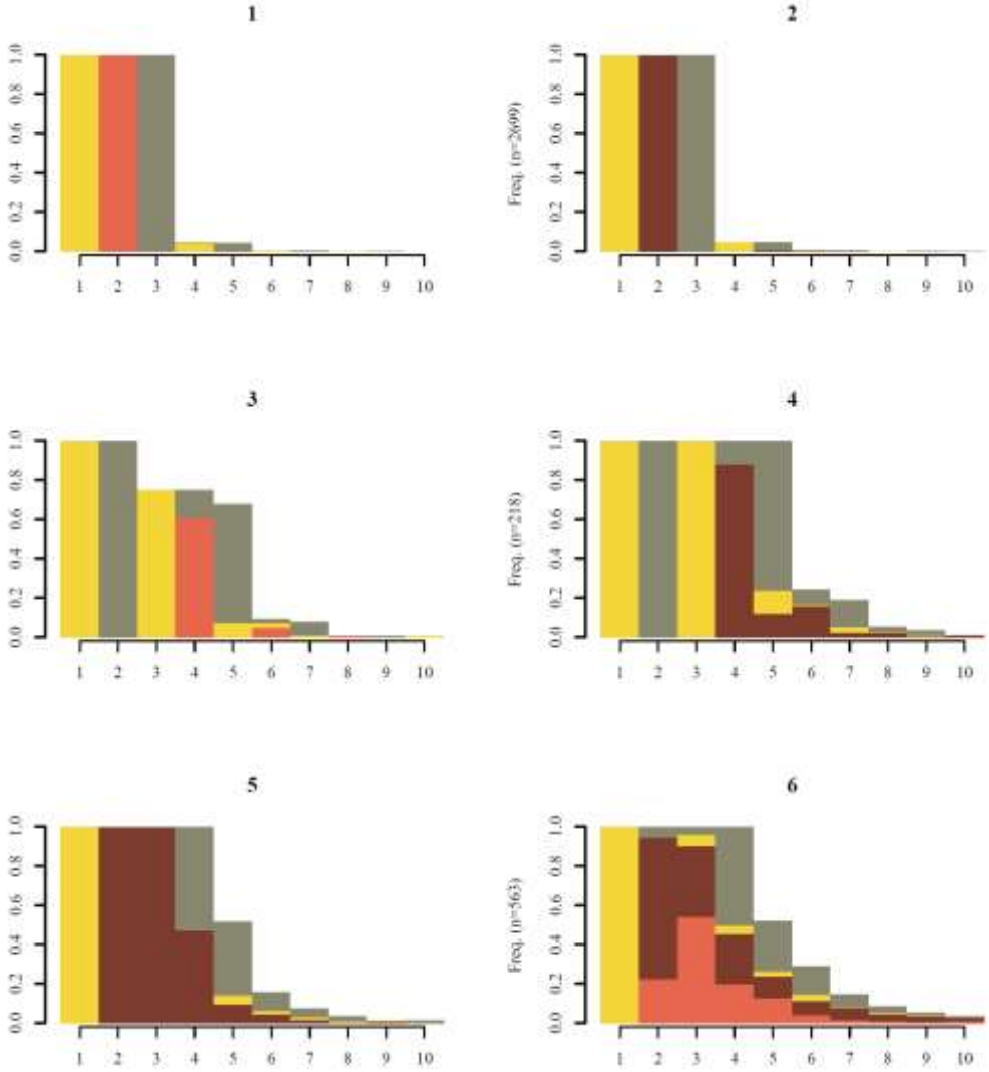
Exploration of "**Clear Answer**" button

Cluster 5:

Alternation between right/wrong choices

Cluster 6:

Alternation + Clear answer



Sequence distribution plots for the 7 clusters on VH266695:

Patterns very similar, w/ a few exceptions:

1. A preview + incorrect response cluster emerged.
2. An alternation between incorrect choices cluster emerged.
3. “Clear Answer” exploration cluster no longer present

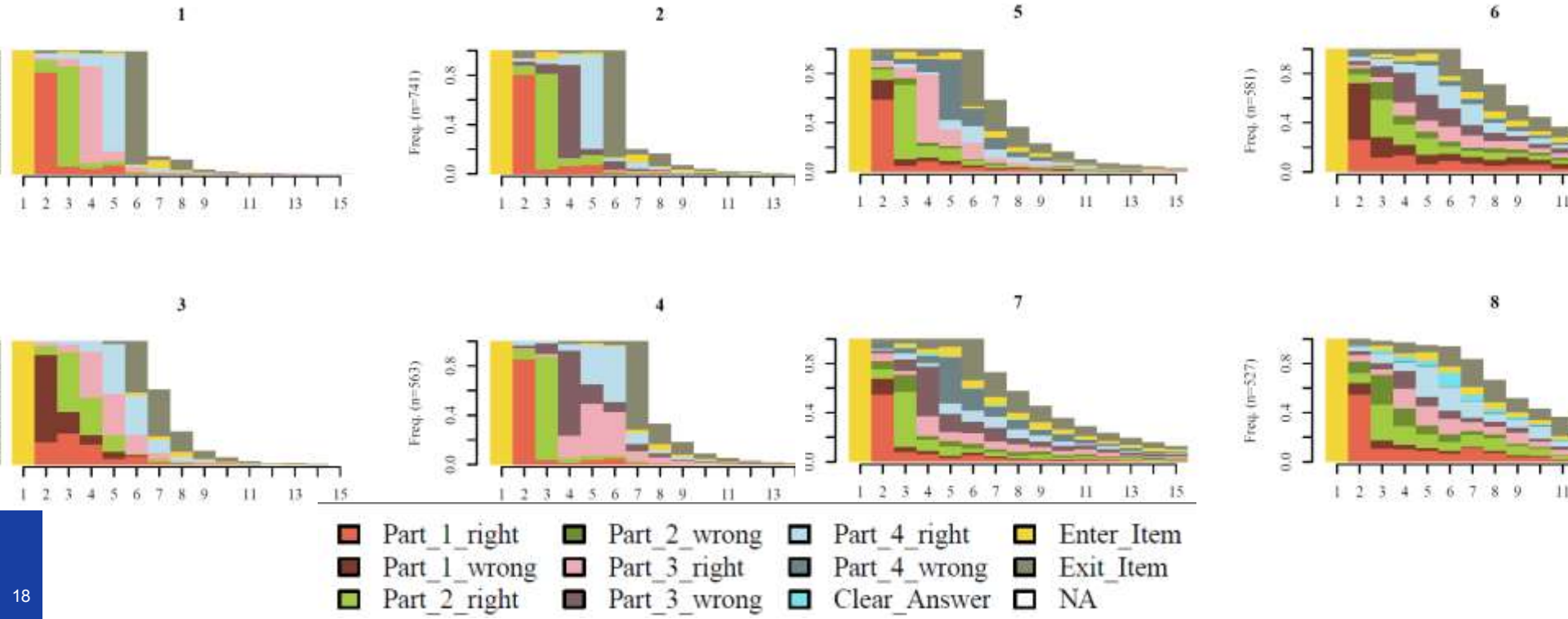
Institute of Education Sciences

2

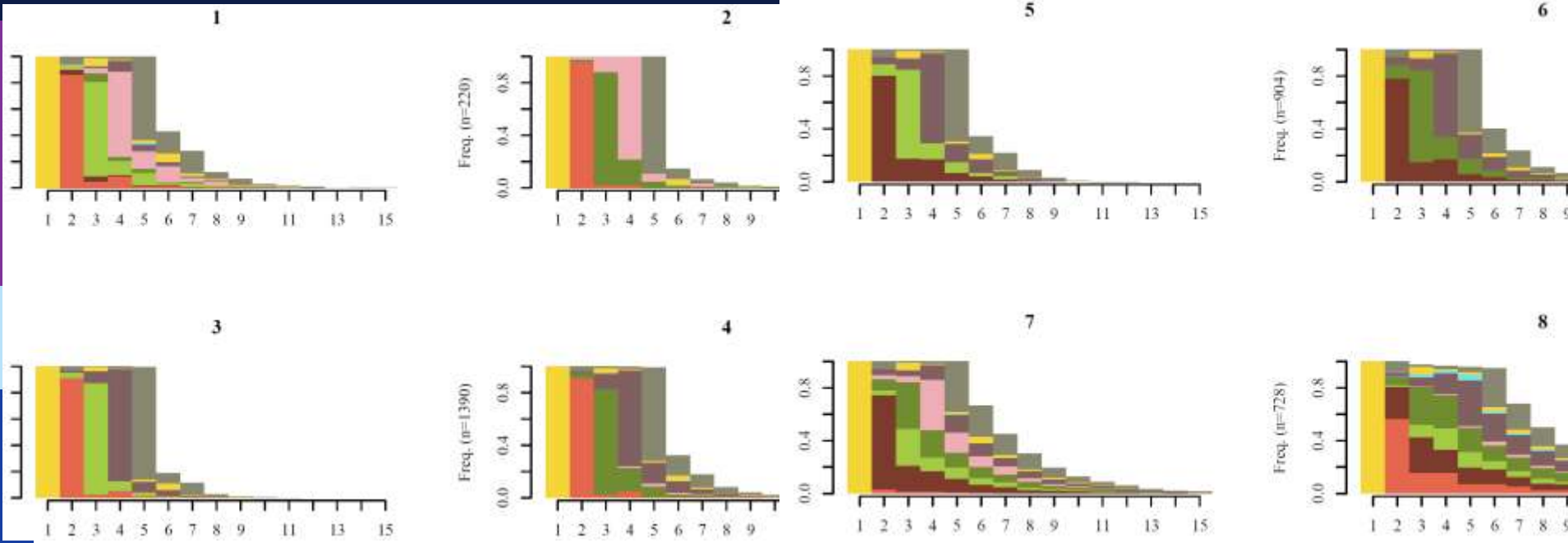
Multiple-part, single-select items

8 clusters on VH263651

Different clusters reflect mistake/hesitation on different parts of a question, and usage of clear answer



8 clusters on VH302900



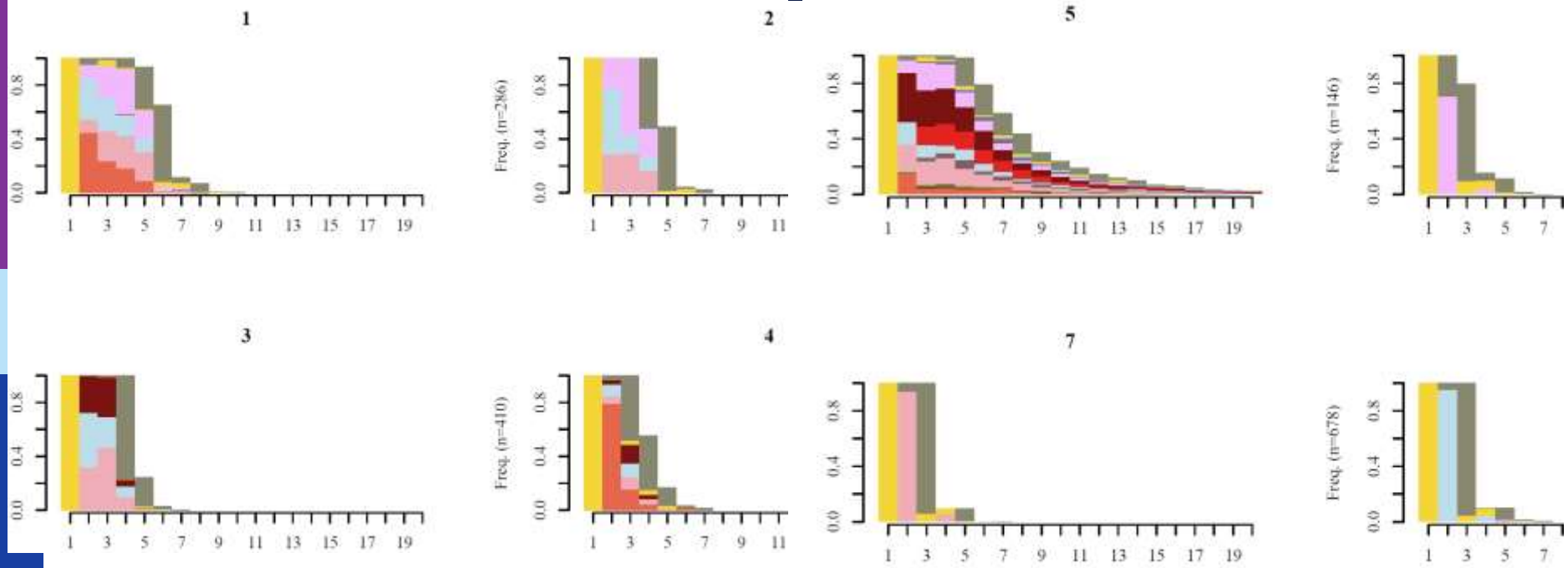
Institute of Education Sciences

3

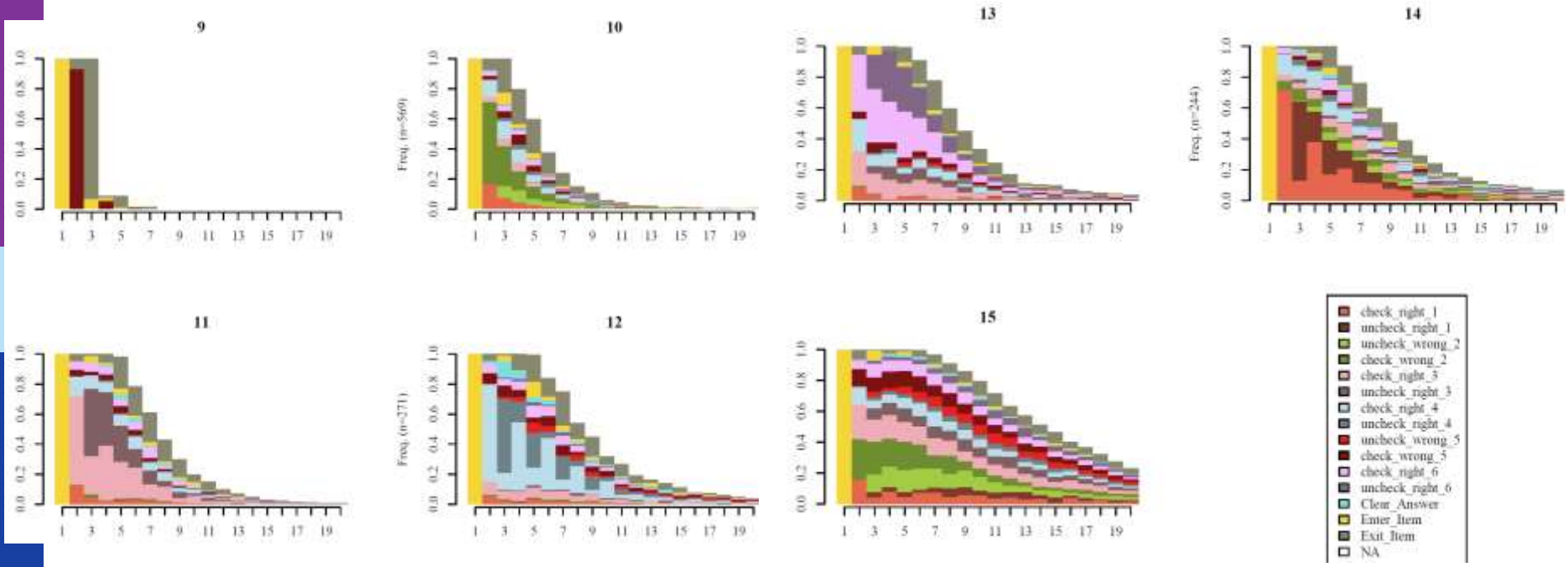
Multiple-select items

15 clusters on VH271388

Different clusters reflect missing certain correct options, mistake/hesitation on different options, and usage of clear answer



15 clusters on VH271388 (ctd.)



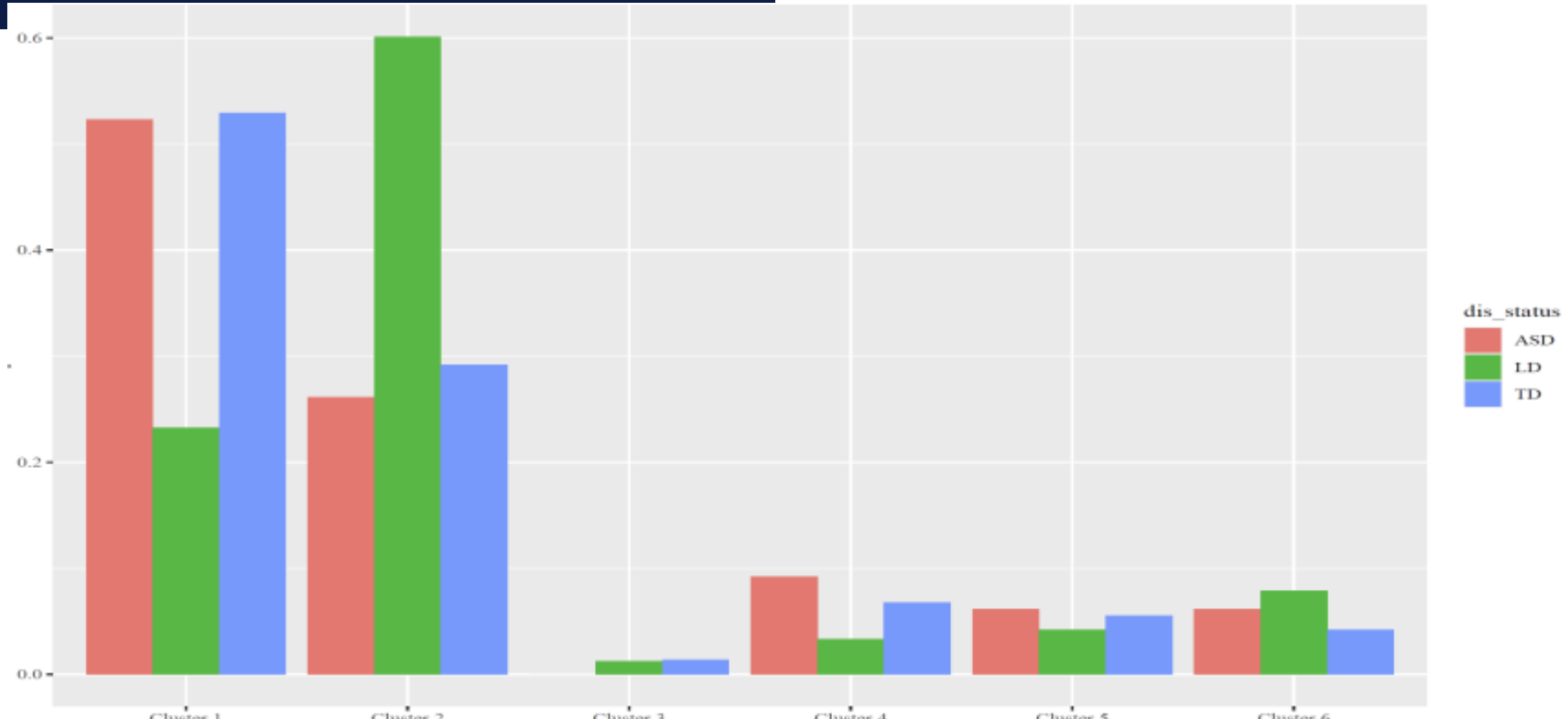
Institute of Education Sciences

4

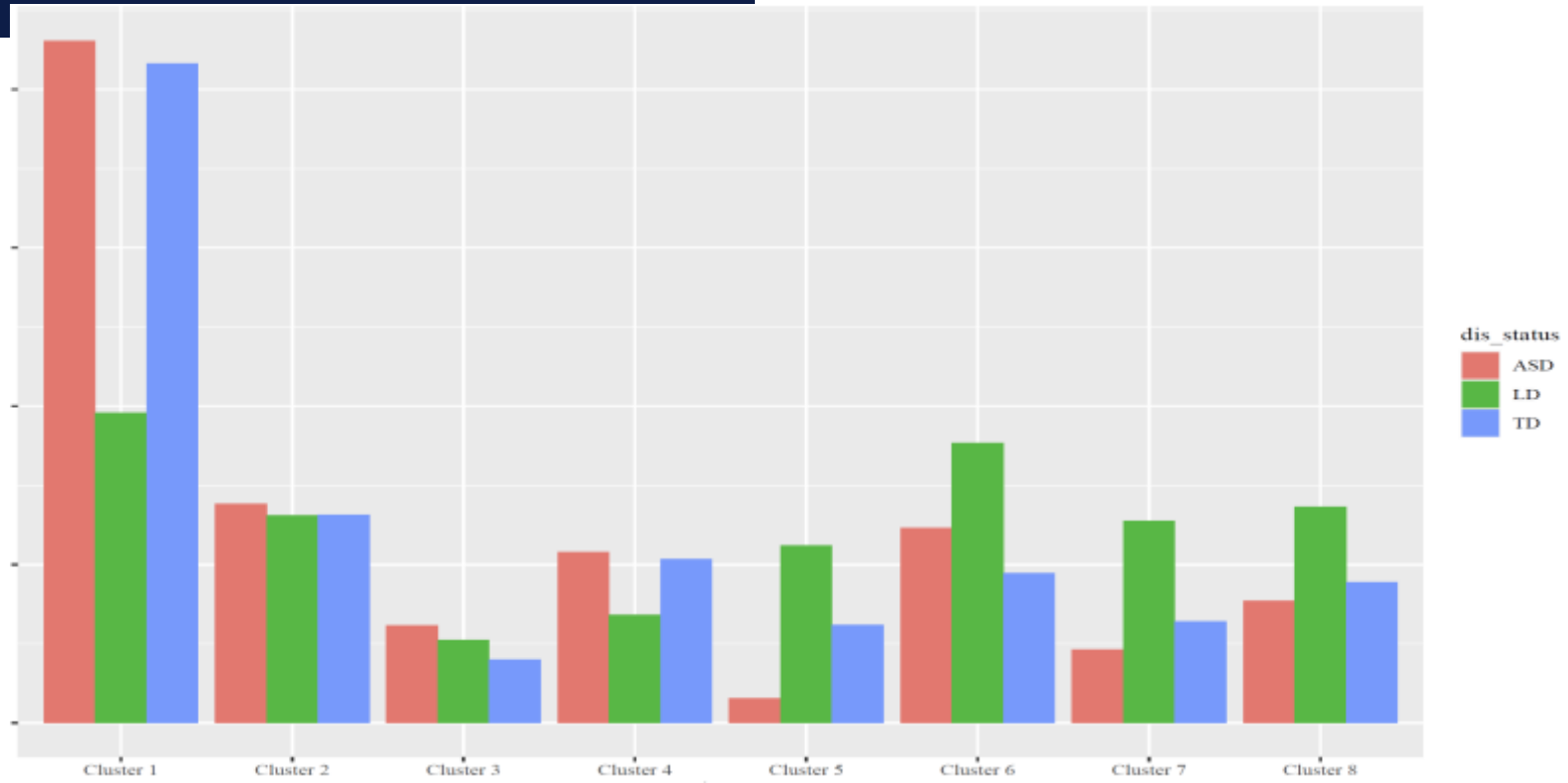
Group differences in Cluster Distribution

Group difference in cluster distribution: VH266695

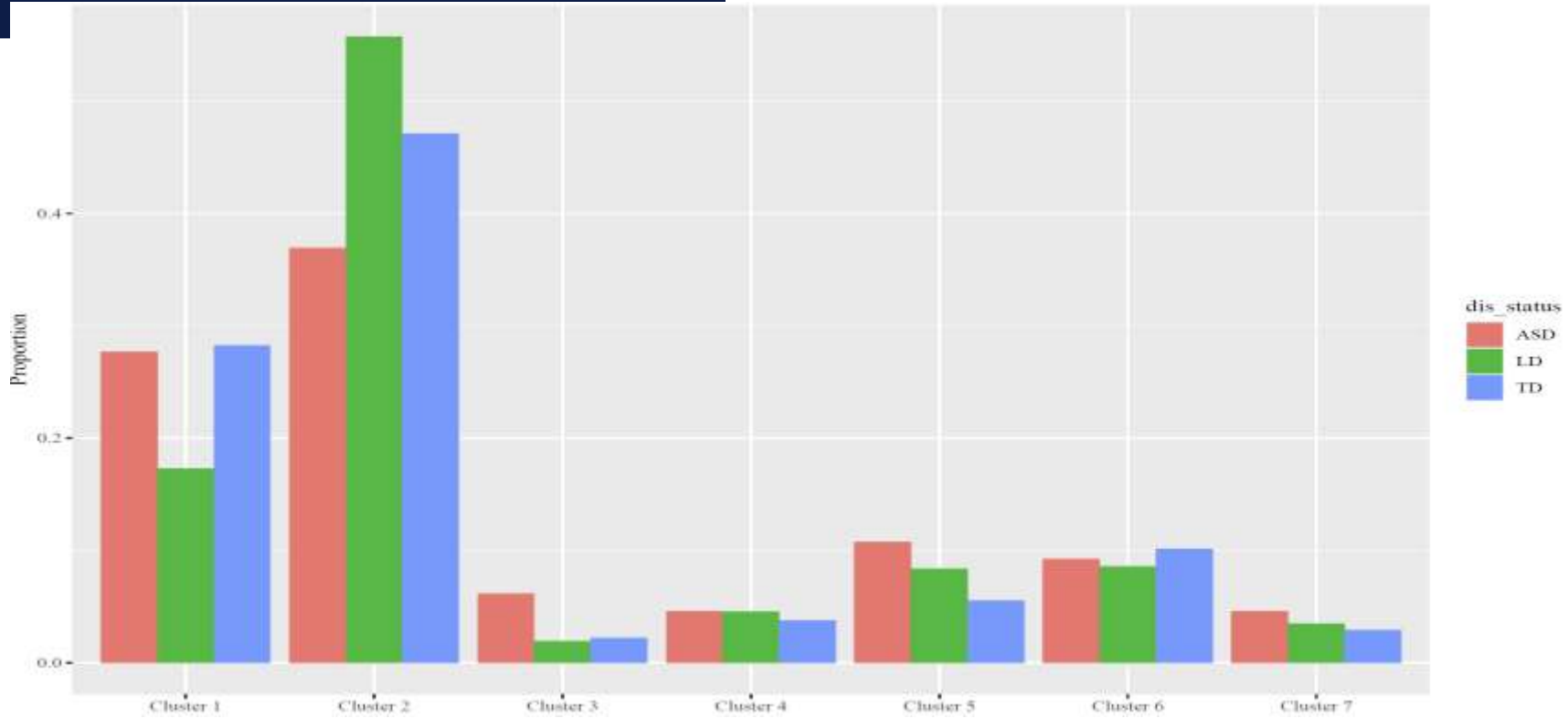
Across all 5 items, chi-squared test of independence between disability status and cluster membership rejected the independence null hypothesis.



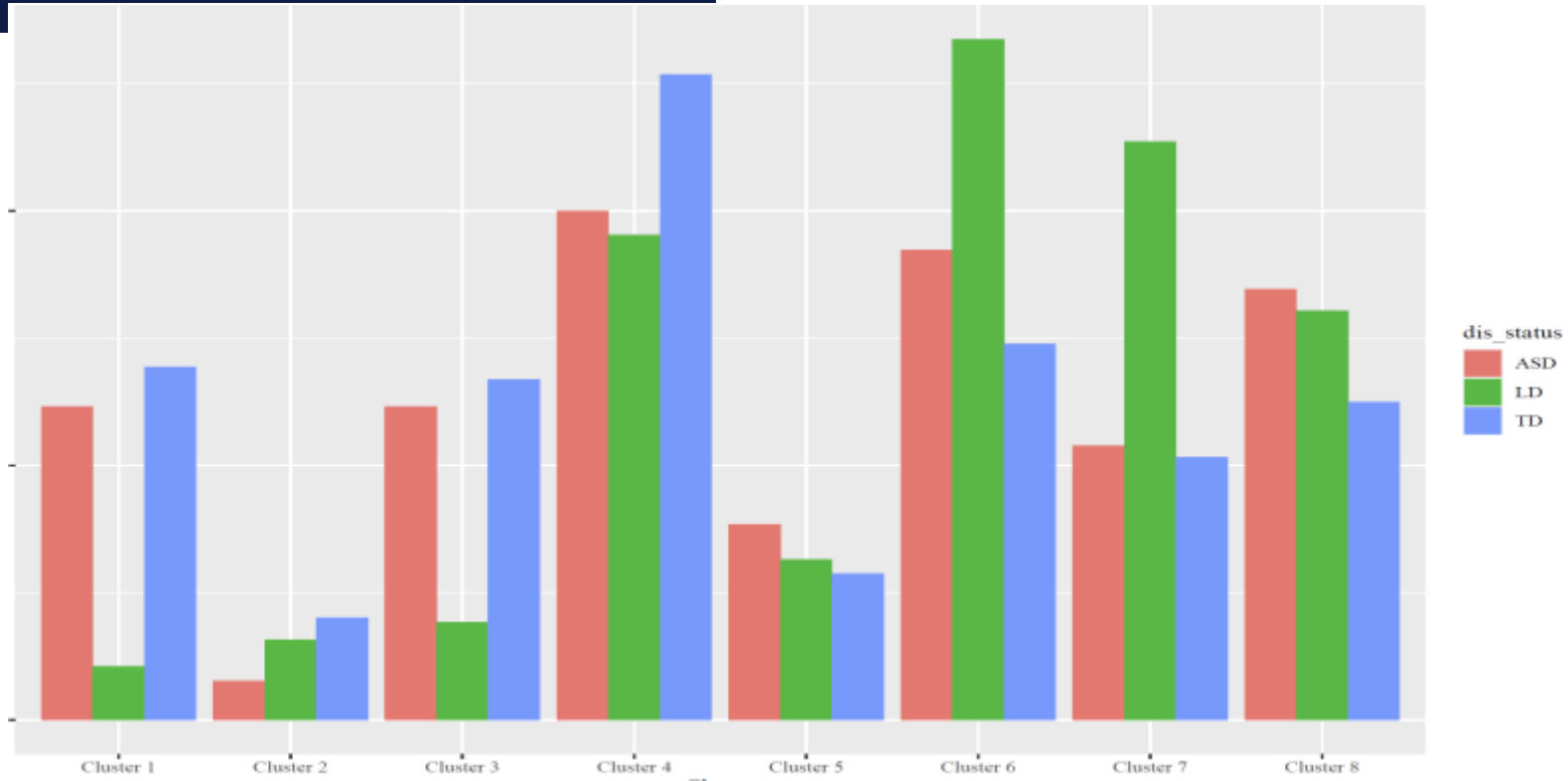
Group difference in cluster distribution: VH263651



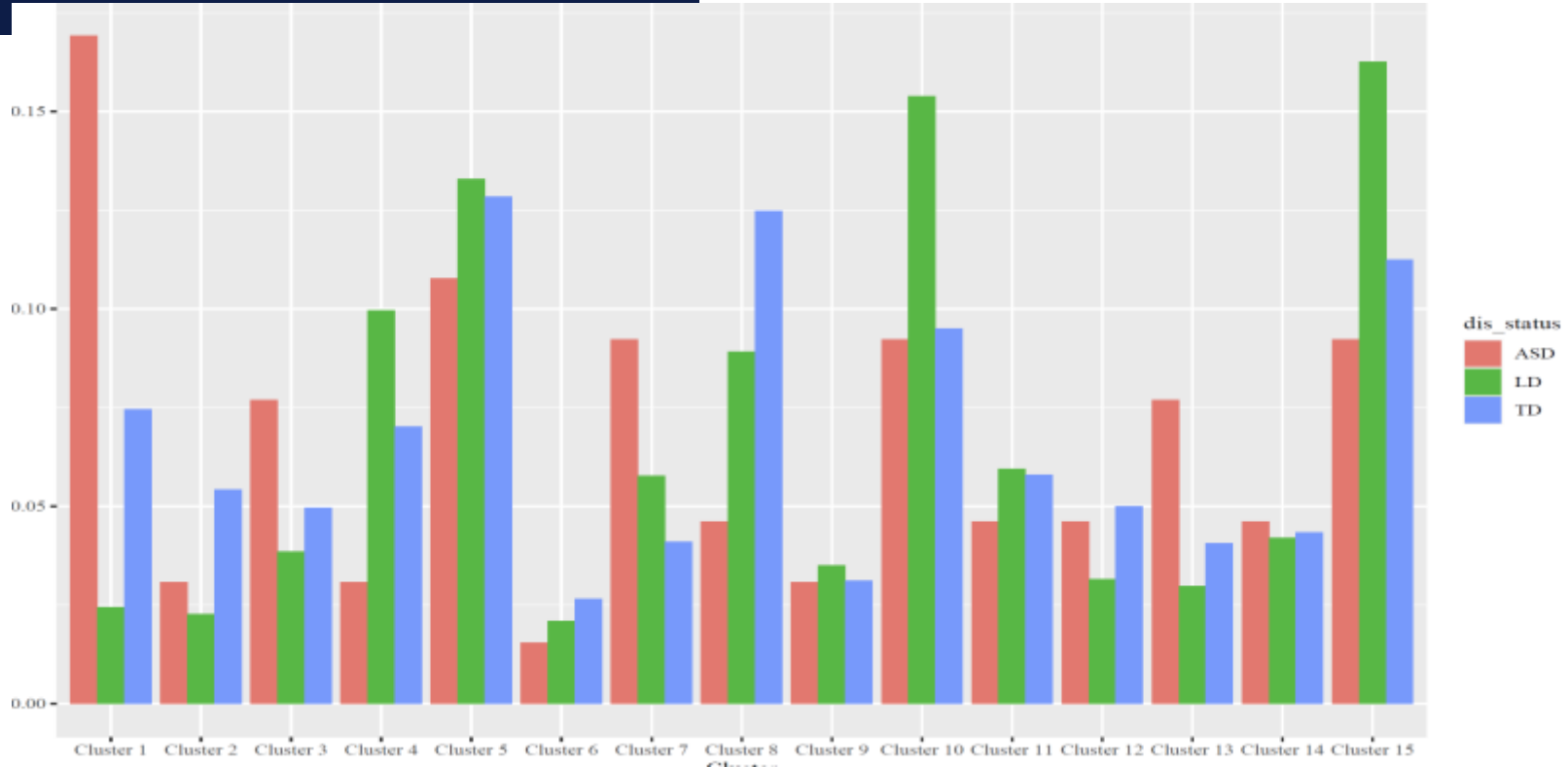
Group difference in cluster distribution: VH287980



Group difference in cluster distribution: VH302900



Group difference in cluster distribution: VH271388



Summary of Observations: TD vs LD

- For both single-part, single-select items, individuals with LD were:
  - Less likely to be firmly correct
  - More likely to be firmly incorrect
  - More likely to alternate different incorrect options

  Prior research indicated that LD students have more difficulties with fractions (Siegler, 2010) and mental rotation (Klonan et al., 2019).
- For multiple-part and/or multiple-select items, individuals with LD:
  - Demonstrated incorrect response or hesitation on different part(s) of the question than the "common misconception" by TD group
- Across all items, individuals with LD were:
  - More likely to use "Clear Answer" together with response alternation

Summary of Observations: TD vs ASD

- For all items, individuals with ASD had comparable (if not higher) proportion of individuals who responded correctly within one attempt compared to their TD peers:
  - On Item VH271388, ASD group had highest proportion (17%) of correct responders (vs. 7.5% in TD learners)

  Consistent with prior research that high-functioning autistic learners are strong in abstract spatial reasoning (Caron et al., 2004)
- However, ASD also demonstrated:
  - Higher proportions of individuals who hesitate or make mistakes on the "uncommon misconceptions"
  - Higher proportion who omitted item VH287980

Summary of Findings

- Clustering of revision and review sequences may preserve contextual information that cannot be achieved through analysis of single actions:
  - E.g., clear answer may either indicate platform functionality exploration or hesitation/starting over on question
- Learners within the same disability group can also demonstrate behavioral and performance heterogeneity. Clustering helps identify proportions of each type.
- The most common misconceptions in the TD group are often not the same as that in groups with disability.
- Use of Clear Answer in combination with response alternation may indicate repeatedly starting over on a question, suggesting lack of continuous and uninterrupted engagement in problem-solving.

THANKS!

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R324P210005 to SRI International and the University of Illinois at Urbana-Champaign. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Data source: U.S. Department of Education, National Center for Education Statistics, “Response Process Data from the NAEP 2017 Grade 8 Math Assessment.”

You can find me at szhan105@illinois.edu