Non-construct Item Features and Response Processes in Computer Science Assessments: Evidence From Think-Alouds and Sequence Analysis

AERA 2021 Annual Meeting

Matt J. Davidson
Min Li

Automatically Synthesizing Valid, Personalized, Formative Assessments of CS1 Concepts (NSF-1735123)

UNIVERSITY *of* WASHINGTON | COLLEGE OF EDUCATION

Land Acknowledgement

Land Acknowledgement

The University of Washington, and all of our lives and institutions, exists on Indigenous land. I acknowledge that I work and live on the ancestral homelands of the Coast Salish Peoples who have lived here from time immemorial. This land acknowledgement is one small act in the ongoing process of working to be in good relationship with the land and the people of the land and, ultimately, toward decolonization.

Adapted from the Banks Center for Educational Justice, https://www.education.uw.edu/cej/

Response processes

Response processes

- Underexplored source of validity evidence

Response processes

- Underexplored source of validity evidence
- Research on CS assessments focused on content validity

Response processes

- Underexplored source of validity evidence
- Research on CS assessments focused on content validity

***Research Question*: How does varying non-construct item features influence response processes?**

Item features framework

Item features framework

Dimension
Cognitive Demands
Openness
Evidence of thinking

Item features framework

Dimension	*Category*			
Cognitive Demands	Reading Syntax	Reading Templates	Writing Syntax	~~Writing Templates~~
Openness				
Evidence of thinking				

Note: Cells with a line through them were not tested as part of this study

Item features framework

Dimension	*Category*			
Cognitive Demands	Reading Syntax	Reading Templates	Writing Syntax	~~Writing Templates~~
Openness	Closed	~~Semi-open~~		Open
Evidence of thinking				

Note: Cells with a line through them were not tested as part of this study

Item features framework

Dimension	*Category*			
Cognitive Demands	Reading Syntax	Reading Templates	Writing Syntax	~~Writing Templates~~
Openness	Closed	~~Semi-open~~		Open
Evidence of thinking	~~None~~	Inferred		Explicit

Note: Cells with a line through them were not tested as part of this study

Item features framework

Dimension	*Category*			
Cognitive Demands	Reading Syntax	Reading Templates	$\boxed{\text{Writing Syntax}}$	~~Writing Templates~~
Openness	Closed	~~Semi-open~~		Open
Evidence of thinking	~~None~~	Inferred		Explicit

Note: Cells with a line through them were not tested as part of this study

Item features framework

Dimension	*Category*			
Cognitive Demands	Reading Syntax	Reading Templates	Writing Syntax	~~Writing Templates~~
Openness	Closed	~~Semi-open~~		Open
Evidence of thinking	~~None~~	Inferred		Explicit

Note: Cells with a line through them were not tested as part of this study

Thinkalouds → Qualitative Codes

Thinkalouds → Qualitative Codes

Process Category
Monitoring
Problem Solving
Reading Semantics
Reading Templates
Writing Semantics
Writing Syntax

Thinkalouds → Qualitative Codes

Process Category	*Detailed Sub-codes (e.g.)*
Monitoring	applying strategy; noticing confusion
Problem Solving	read task requirements; eliminate answers
Reading Semantics	multistructural reading; tracing values
Reading Templates	relational reading; recognizing template
Writing Semantics	global planning; revising semantics
Writing Syntax	typing syntax; revising syntax

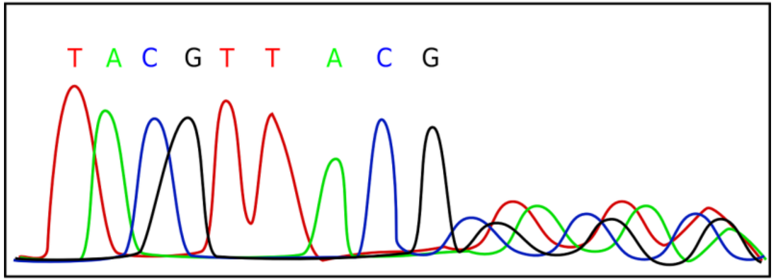
Thinkalouds → Qualitative Codes

Process Category	*Detailed Sub-codes (e.g.)*
Monitoring	applying strategy; noticing confusion
Problem Solving	read task requirements; eliminate answers
Reading Semantics	multistructural reading; tracing values
Reading Templates	relational reading; recognizing template
Writing Semantics	global planning; revising semantics
Writing Syntax	typing syntax; revising syntax

Data Analysis

Data Analysis

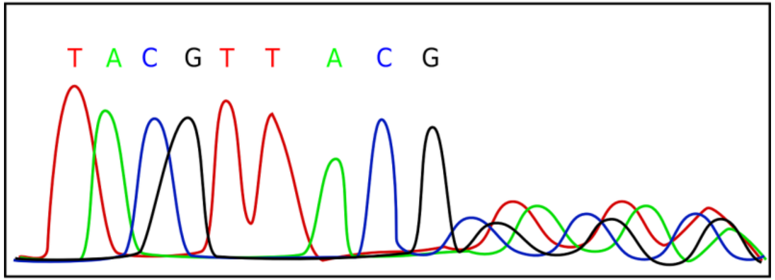
- Proportional frequency



Data Analysis

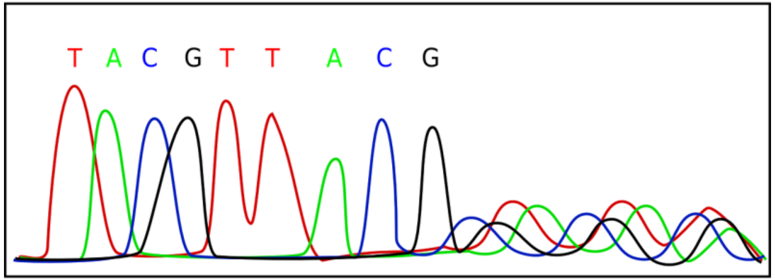
> Proportional frequency

> Sequence Dissimilarity



Data Analysis

- Proportional frequency
- Sequence Dissimilarity
- Sequential Pattern Mining



Proportional frequency

test for the frequency of each process category

Proportional frequency

Test for the frequency of each process category

Broad differences in sequences, ignoring order

Proportional frequency

Broad differences in sequences, ignoring order

> ***yrogetac ssecorp hcae fo ycne***
> ***test for the frequency of each process category***

	Monitoring	Problem Solving	Reading Semantics
Open	.67	.75	.42
Closed	.23	.65	.34

Proportional frequency

Broad differences in sequences, ignoring order

- > *yrogetac ssecorp hcae fo ycne*
  test for the frequency of each process category
- > ***p*-value correction within each item feature**

	Monitoring	Problem Solving	Reading Semantics
Open	.67	.75	.42
Closed	.23	.65	.34

Sequence Dissimilarity

$= +$

Sequence Dissimilarity

= +

Broad differences in sequences, taking order into account

Sequence Dissimilarity

Broad differences in sequences, taking order into account

= +

A: Monitoring – RS – RT – Problem Solving

B: Monitoring – RS – WS – Monitoring

C: Problem Solving – RS – WS - RT

Sequence Dissimilarity

Broad differences in sequences, taking order into account

A: Monitoring – RS – RT – Problem Solving

B: Monitoring – RS – WS – Monitoring

C: Problem Solving – RS – WS - RT

> $\boldsymbol{LCS}\ \mathbf{A, B}$

> $\boldsymbol{BBLLCCSS}\ \ \boldsymbol{LCS}\ \mathbf{A, B\ AA,}\ \ \boldsymbol{LCS}\ \mathbf{A, B}\ \ *$
$\mathbf{2} - \boldsymbol{L}\ \mathbf{B}\ \boldsymbol{LL}\ \boldsymbol{L}\ \mathbf{B\ BB}\ \boldsymbol{L}\ \mathbf{B}\ \ \ \boldsymbol{L}\ \mathbf{A}\ \boldsymbol{LL}\ \boldsymbol{L}\ \mathbf{A\ AA}\ \boldsymbol{L}\ \mathbf{A} + =$
$\mathbf{BB}\ \boldsymbol{distance}\ \mathbf{A, B}\ \ \ \boldsymbol{distance}\ \mathbf{A,}\ \boldsymbol{stance}\ \mathbf{A, B\ A, B}$
$= \boldsymbol{L}_{\mathbf{A}} + \boldsymbol{L}_{\mathbf{B}} - \mathbf{2} * \boldsymbol{LCS}_{\mathbf{A,B}}$

Sequence Dissimilarity

Broad differences in sequences, taking order into account

A: Monitoring – RS – RT – Problem Solving

B: Monitoring – RS – WS – Monitoring

C: Problem Solving – RS – WS - RT

- $LCS\ \mathrm{A, B}$
- $BBLLCCSS\ \ LCS\ \mathrm{A, B\ AA},\ \ LCS\ \mathrm{A, B}\ \ *$
  $2 - \ L\ \mathrm{B}\ LL\ L\ \mathrm{B\ BB}\ L\ \mathrm{B}\quad L\ \mathrm{A}\ LL\ L\ \mathrm{A\ AA}\ L\ \mathrm{A}+=$
  $\mathrm{BB}\ distance\ \mathrm{A, B}\quad distance\ \mathrm{A},stance\ \mathrm{A, B\ A, B}$
  $= L_{\mathrm{A}} + L_{\mathrm{B}} - 2 * LCS_{\mathrm{A,B}}$

	A	**B**	**C**
A	-	2	4
B	2	-	2
C	4	2	-

Sequence Dissimilarity

Broad differences in sequences, taking order into account

A: Monitoring – RS – RT – Problem Solving

B: Monitoring – RS – WS – Monitoring

C: Problem Solving – RS – WS - RT

- LCS A, B
- $BBLLCCSS$ LCS A, B AA, LCS A, B $*$
  2 – L B LL L B BB L B L A LL L A AA L A+ =
  BB $distance$ A, B $distance$ A, $stance$ A, B A, B
  $= L_{A} + L_{B} - 2 * LCS_{A,B}$
- $SS =$ sum of pairwise distances

	A	B	C
A	-	2	4
B	2	-	2
C	4	2	-

Sequence Dissimilarity

Broad differences in sequences, taking order into account

A: Monitoring – RS – RT – Problem Solving

B: Monitoring – RS – WS – Monitoring

C: Problem Solving – RS – WS - RT

- $SS\ within\ SSSS\ SS\ within\ wwiitthhiinn\ SS\ within$
- $SS\ between\ \ /\ \ LCS\ \mathrm{A,B}$
- $BBLLCCSS\ \ LCS\ \mathrm{A,B\ AA},\ \ LCS\ \mathrm{A,B}\ \ *$
  $2 - \ L\ \mathrm{B}\ LL\ L\ \mathrm{B\ BB}\ L\ \mathrm{B}\ \ \ L\ \mathrm{A}\ LL\ L\ \mathrm{A\ AA}\ L\ \mathrm{A} + =$
  $\mathrm{BB}\ distance\ \mathrm{A,B}\ \ distance\ \mathrm{A},\ stance\ \mathrm{A,B\ A,B}$
  $= L_{\mathrm{A}} + L_{\mathrm{B}} - 2 * LCS_{\mathrm{A,B}}$
- $SS =$ sum of pairwise distances

	A	B	C
A	-	2	4
B	2	-	2
C	4	2	-

- $F = \ SS\ between\ en\ between\ /\ SS_{within}$

Sequential Pattern Mining

Sequential Pattern Mining

Subsequences that are associated with an item feature

Sequential Pattern Mining

Subsequences that are associated with an item feature

> **Discover frequent subsequences**

Sequential Pattern Mining

Subsequences that are associated with an item feature

- **Discover frequent subsequences**
  - **50% support**

Sequential Pattern Mining

Subsequences that are associated with an item feature

- **Discover frequent subsequences**
  - **50% support**
  - **Max length of 3**

Sequential Pattern Mining

Subsequences that are associated with an item feature

- **Discovertest for each subsequence**
  - **50% support**
  - **Max length of 3**

- **χ 2 2 2 2 test for each subsequence**

Sequential Pattern Mining

Subsequences that are associated with an item feature

- **Discovertest for each subsequence**
  - **50% support**
  - **Max length of 3**

Subseq 1	Present	Absent
Open	58	20
Closed	13	39

- **$\chi\ 2 \quad 2\ 2\ 2$ test for each subsequence**

Sequential Pattern Mining

Subsequences that are associated with an item feature

- Discovertest for each subsequence
  - 50% support
  - Max length of 3
- χ 2   2 2 2 test for each subsequence

Subseq 1	Present	Absent
Open	58	20
Closed	13	39

Subseq 2	Present	Absent
Open	39	39
Closed	30	22

Results

Results

> **Openness**

Results

> **Openness**

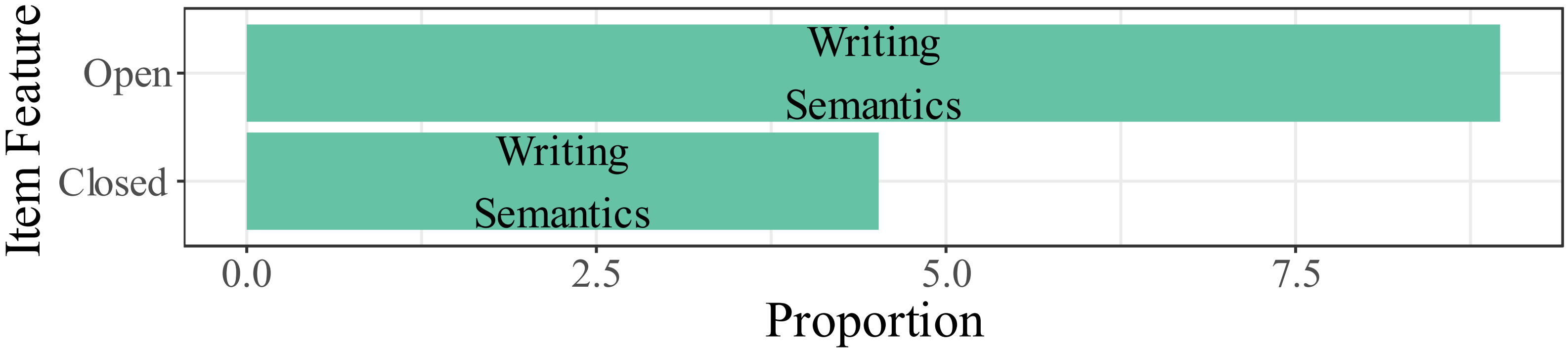
> **Evidence of thinking**

Results

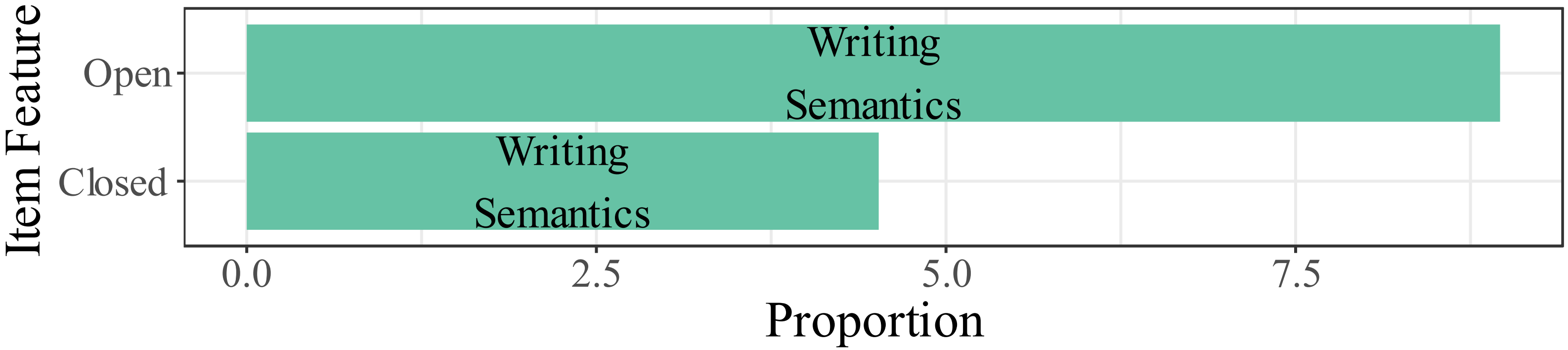
- Openness
- Evidence of thinking
- Cognitive demands

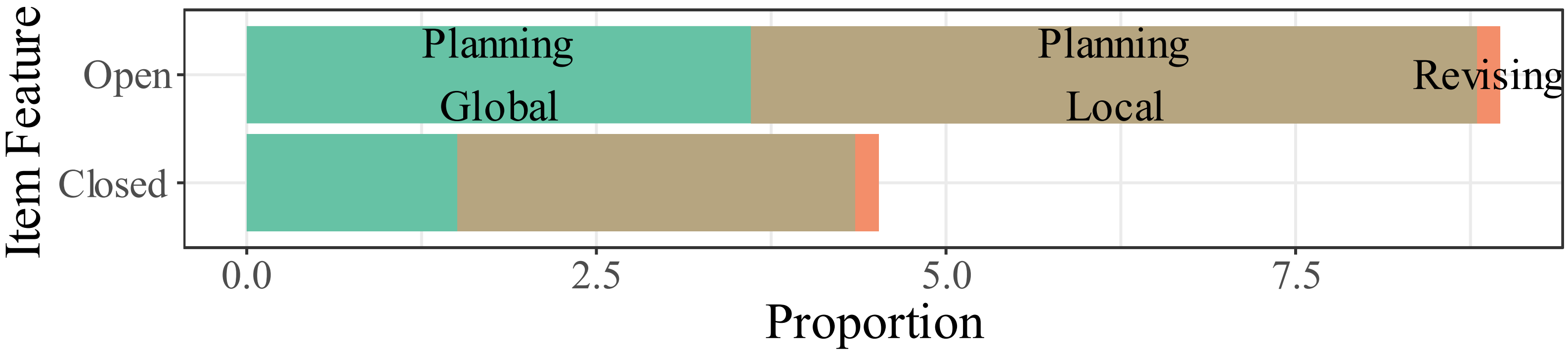
Results: Proportional Frequency

Results: Proportional Frequency



Results: Proportional Frequency





Results: Sequence Dissimilarity

Results: Sequence Dissimilarity

Summary of Pseudo-ANOVA results for sequence dissimilarity measure

	Sum of Squares	*df*	Mean Square	*pseudo- F*	*p*
Openness (open, closed)					
Explained	0.56	1	0.55	2.42	0.004
Residual	29.74	129	0.23		
Evidence (inferred, explicit)					
Explained	0.31	1	0.31	1.32	0.141
Residual	29.99	129	0.23		
Cognitive Demands (RS, RT, WS)					
Explained	5.17	2	2.59	13.18	0.001
Residual	25.13	128	0.20		

Results: Sequence Dissimilarity

Summary of Pseudo-ANOVA results for sequence dissimilarity measure

	Sum of Squares	*df*	Mean Square	*pseudo- F*	*p*
Openness (open, closed)					
Explained	0.56	1	0.55	2.42	0.004
Residual	29.74	129	0.23		
Evidence (inferred, explicit)					
Explained	0.31	1	0.31	1.32	0.141
Residual	29.99	129	0.23		
Cognitive Demands (RS, RT, WS)					
Explained	5.17	2	2.59	13.18	0.001
Residual	25.13	128	0.20		

Results: Sequential Pattern Mining

Results: Sequential Pattern Mining

Subsequences that discriminated best between open and closed writing items

Subsequence	*p*	Open	Closed
(typing syntax)-(planning: local)-(reread program specification)	0.23	0.75	0.19
(planning: local)-(typing syntax)-(reread program specification)	0.44	0.71	0.19
(planning: local)-(reread program specification)	0.61	0.75	0.25
(read program specification)-(planning: local)-(reread program specification)	0.61	0.75	0.25
(read task requirements)-(planning: local)-(reread program specification)	0.61	0.75	0.25
(typing syntax)-(planning: local)-(planning: local)	0.61	0.75	0.25
(planning: local)-(reread program specification)-(%)	0.87	0.71	0.25
(planning: local)-(reread program specification)-(typing syntax)	0.87	0.71	0.25
(typing syntax)-(reread program specification)-(planning: local)	0.95	0.75	0.31
(planning: local)-(typing syntax)-(typing syntax)	1.00	0.86	0.50

Results: Sequential Pattern Mining

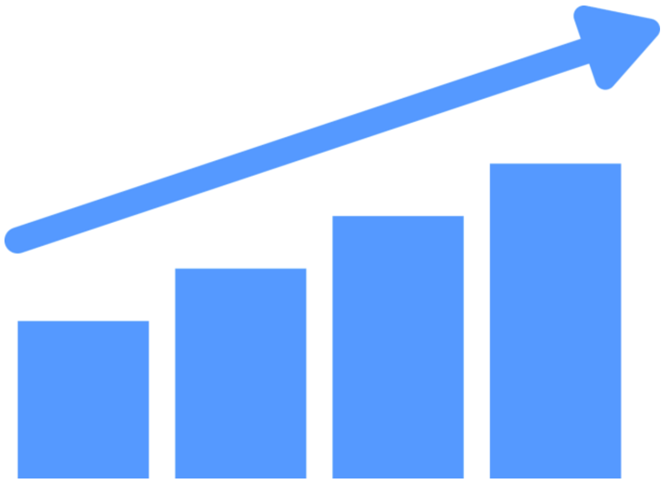
Subsequences that discriminated best between open and closed writing items

Subsequence	*p*	Open	Closed
(typing syntax)-(planning: local)-(reread program specification)	0.23	0.75	0.19
(planning: local)-(typing syntax)-(reread program specification)	0.44	0.71	0.19
(planning: local)-(reread program specification)	0.61	0.75	0.25
(read program specification)-(planning: local)-(reread program specification)	0.61	0.75	0.25
(read task requirements)-(planning: local)-(reread program specification)	0.61	0.75	0.25
(typing syntax)-(planning: local)-(planning: local)	0.61	0.75	0.25
(planning: local)-(reread program specification)-(%)	0.87	0.71	0.25
(planning: local)-(reread program specification)-(typing syntax)	0.87	0.71	0.25
(typing syntax)-(reread program specification)-(planning: local)	0.95	0.75	0.31
(planning: local)-(typing syntax)-(typing syntax)	1.00	0.86	0.50

Conclusions

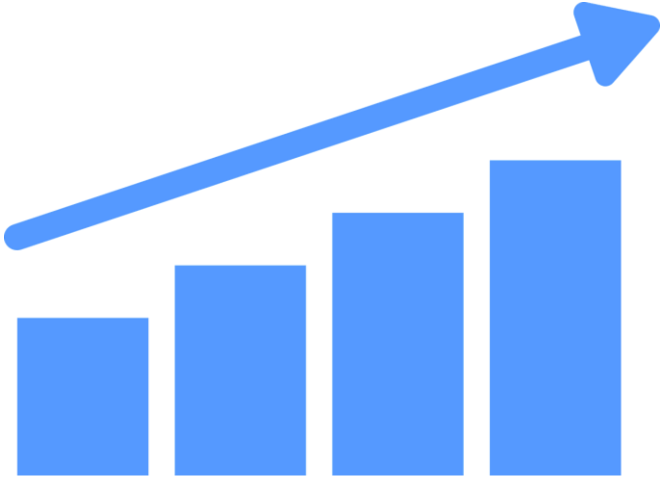
Conclusions

> Open items → more planning



Conclusions

- Open items → more planning
- Implications for the construct

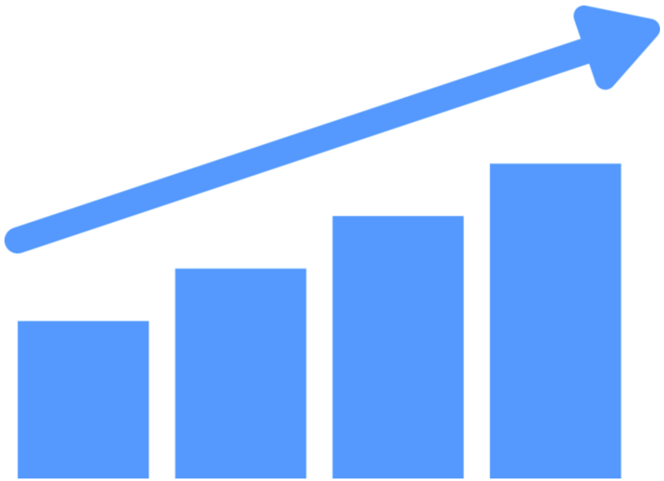


Conclusions

> Open items → more planning

> Implications for the construct

> Response process validity



Non-construct Item Features and Response Processes in Computer Science Assessments: Evidence From Think-Alouds and Sequence Analysis

Matt Davidson | mattjd@uw.edu | @dontmattme | mattjohndavidson.github.io

> **Varying an item's openness influenced students planning and writing processes on code writing items.**

> **Sequence analysis techniques can be used to analyze response process data.**

More details about the methods in blog post at:

bit.ly/AERA21-sequences

Details on qualitative codes and results for all item features in the iPresentation:

bit.ly/AERA21-poster