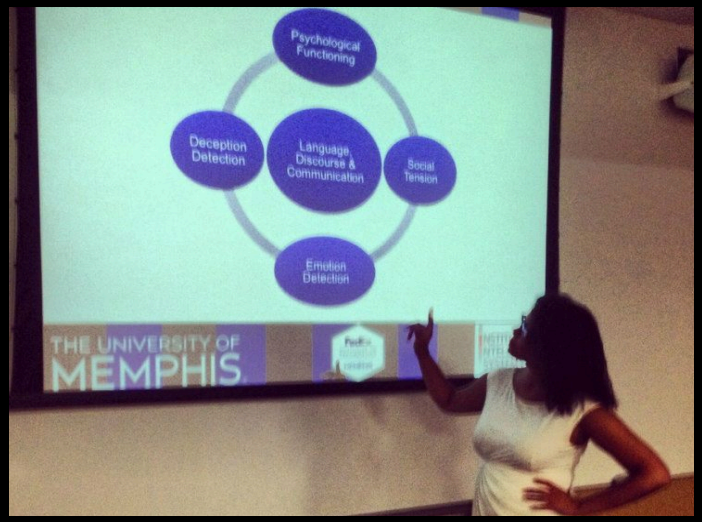
Group Communication Analysis

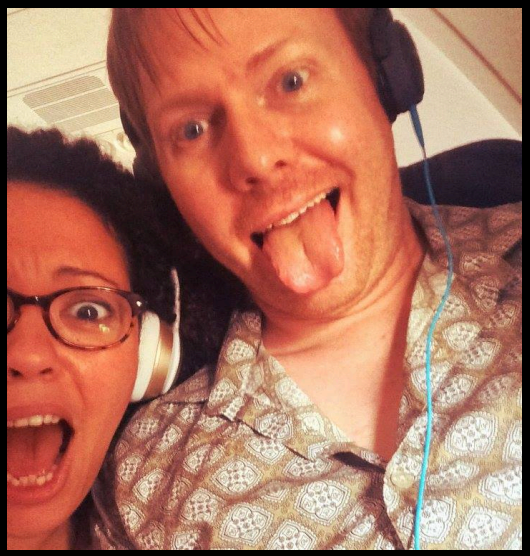
A computational linguistics approach for detecting sociocognitive roles in multi-party interactions

Dr. Nia Dowell

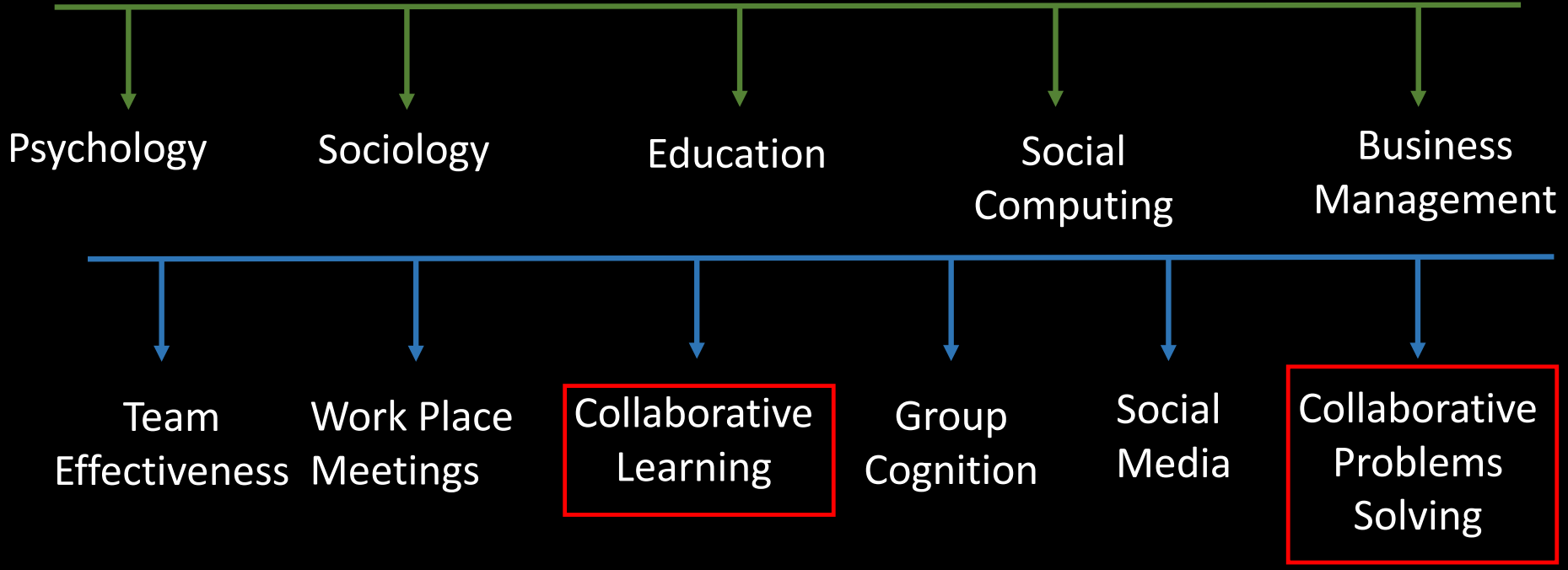
Postdoctoral Research Fellow
School of Information
Digital Innovation Greenhouse
University of Michigan
ndowell@umich.edu
http://niadowell.com


Psychological Functioning
Deception Detection
Language, Discourse & Communication
Social Tension
Emotion Detection
THE UNIVERSITY OF MEMPHIS

Gala



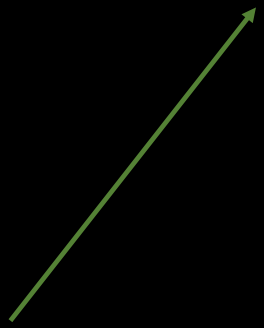
Social Roles Research



Prominent Perspectives on Roles

Assigned Roles

A position to which a person is assigned and then performs the behavior associated with that position

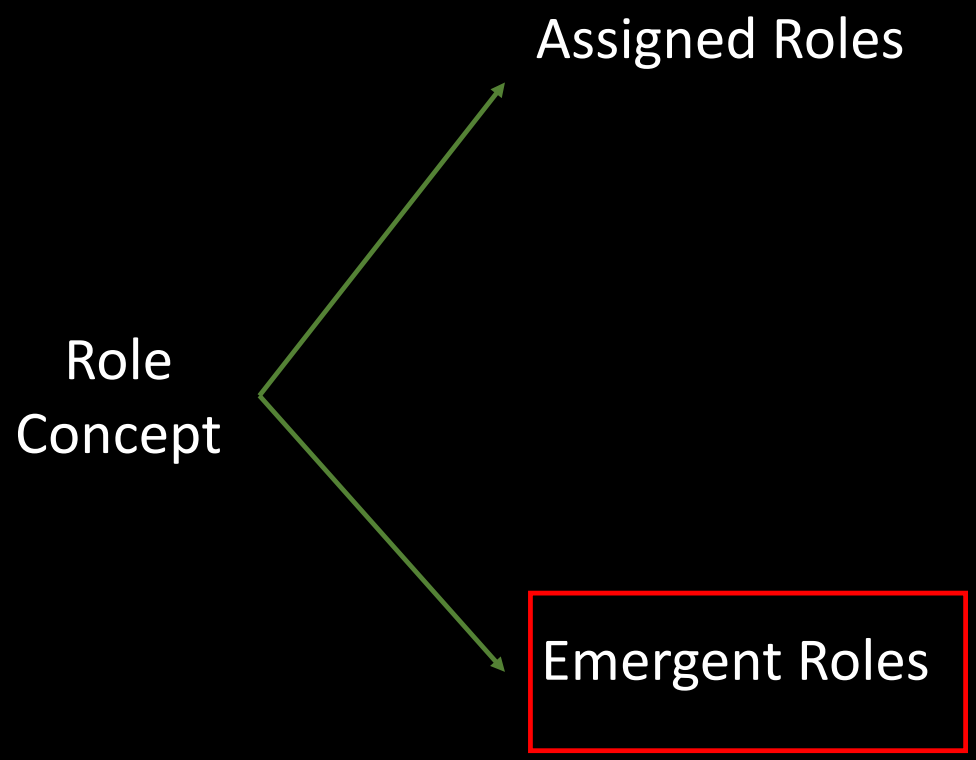


Role Concept

Concerns

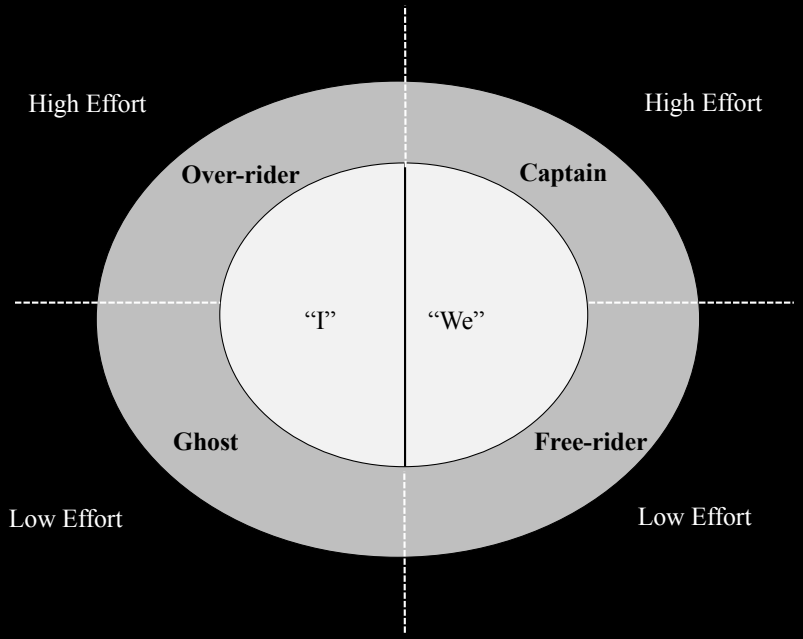
- Dysfunctional group roles
- What is actually captured in role assignment research?
- A single role inhibits role and group flexibility, and the potential advantages of this
- Disregards the dynamic and interactive way in which roles are created, negotiated, and evolve among group members during social interaction

Prominent Perspectives on Roles

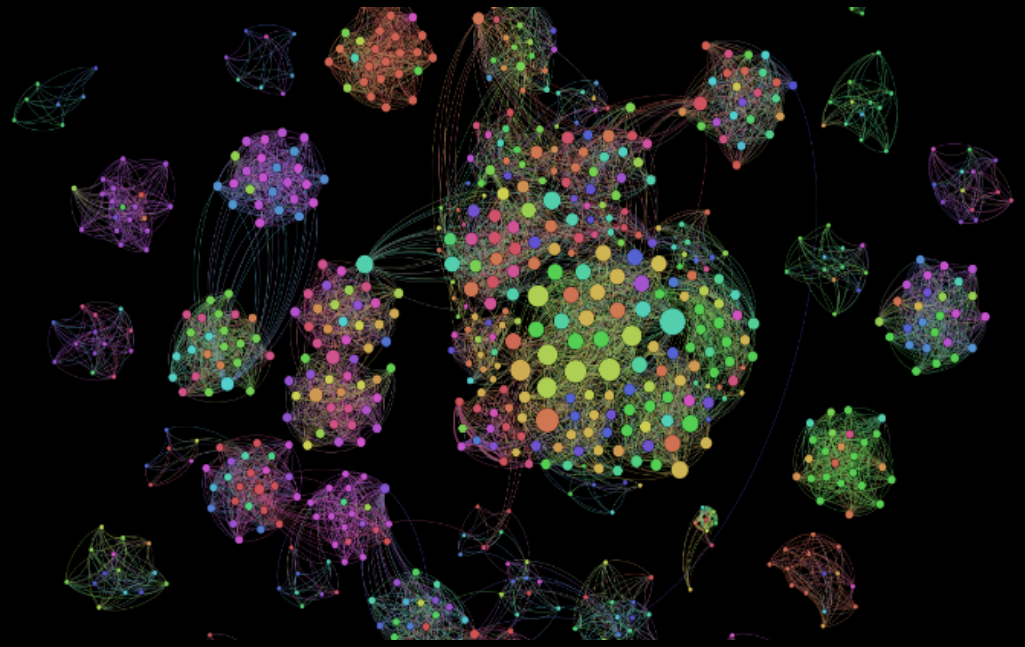


A position to which a person is assigned and then performs the behavior associated with that position

Develop naturally out of the interpersonal interaction without any prior instruction or assignment, and are characterized by their behavioral proximity (similarities and differences) to other interactional partners



Strijbos & De Laat (2010)



Marcos-Garcia et al., 2015

Can we automatically identify the roles students take on during collaborative interactions?

Language and Discourse

Online Multi-party Interaction

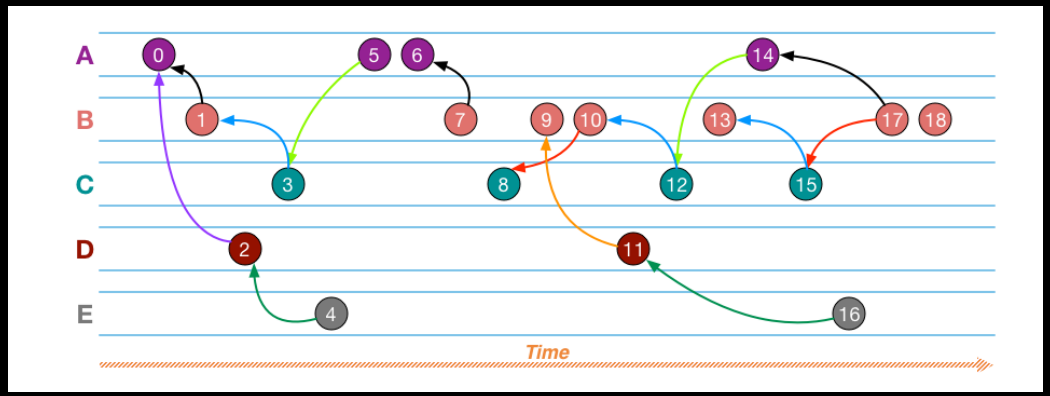
Speaker **Time** **Discourse**

D6

	A	B	C	D
	person_id	chat_room_i	chat_time	chat_text
2	33	953	11/9/11 13:29	hi
3	31	953	11/9/11 13:29	hi
4	31	953	11/9/11 13:29	hmm I don't see the discussion questions
5	20	953	11/9/11 13:29	So they cannot progress until they are at least in a room that starts.
6	33	953	11/9/11 13:29	john, do i need to remove the timer...in case they have problems?
7	31	953	11/9/11 13:29	do you see them?
8	20	953	11/9/11 13:30	oh. 1 sec.
9	33	953	11/9/11 13:30	i don't know what questions you are talking about
10	31	953	11/9/11 13:30	the discussion questions
11	31	953	11/9/11 13:30	the ones they are suppose to be chatting about
12	31	953	11/9/11 13:30	I can send them again
13	31	953	11/9/11 13:30	is it possible to add these above the chat box?
14	31	953	11/9/11 13:31	This is what the discussion questions are suppose to say
15	31	953	11/9/11 13:31	What are the characteristics that individuals must display to be consider
16	33	953	11/9/11 13:31	i'm guessing that john knows about these?
17	31	953	11/9/11 13:33	i'm not sure
18	20	953	11/9/11 13:33	refresh
19	33	953	11/9/11 13:33	ok
20	31	953	11/9/11 13:33	ah great
21	31	953	11/9/11 13:33	thanks
22	20	953	11/9/11 13:33	funny saw you blink out when you refreshed.
23	20	953	11/9/11 13:34	I put the text in the chat brief.
24	33	953	11/9/11 13:34	so, where do I put the instructions?
25	31	953	11/9/11 13:35	Would it be possible to add above these questions: Please use this time
26	20	953	11/9/11 13:35	in the brief for the pretest....
27	31	953	11/9/11 13:36	not sure if you are responding to jason or my comment.
28	31	953	11/9/11 13:36	I was talking about brief instructions above these chat discussion questi
29	20	953	11/9/11 13:37	jason
30	31	953	11/9/11 13:37	ok, makes sense
31	31	953	11/9/11 13:37	did jason drop out of the chat?
32	20	953	11/9/11 13:38	I think the instructions would fit best in the followup for the pretest. bef
33	31	953	11/9/11 13:38	you mean the ones I just posted?

How do we go from this semi-unstructured data to something meaningful, something that allows us to capture the important sociocognitive processes taking place within the interaction.

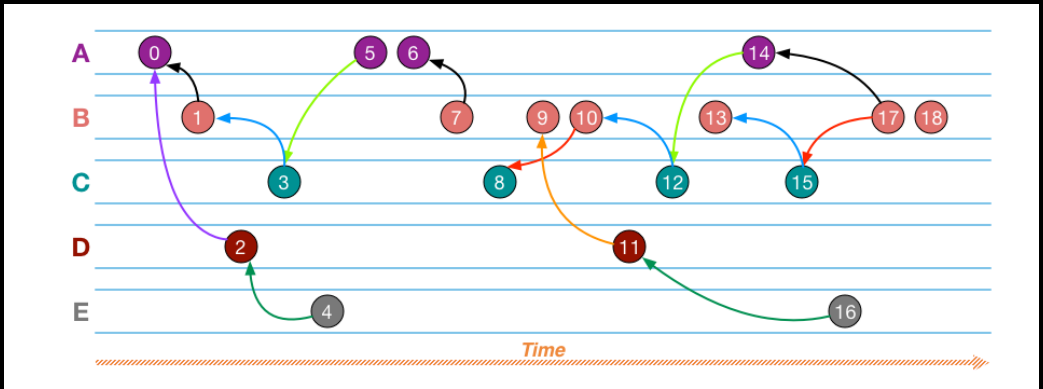
Infer semantic relationship among students' contributions

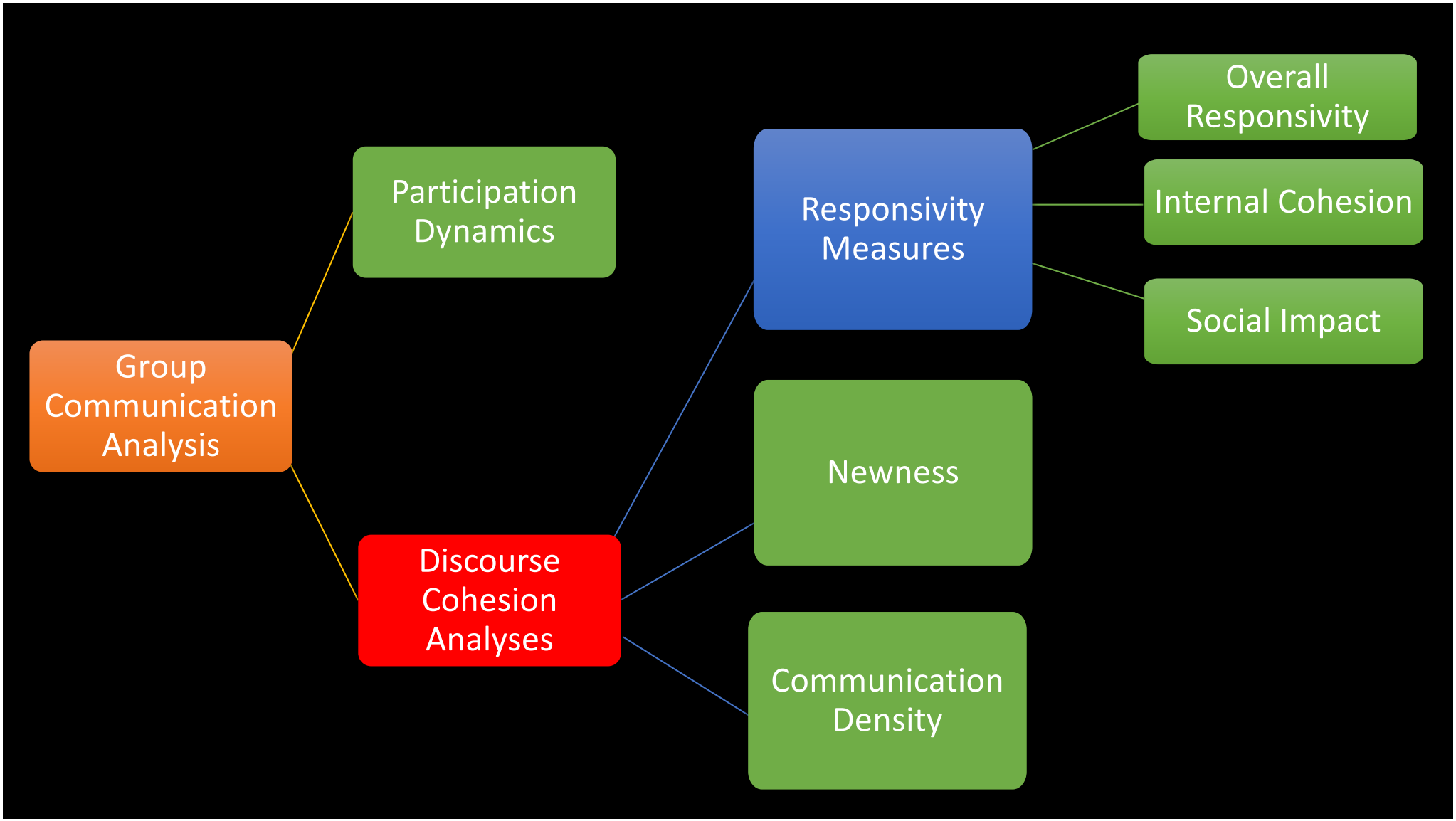


Discourse Cohesion

Latent Semantic Analysis

This similarity measure represents the semantic and conceptual meanings of individual words, utterances, texts, and larger stretches of discourse based on the statistical regularities between words in a large corpus of naturalistic text



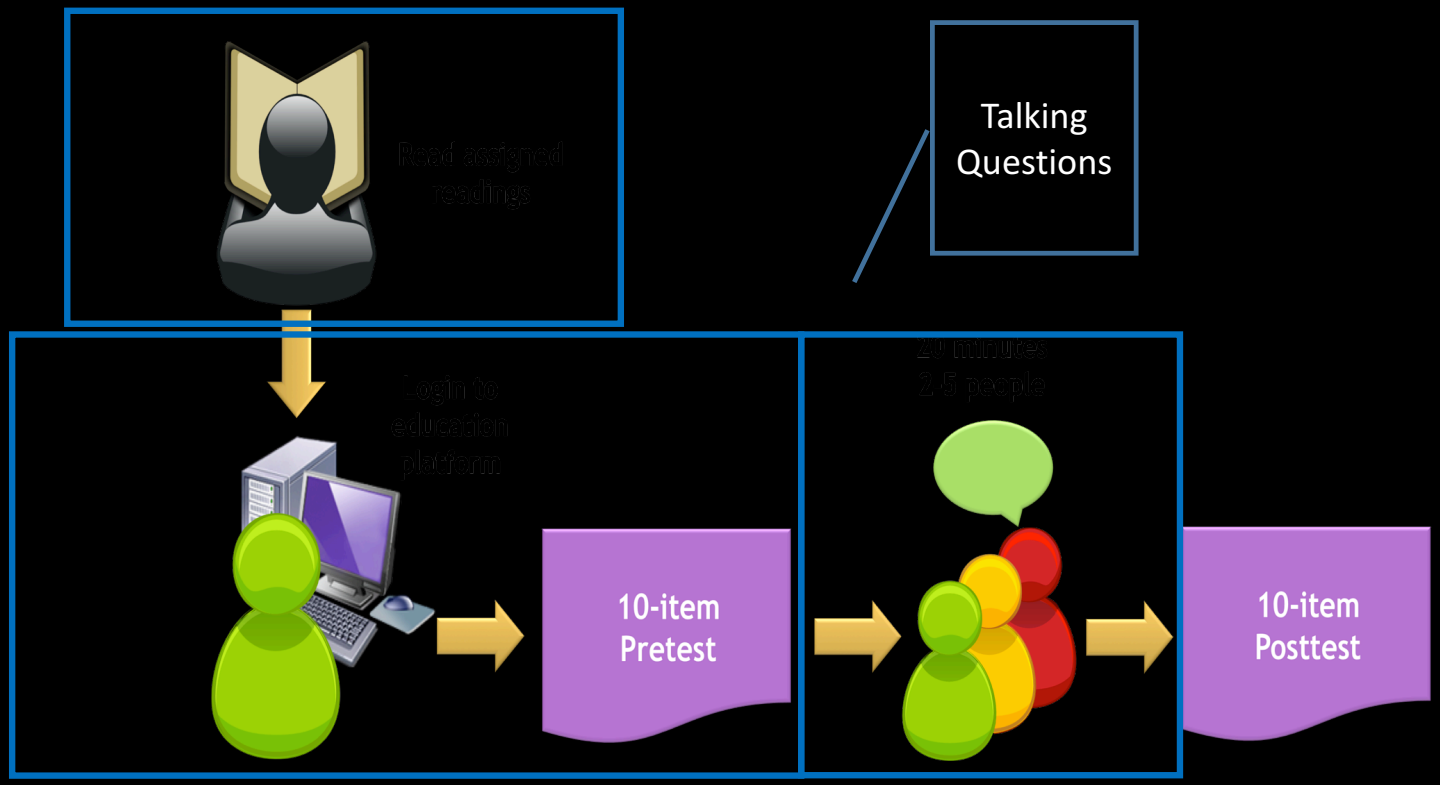

Group Communication Analysis
Participation Dynamics
Discourse Cohesion Analyses
Responsivity Measures
Newness
Communication Density
Overall Responsivity
Internal Cohesion
Social Impact

Jamie Pennebaker + Team

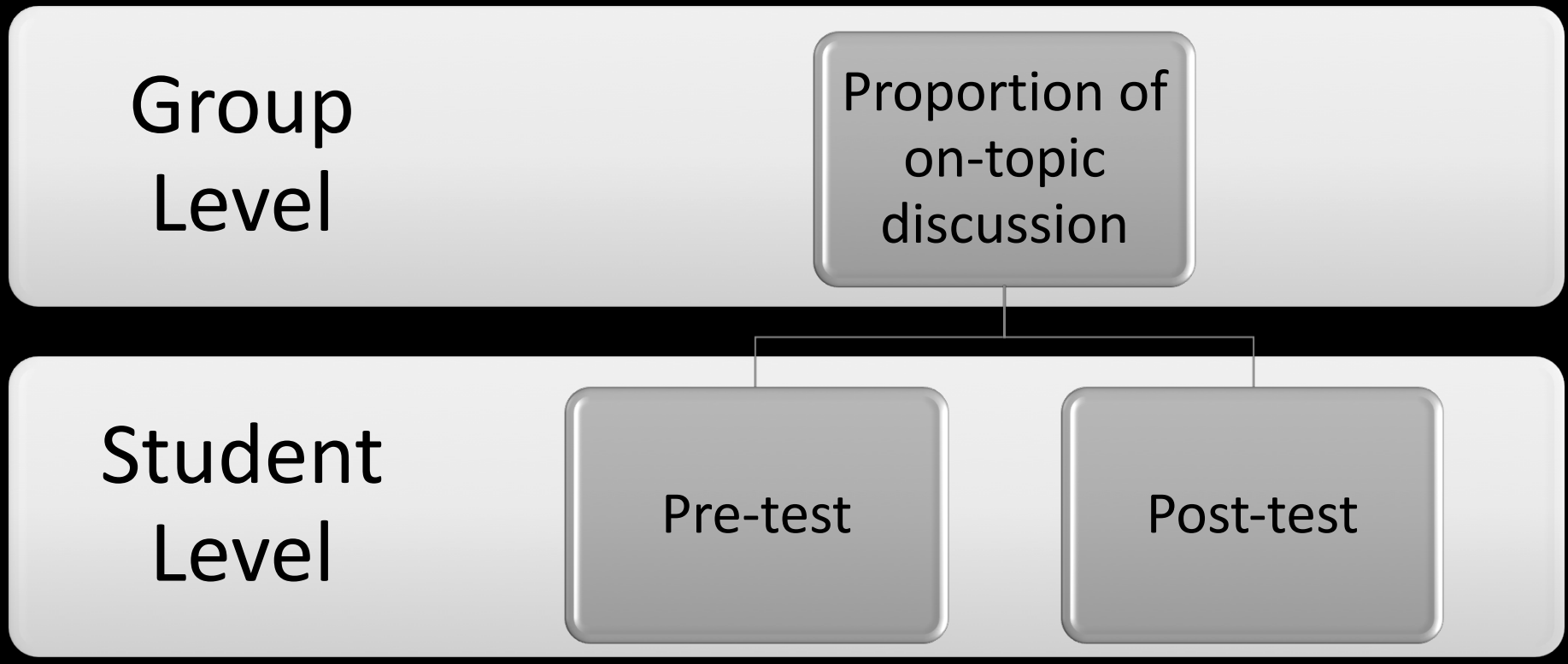
THE UNIVERSITY OF TEXAS AT AUSTIN

Participants: 840 undergraduates in an introductory-level psychology course
Groups: 184 randomly assigned groups

THE UNIVERSITY OF TEXAS AT AUSTIN



Measuring Performance



[% Posttest - % Pretest] / [1 - % Pretest]

Detecting Emergent Roles

Pre Clustering

Testing

Training

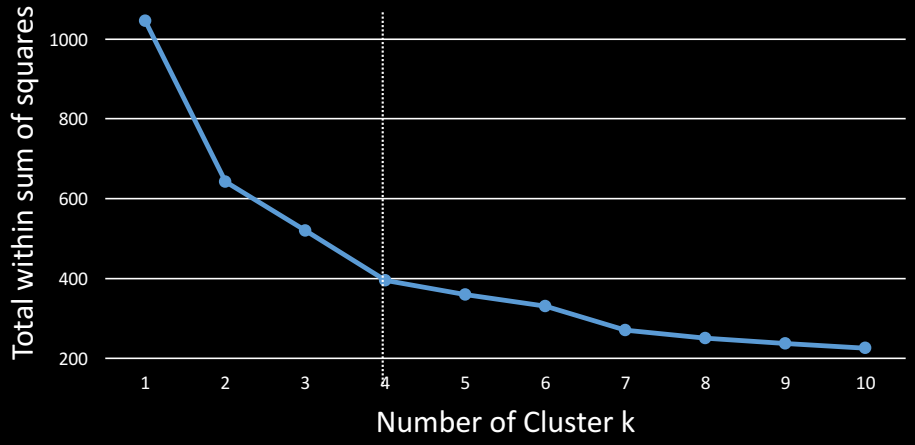
Multicollinearity

Cluster Tendency

Hopkins statistic = .15

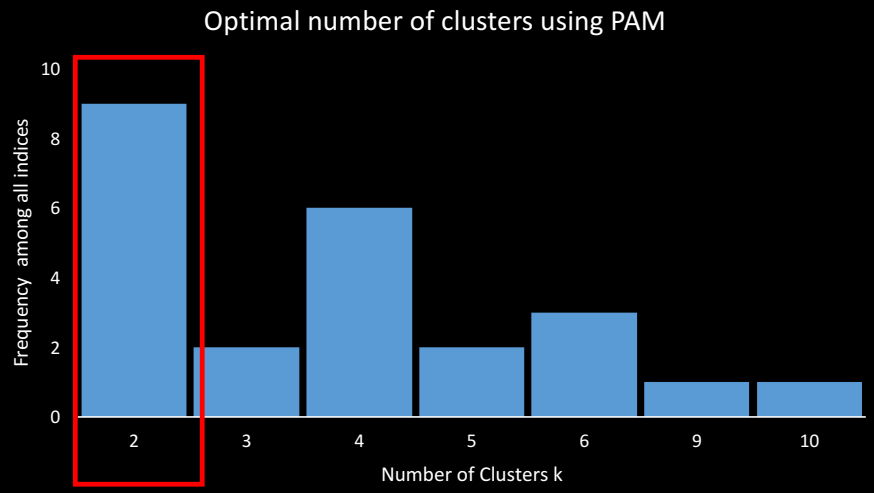
Optimal Number of Clusters

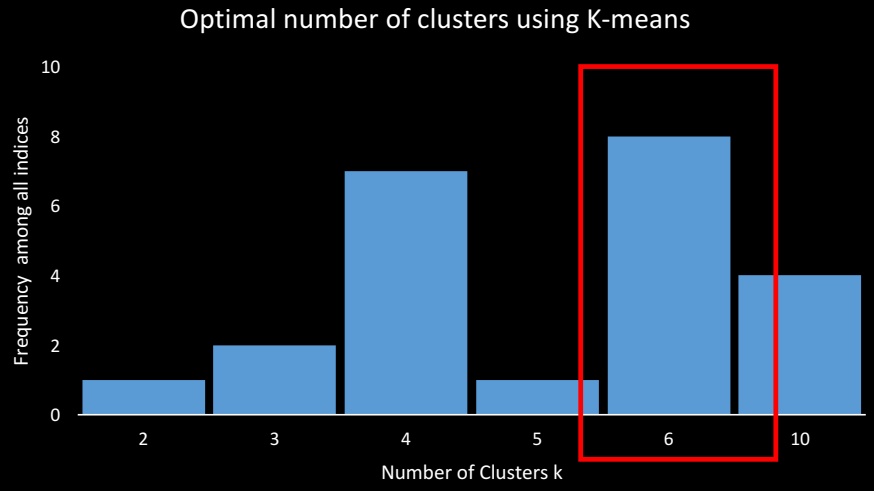
WSS



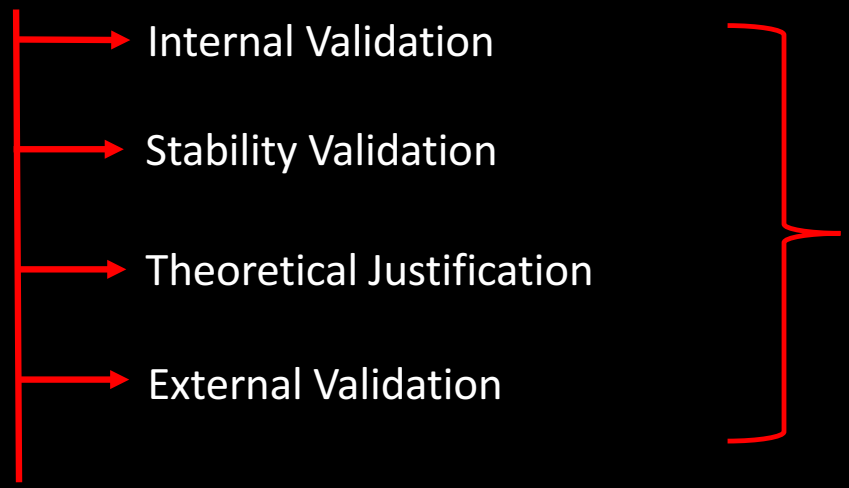
The disadvantage of elbow and similar methods is that, they measure a global clustering characteristic only

Majority rule



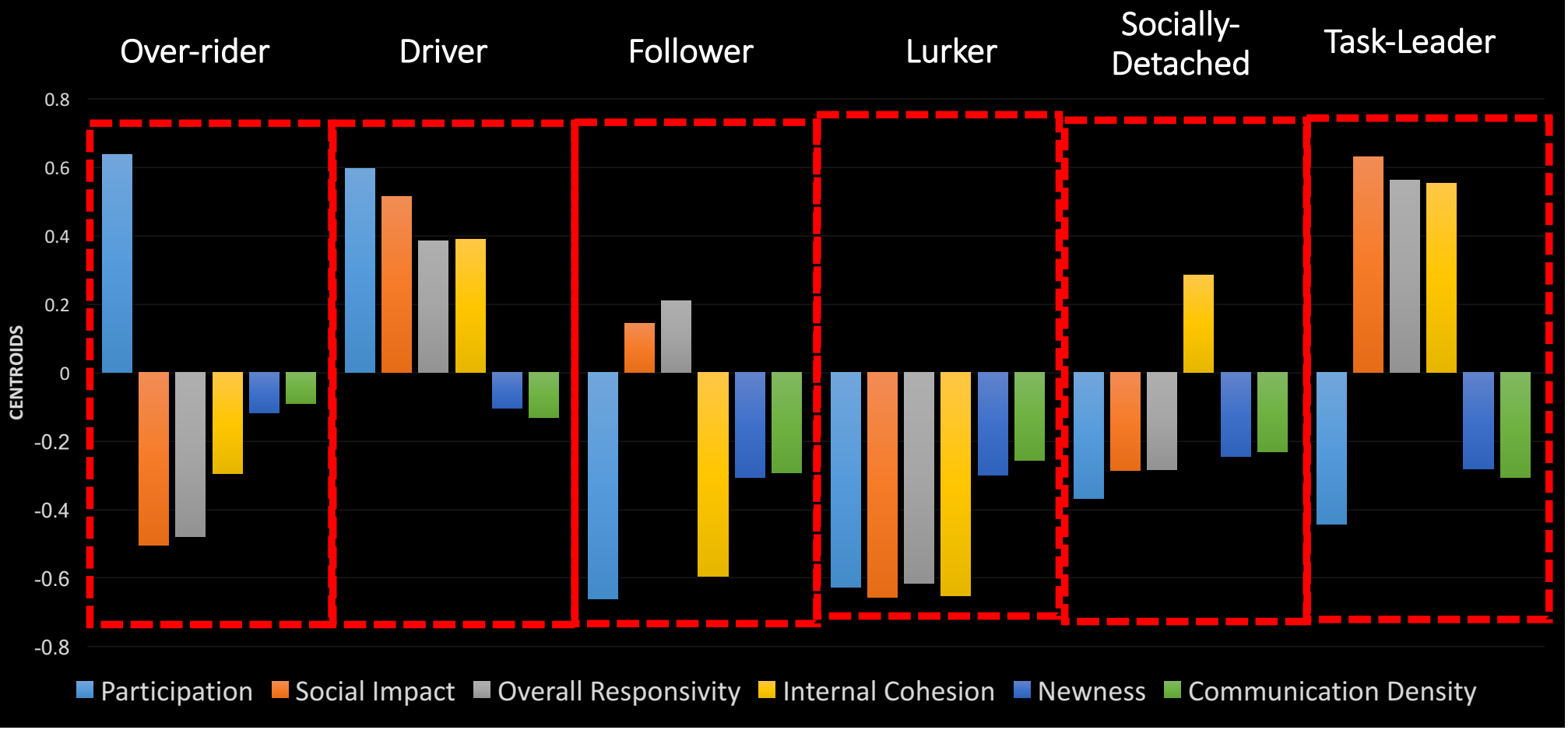


Cluster Evaluation and Validation



4 Cluster Model
&
6 Cluster Model

From Model to Meaning



Student Roles and Learning

Computer-Supported Collaborative Learning (2013) 8:1–12
DOI 10.1007/s11412-013-9169-0

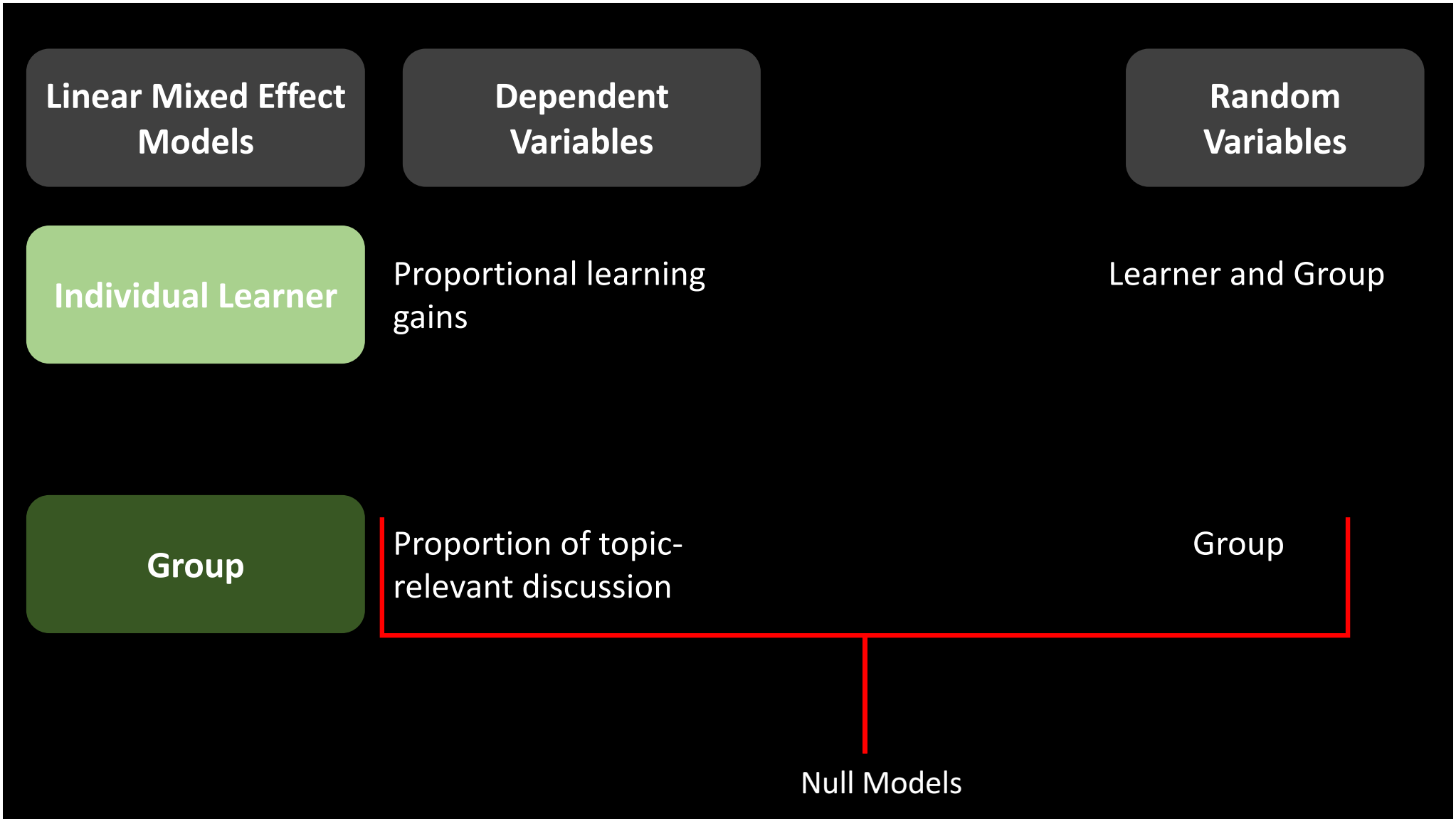
Learning across levels

Gerry Stahl

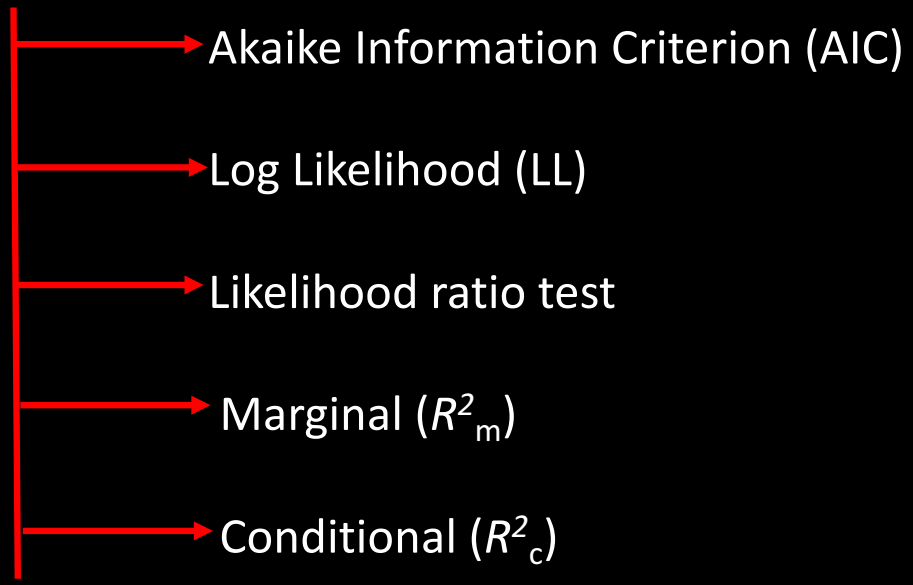
Published online: 16 February 2013
© International Society of the Learning Sciences, Inc. and Springer Science+Business Media New York 2013

The theme of this year's CSCL 2013 conference—"To see the world *and* a grain of sand: Learning across levels of space, time and scale"—targets a provocative challenge for CSCL, namely that the interactions of collaborative learning be understood, supported and analyzed at multiple levels. As the conference call puts it, "the attention to the theoretical, methodological and technological issues of addressing research at multiple levels is highly relevant to current research in CSCL, as well as to developing an emerging understanding of the epistemological and methodological issues that will shape our intellectual efforts well into the future" (http://isls.org/cscl2013).

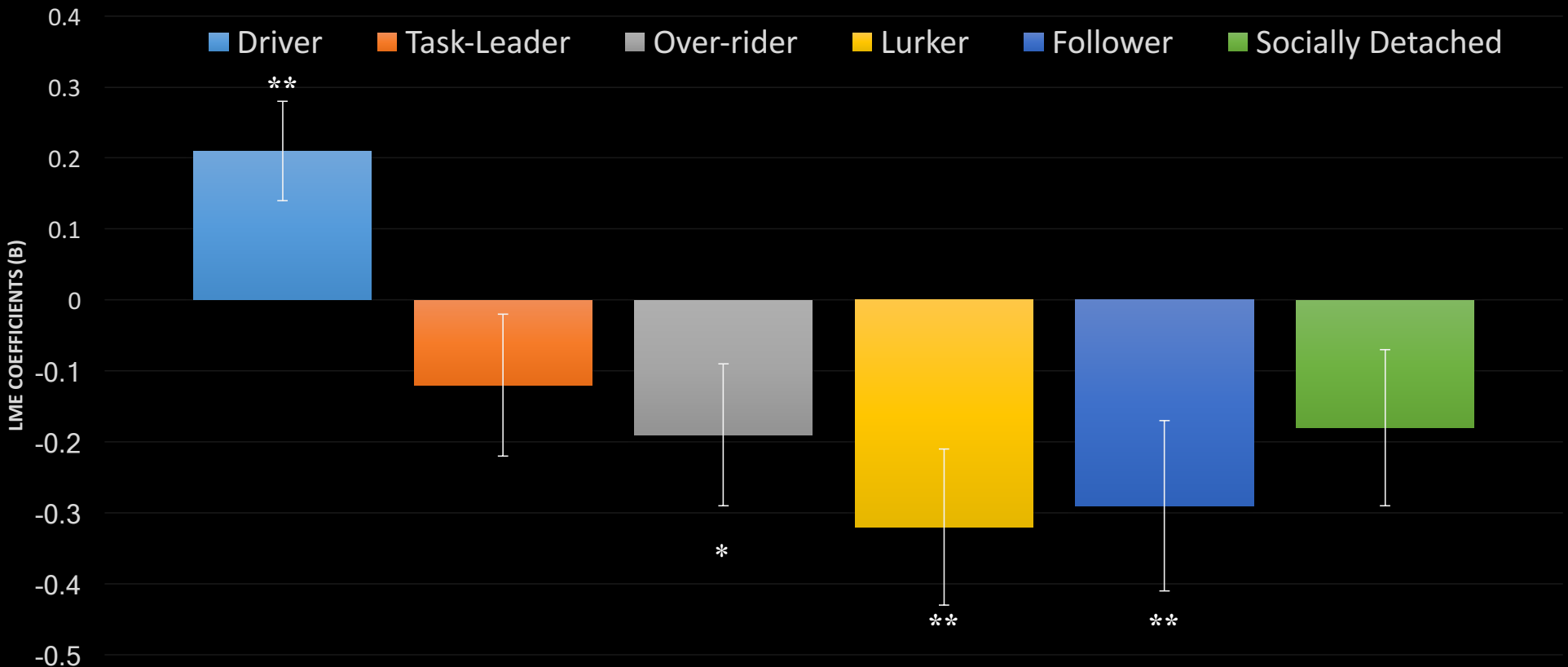
Linear Mixed Effect Models	Dependent Variables	Independent Variables	Random Variables
Individual Learner	Proportional learning gains	Identified roles	Learner and Group
Group	Proportion of topic-relevant discussion	Proportional occurrence of each identified role	Group


Linear Mixed Effect Models
Dependent Variables
Random Variables
Individual Learner
Proportional learning gains
Learner and Group
Group
Proportion of topic-relevant discussion
Group
Null Models

Linear Mixed Effect Models Evaluation



How do learners' roles influence individual learners' performance?

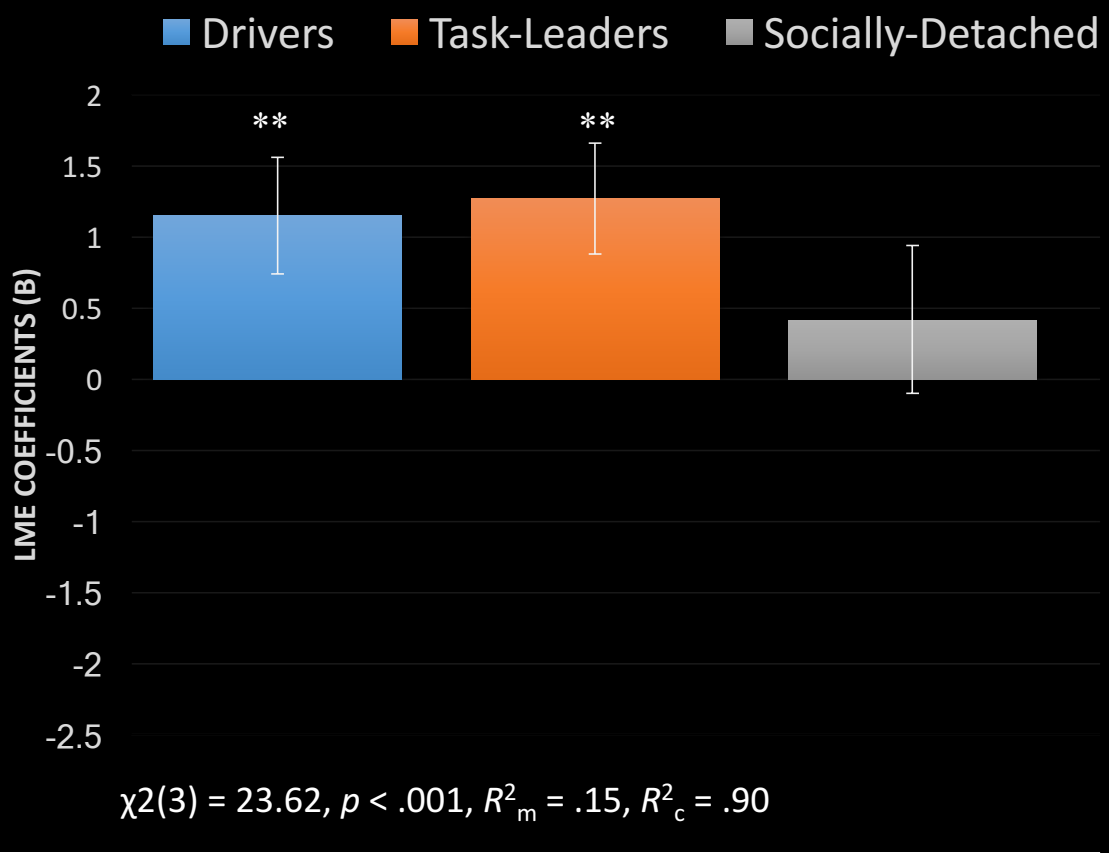


* $p < .05$, ** $p < .01$, *** $p < .001$; $N = 704$

$\chi 2(7) = 14.93$, $p = .001$, $R^2_m = .02$, $R^2_c = .95$.

How do learners' roles influence overall group performance?

Productive roles model



$\chi 2(3) = 23.62$, $p < .001$, $R^2_m = .15$, $R^2_c = .90$

Take Home

- Roles influence student and group outcomes
  - Drivers > Lurkers
  - Drivers = Task leaders and Socially-detached learners
    - *Difference in learning is not a result of the students simply being more prolific*

Driver

Task-leader

Socially-detached

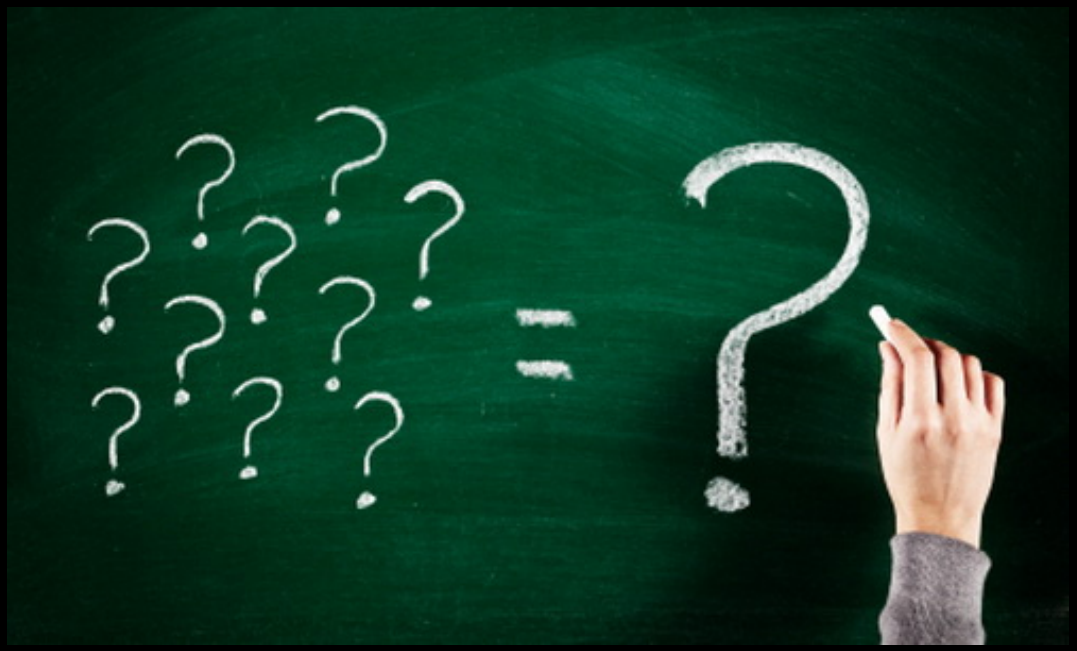
Over-rider

Follower

Lurker

- Optimal group composition ≠ simply high participating learners
  - Optimal group composition = high and low participators aware of and invested in the social climate of the group interaction
- Effect size differences

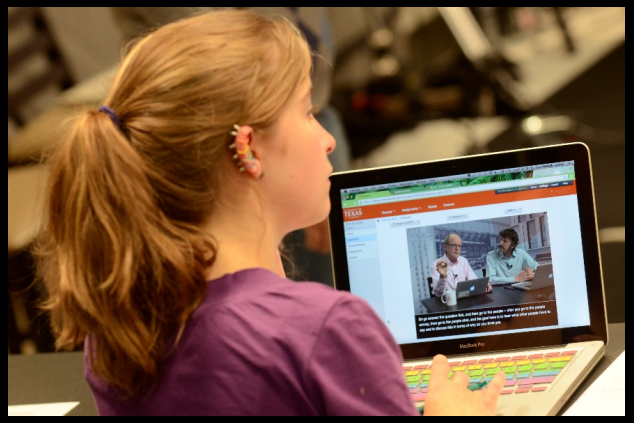
How well the identified clusters generalize to held out and completely different computer-mediated collaborative learning contexts?



SMOC: Synchronous Massive Online Class

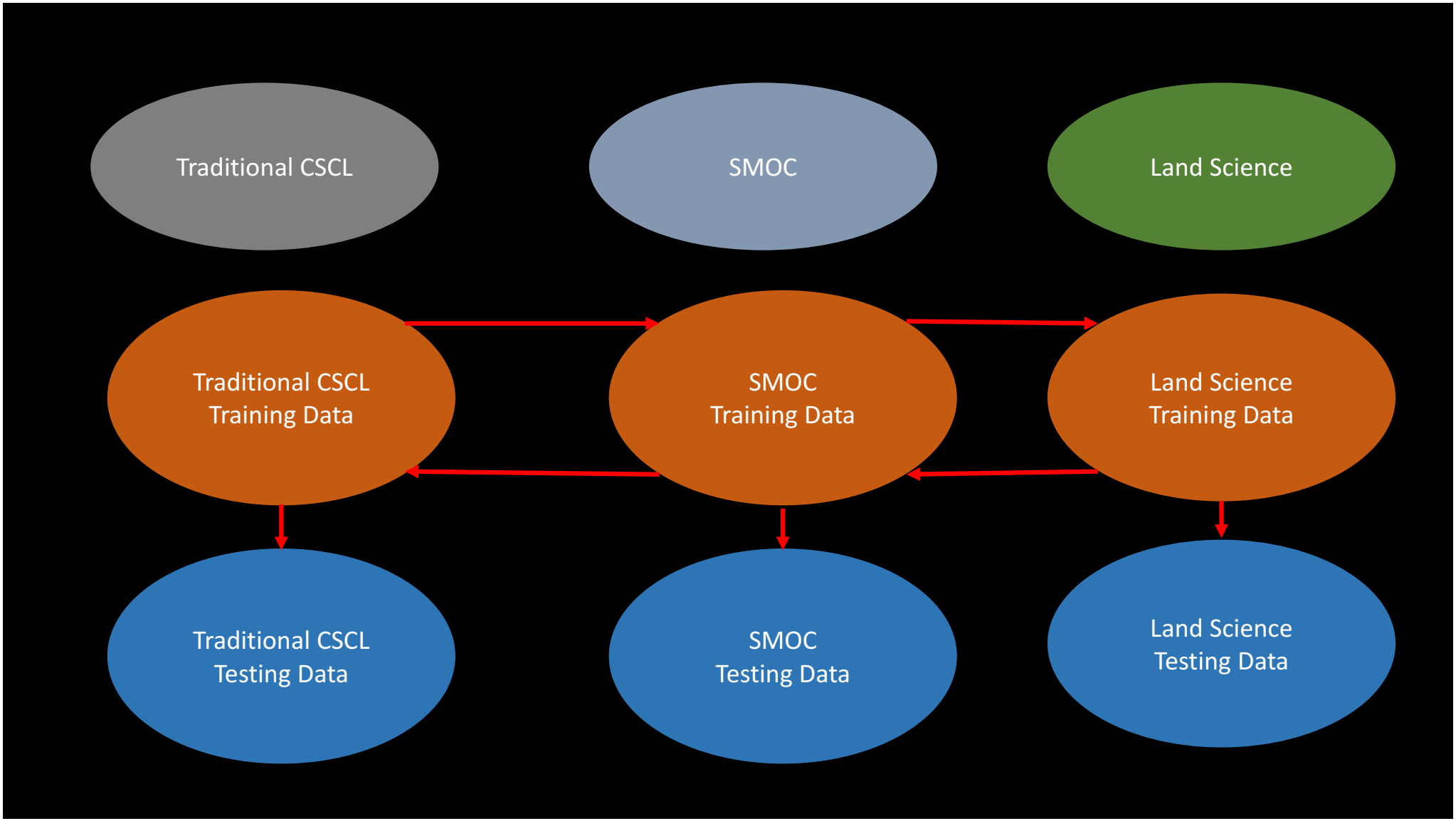
- Intro psychology course
- Students randomly assigned to groups
- 200-300 groups of 4-5 students per day
  - learner *N* = 1,713, group *N* = 3,380
- Interactions last 3-9 minutes, averaging 5 minutes
- Over 26 different chat topics
- Similar to the Traditional CSCL dataset, but larger and more distributed in terms of people and topics
- Students were in 9 chats groups throughout the semester

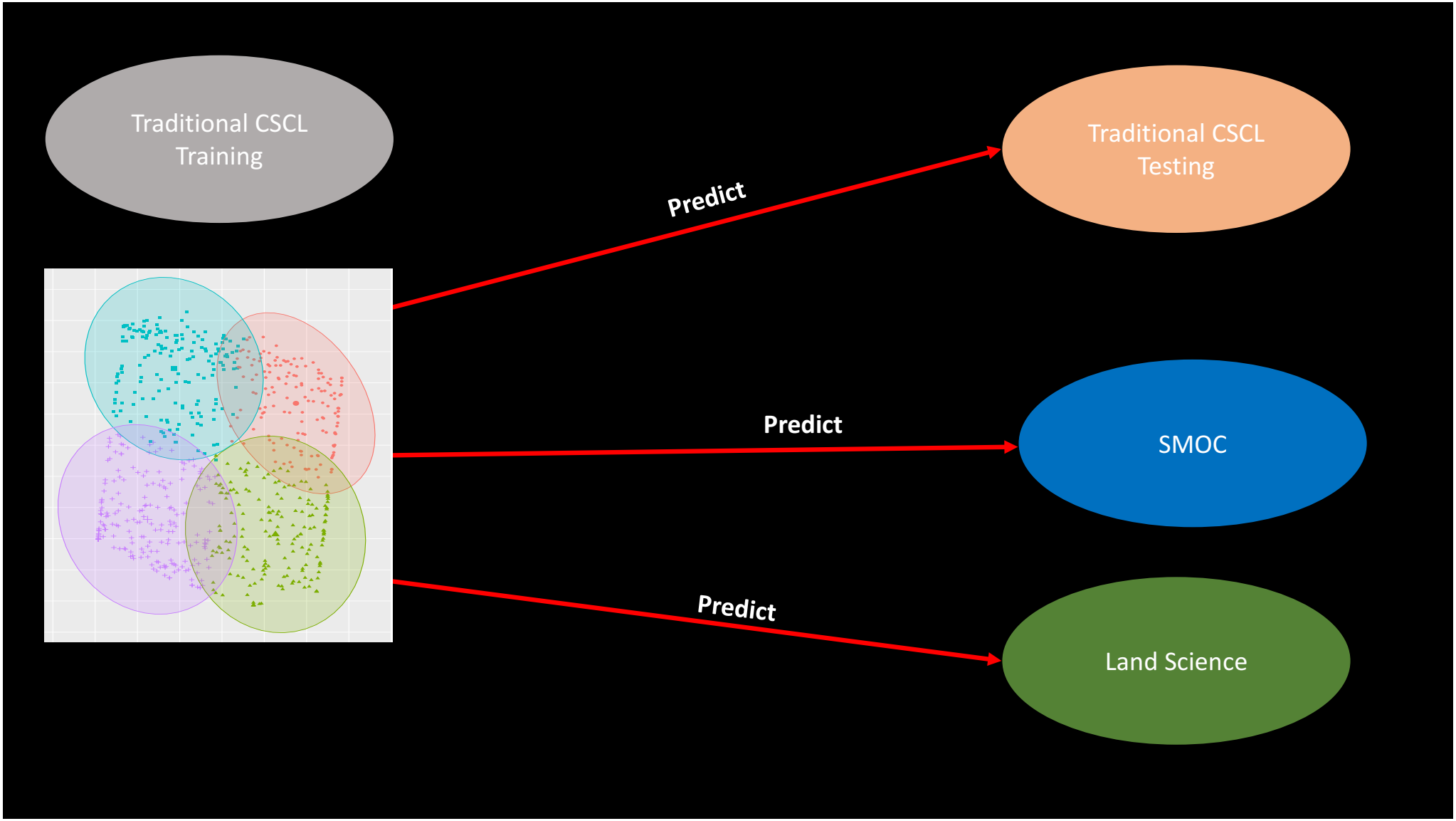
THE UNIVERSITY OF TEXAS AT AUSTIN



Land Science: A Virtual Internship

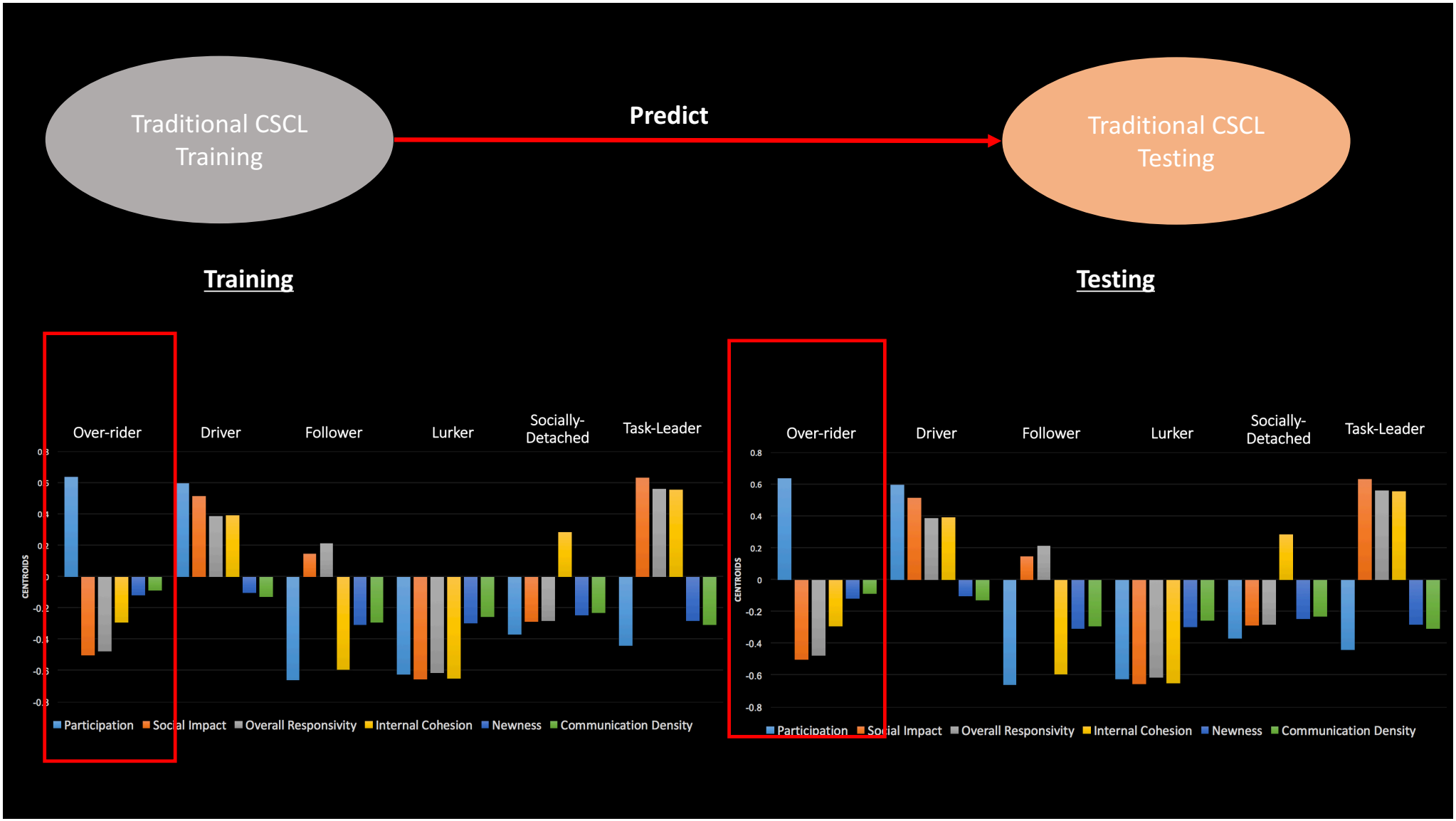
- Land Science is an interactive urban-planning simulation with **collaborative problem-solving** in an simulation environment
- Interns receive instructions and coaching from Mentors
- Interns participate in collaborative problem solving chat sessions to achieve collective goals
- learner *N* = 38, group *N* = 630

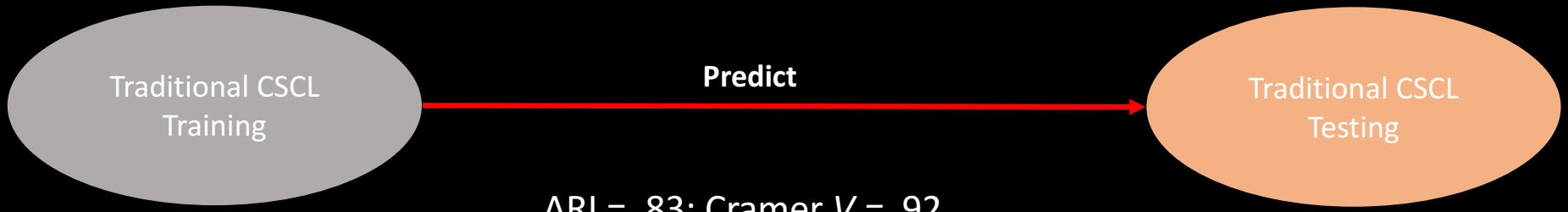

Traditional CSCL
SMOC
Land Science
Traditional CSCL Training Data
SMOC Training Data
Land Science Training Data
Traditional CSCL Testing Data
SMOC Testing Data
Land Science Testing Data


Traditional CSCL Training
Predict
Traditional CSCL Testing
Predict
SMOC
Predict
Land Science

Prediction Evaluation

- Cross-tabulation assessment
- Adjusted Rand Index (ARI)
  - computes the proportion of agreement between 2 cluster partitions & penalizes for any randomness in the overlap
  - Steinley (2004) considers ARI values greater than 0.90 - excellent, values greater than 0.80 - good, values greater than 0.65 - moderate, and values less than 0.65 - poor
- Cramer *V*
  - Effect size for the strength of the relationship between 2 cluster partitions

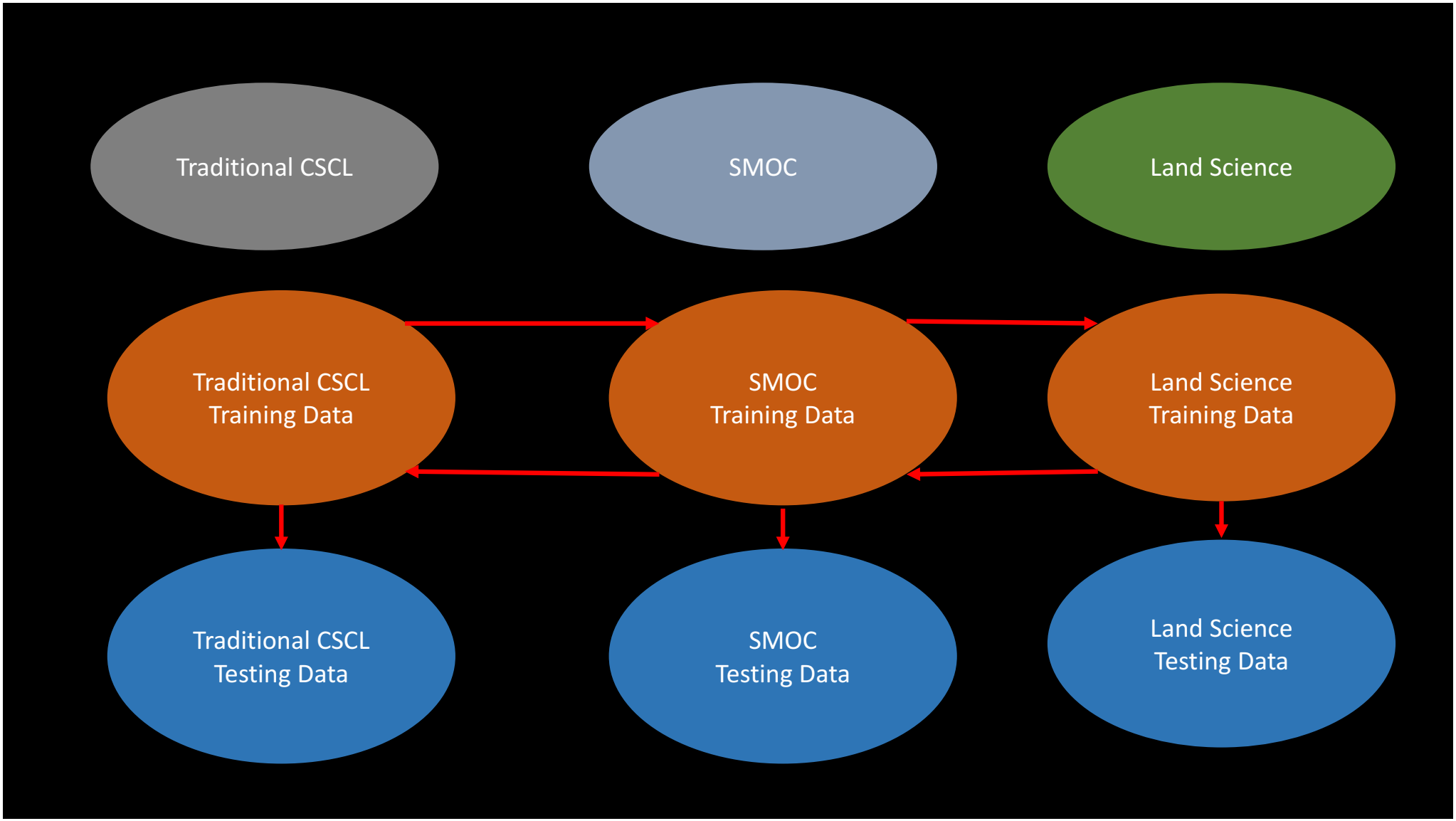

Traditional CSCL Training
Predict
Traditional CSCL Testing
Training
Testing
Over-rider
Driver
Follower
Lurker
Socially-Detached
Task-Leader
CENTROIDS
Participation
Social Impact
Overall Responsivity
Internal Cohesion
Newness
Communication Density



ARI = .83; Cramer *V* = .92

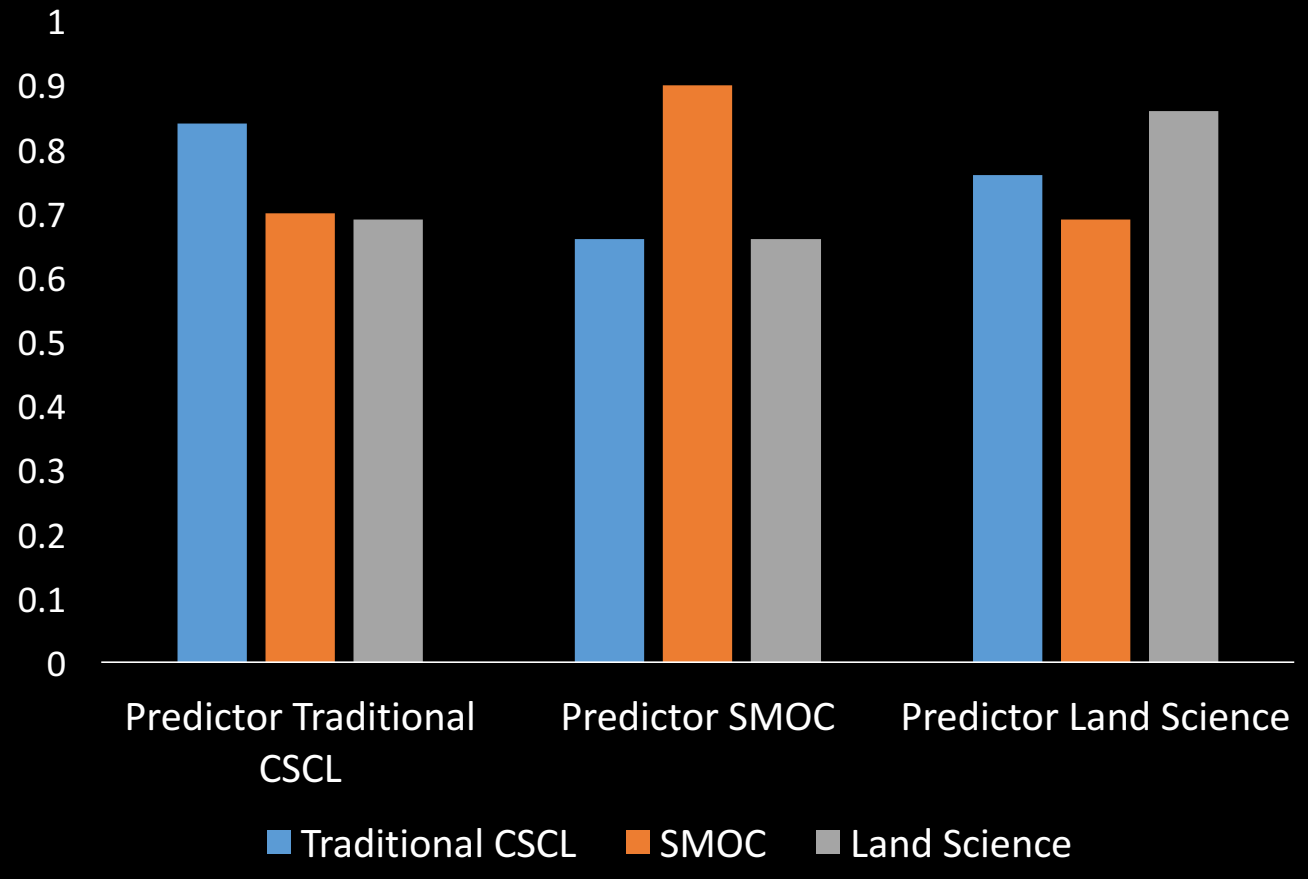
Cross-tabulation of the predicted and actual cluster assignments

Testing Clusters	Training Predicted Clusters					
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Cluster 1	32	0	0	0	0	0
Cluster 2	2	29	0	0	0	0
Cluster 3	0	0	15	2	1	0
Cluster 4	0	0	0	18	0	0
Cluster 5	4	0	0	1	13	0
Cluster 6	0	0	0	0	0	19


Traditional CSCL
SMOC
Land Science
Traditional CSCL Training Data
SMOC Training Data
Land Science Training Data
Traditional CSCL Testing Data
SMOC Testing Data
Land Science Testing Data

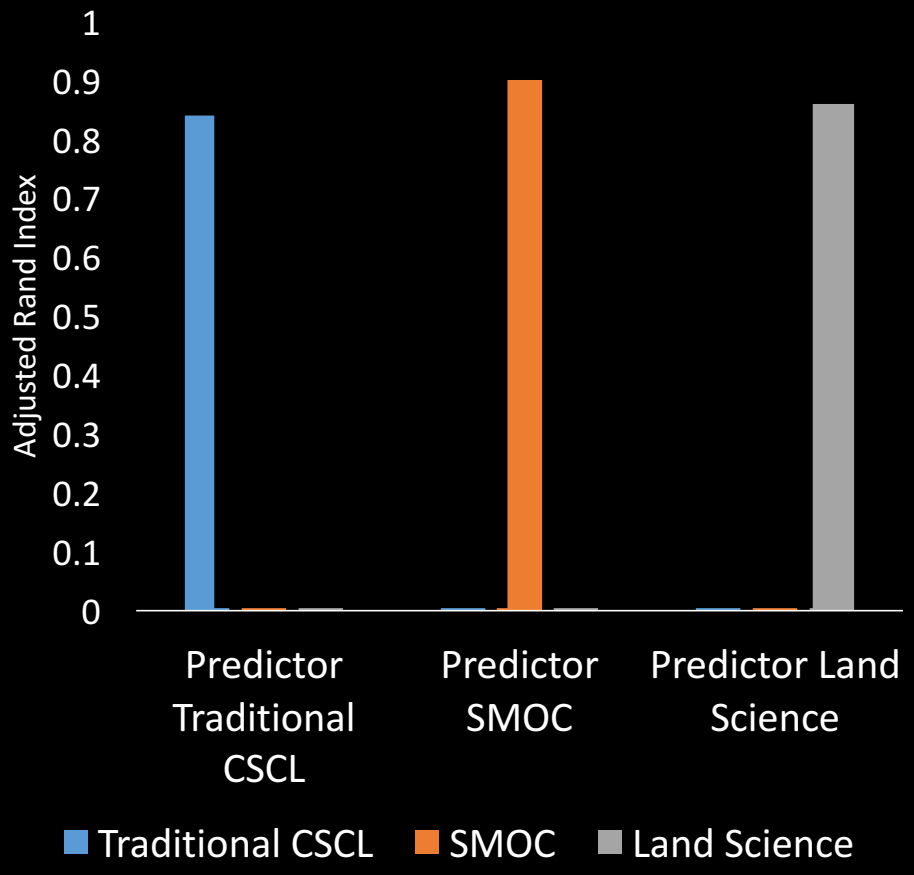
Internal & External Generalization

Six-Cluster Model



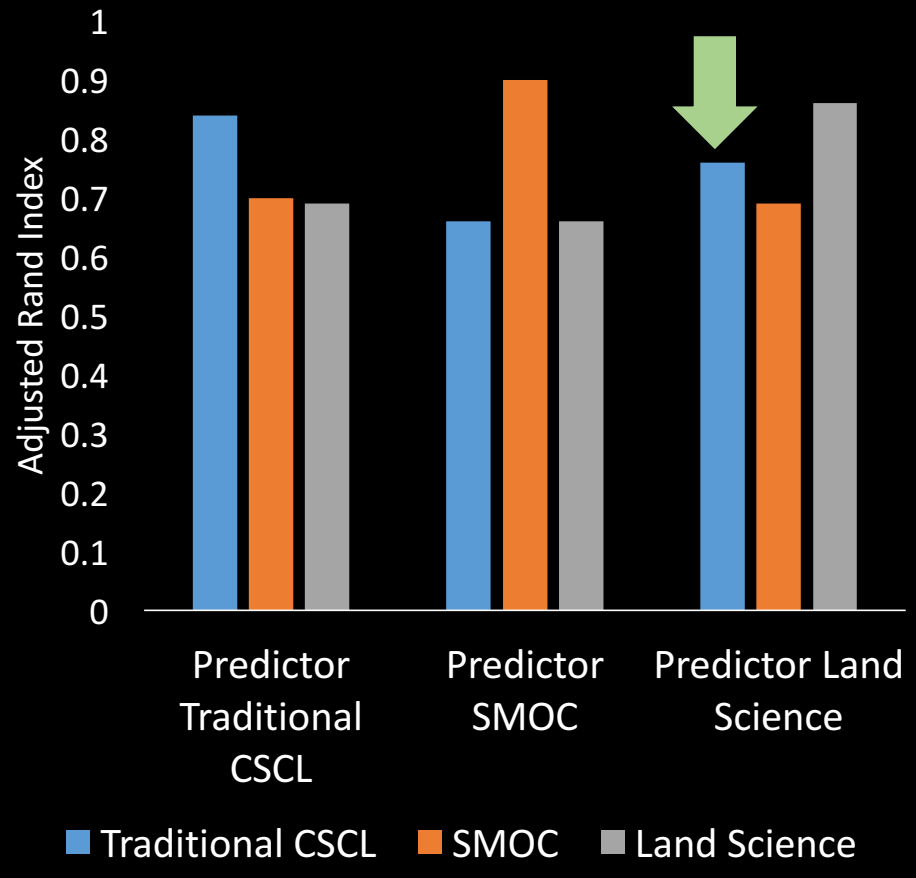
Internal & External Generalization

Six-Cluster Model



Internal & External Generalization

Six-Cluster Model



Take-Away

- The GCA method appears to be a robust method for identifying conversational roles
- We see good generalization of the roles both within and between datasets
  - But the roles seem to be context dependent, which is seen in how they do not generalize as well to the Land Science collaborative problem solving interactions
  - This does not mean the GCA is not a valid approach for identifying roles, just that care should be taken when transferring roles from one type of interaction to another

Onward and Upward: Preliminary findings

If roles are indeed an emergent property of interactions, then they will exhibit certain properties:

1. They should not be consistently or highly associated with trait-based characteristics
2. They will not be static, but instead will change in different context

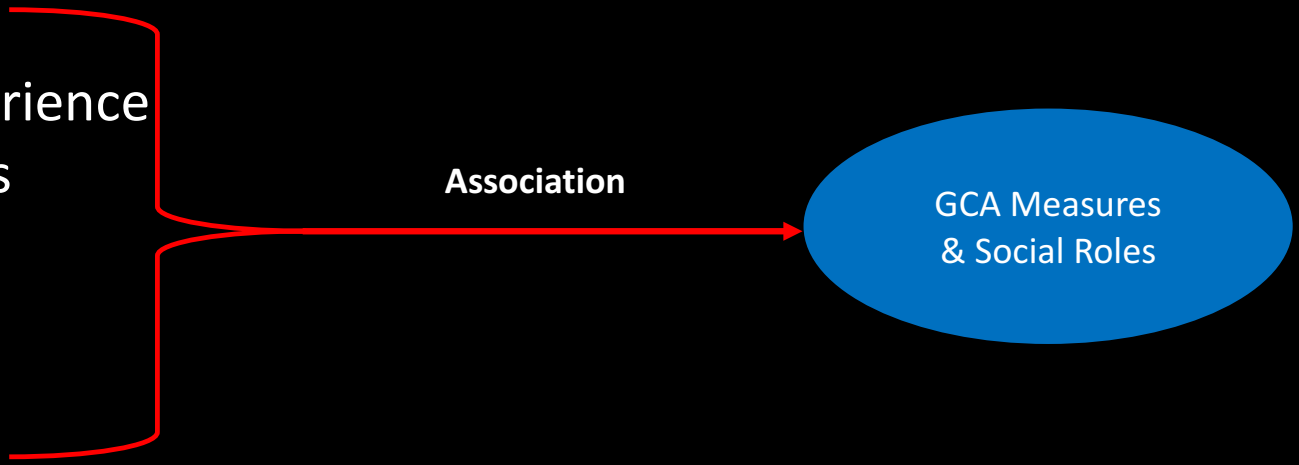
Dowell, et al., in prep

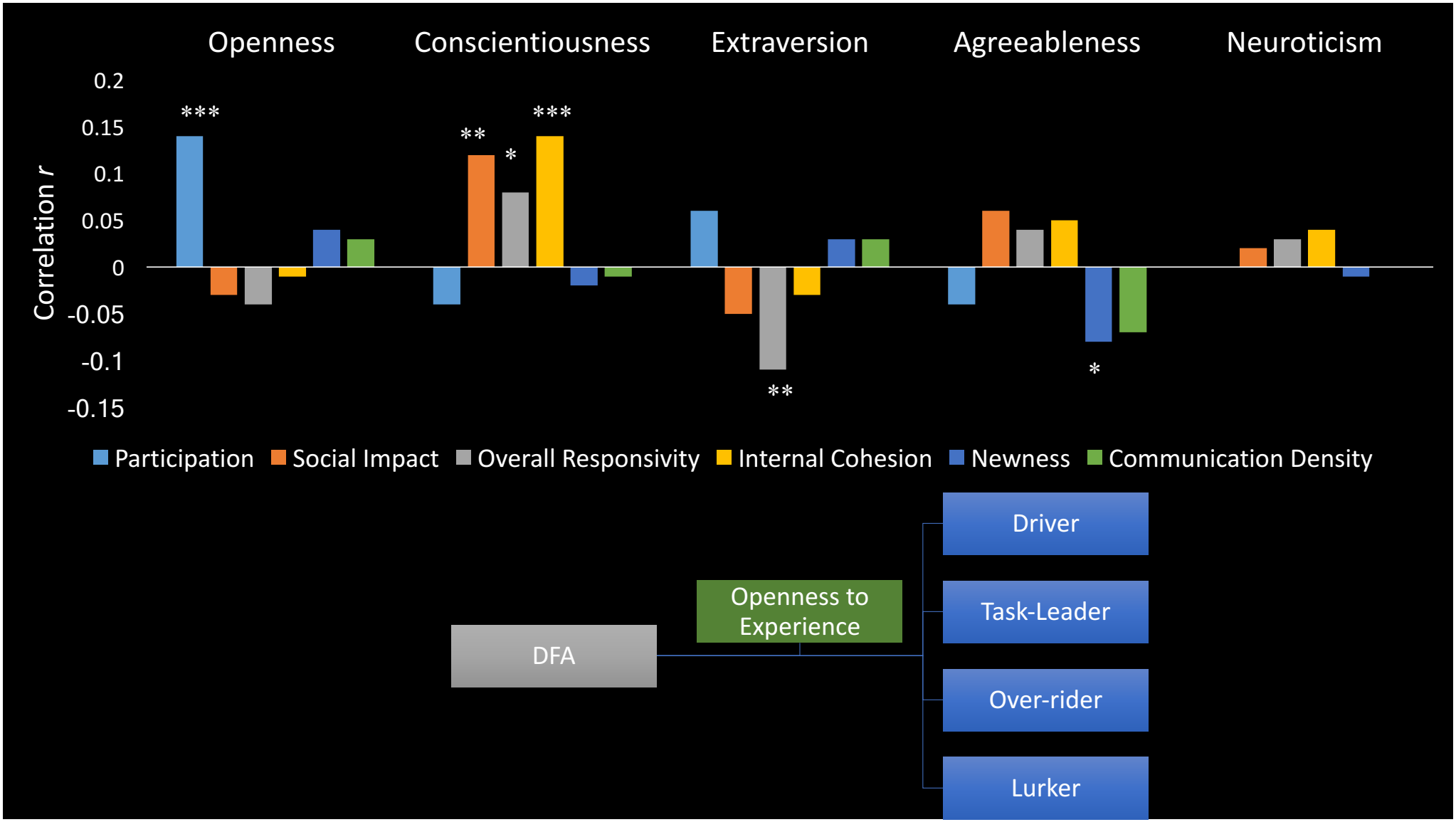
Claim 1. They should not be consistently or highly associated with trait-based characteristics

Correlation and Linear Discriminate Function Analyses

Traditional CSCL Big Five Personality Measures

1. Openness to Experience
2. Conscientiousness
3. Extroversion
4. Agreeableness
5. Neuroticism



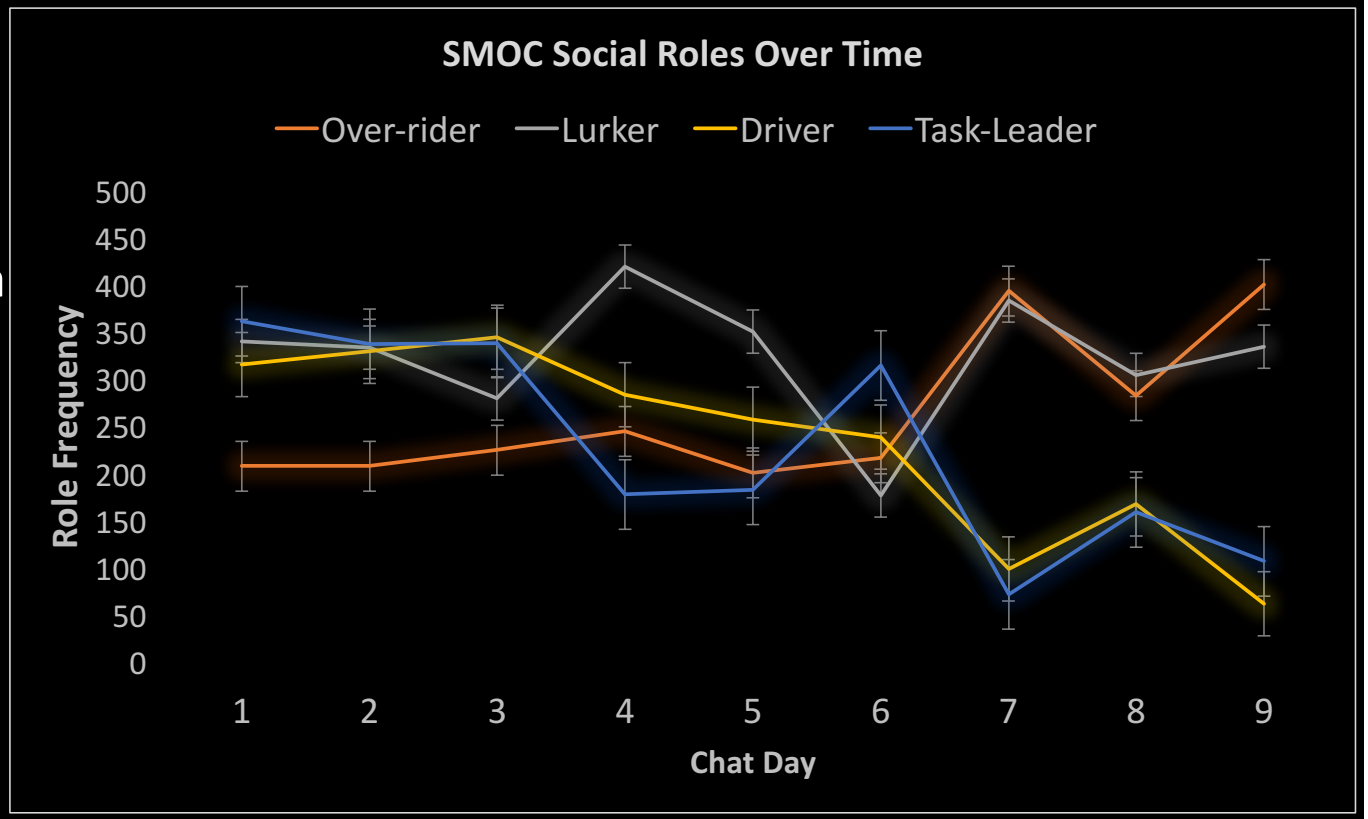

Openness
Conscientiousness
Extraversion
Agreeableness
Neuroticism
Correlation r
0.2
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
***
**
*
***
**
*
Participation
Social Impact
Overall Responsivity
Internal Cohesion
Newness
Communication Density
DFA
Openness to Experience
Driver
Task-Leader
Over-rider
Lurker

Claim 2. They will not be static, but instead will change in different contexts

SMOC Data Set

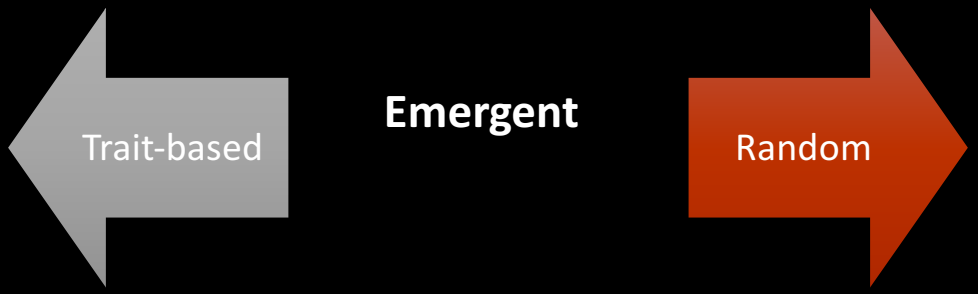
1. Qualitative look at the data

2. State Transition Networks



Take-Away

- The roles do not appear to be highly or consistently related to trait-based characteristics
- The roles are not static, but instead change in different contexts
  - Most of those transitions appear to support a more emergent perspective



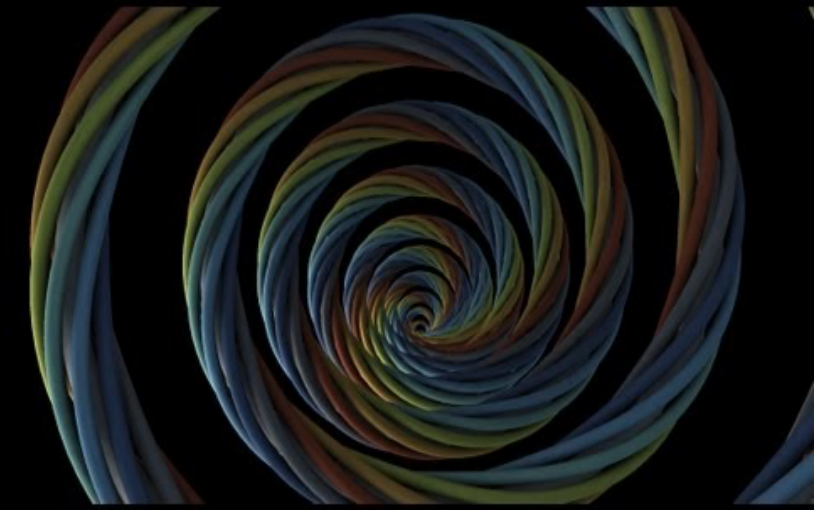
Conclusions

- The GCA appears to be a robust method for identifying conversational roles

- The identified roles have practical value in adding to our understanding of why some groups and students perform better than others

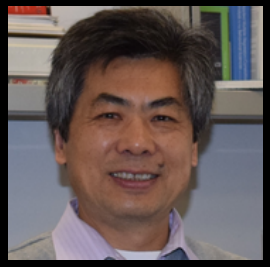
Next Steps: Diving Deeper

- Temporal dynamics
  - Right now we are looking at averages
  - It is possible that an individual shifts roles throughout an interaction or over longer periods of time as they gain experience
- Other variables
  - Internal (linguistic)
    - Affect, Topic Relevance
  - External (individual/contextual)
    - #s of resources viewed
    - Other demographic variables
- Other contexts and outcomes
    - Crowd sourcing design interactions
      - OPEN IDEO- creativity



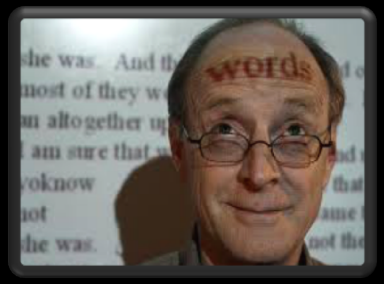
Many Thanks!

Art Graesser



Zhiqiang Cai

THE UNIVERSITY OF MEMPHIS | Institute for Intelligent Systems



Jamie Pennebaker

Tristan Nixon