Introduction to Qualitative Coding with nCoder

Amanda Siebert-Evenstone, Ph.D.
alevenstone@wisc.edu

Seung B. Lee, Ph.D.
seung.lee@pepperdine.edu

Brendan Eagan, Ph.D.
beagan@wisc.edu

Nelson Institute for Environmental Studies
UNIVERSITY OF WISCONSIN–MADISON

PEPPERDINE
UNIVERSITY

epistemic analytics
turning data into meaning


Quantitative Research
Qualitative Research

Categories of Spoons

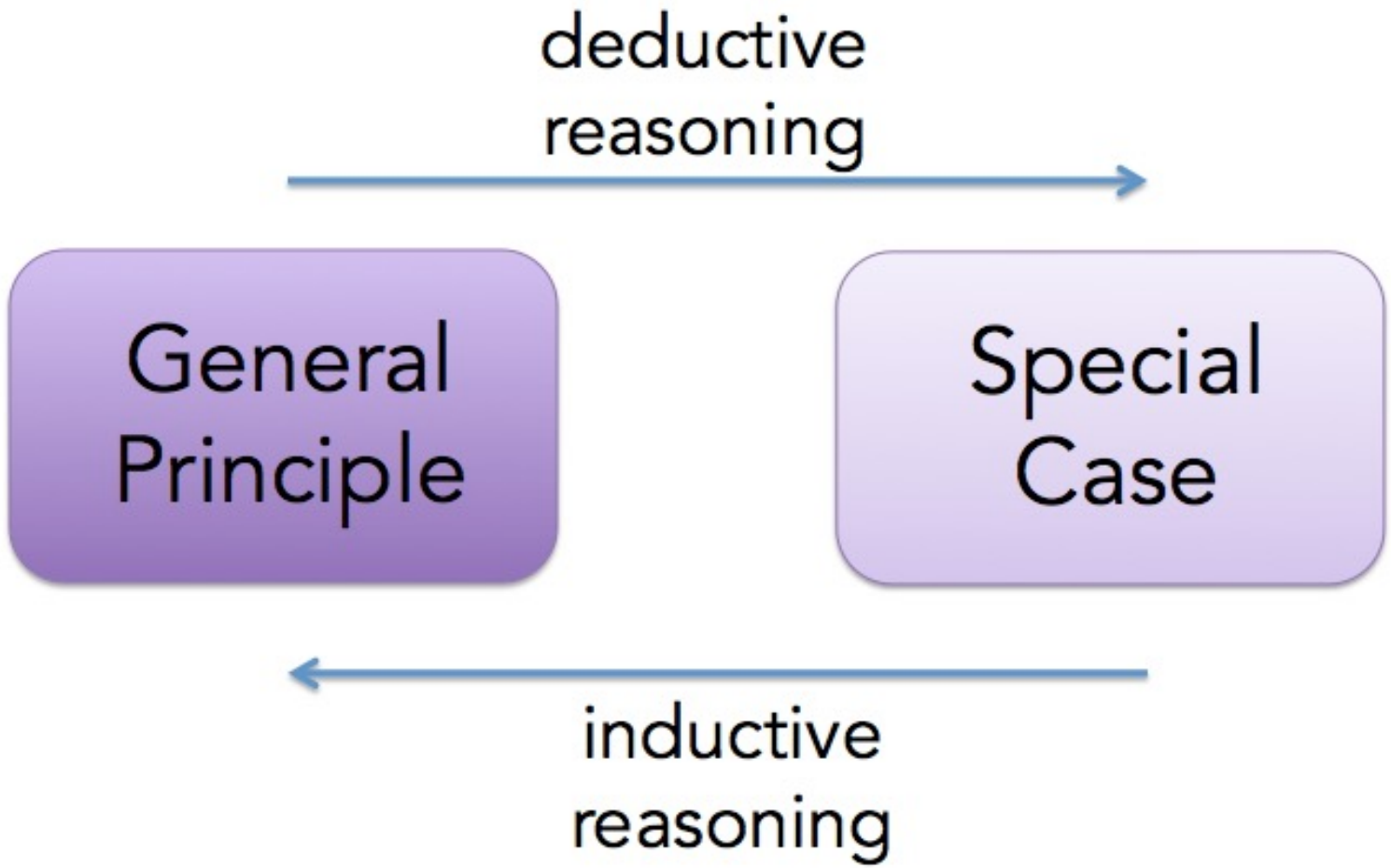
https://go.wisc.edu/0nxl70

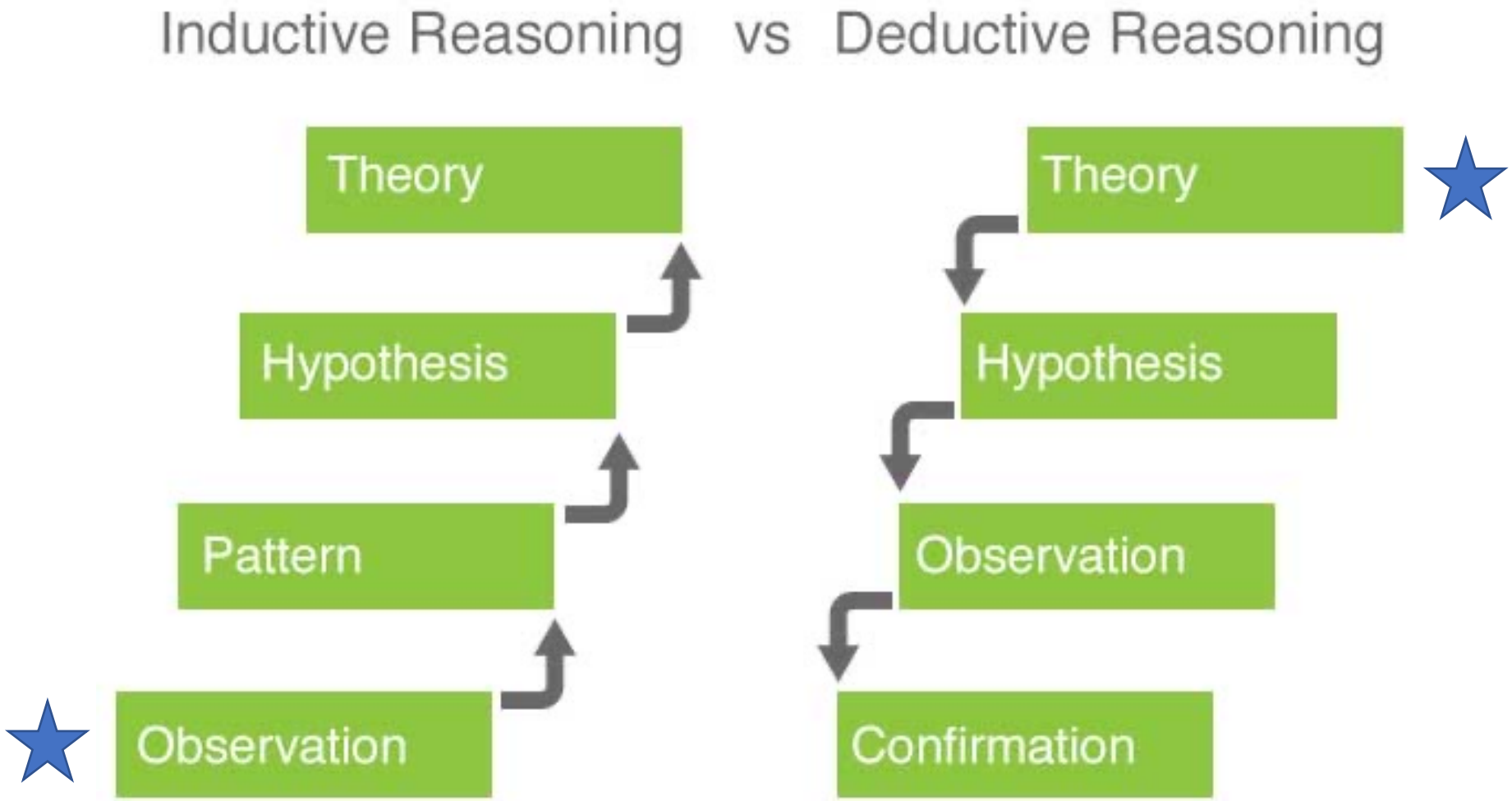
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

Storytelling

Storytelling

What is our systematic explanation?


deductive
reasoning
General
Principle
Special
Case
inductive
reasoning


Inductive Reasoning vs Deductive Reasoning
Theory
Hypothesis
Pattern
Observation
Theory
Hypothesis
Observation
Confirmation

Top-Down

(aka a priori, theoretical, deductive)

Start with theory

Synonyms or Word Associations

Existing coding schemes

Top-Down

(aka a priori, theoretical, deductive)

Start with theory

Synonyms or Word Associations

Existing coding schemes

Bottom-Up

(aka Grounded theory, emergent coding, inductive)

N-grams

TFIDF

Topic Models

Word Counter or TextRazor

Top-Down

(aka a priori, theoretical, deductive)

Start with theory

Synonyms or Word Associations

Existing coding schemes

Bottom-Up

(aka Grounded theory, emergent coding, inductive)

N-grams

TFIDF

Topic Models

Word Counter or TextRazor

*Actually Reading your Data

SIEBERT-EVENSTONE'S MAXIM

WHEN IN DOUBT, **READ YOUR DATA.**

Imagine that you have a special instrument that allows you to see what makes up odor.

The large circle in the drawing represents a spot that is magnified many times, so you can see it up close.

Create a model of what you would see if you could focus on one tiny spot in the area between the jar and your nose.

What is this about?

Imagine that you have a special instrument that allows you to see what makes up odor.

The large circle in the drawing represents a spot that is magnified many times, so you can see it up close.

Create a model of what you would see if you could focus on one tiny spot in the area between the jar and your nose.

What is this about?

Imagine that you have a special instrument that allows you to see what makes up odor.

The large circle in the drawing represents a spot that is magnified many times, so you can see it up close.

Create a model of what you would see if you could focus on one tiny spot in the area between the jar and your nose.

What is this about?

Science

Drawing

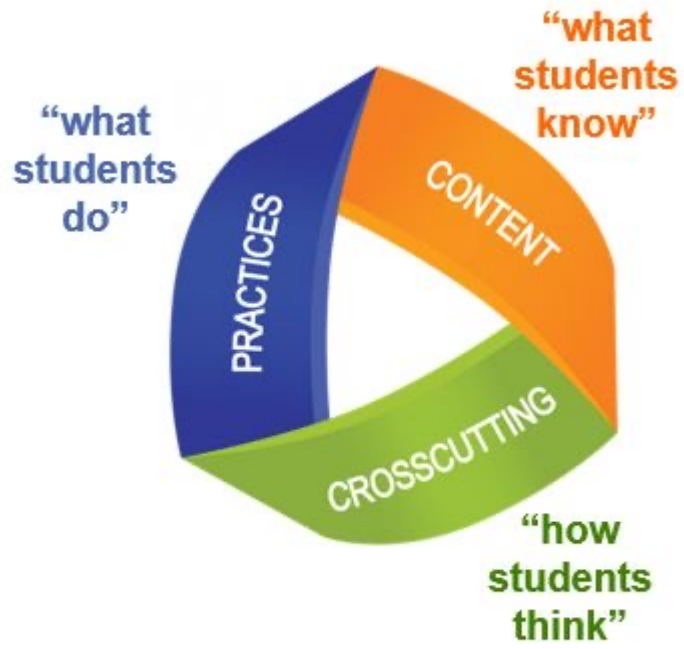
Modeling

Hypothesizing

“imagine”
-in vivo code

Coding Process

1. Identify common theories or ideas about the topic
1. Read and get to know the data



Quoted text from Peter A'Hearn

Top Keyword Density

Top 10 Exclude grammar words ON

1 Word | 2 Words | 3 Words

1. data	80 *(1.5%)*
2. learning	51 *(0.9%)*
3. researchers	36 *(0.7%)*
4. codes	35 *(0.7%)*
5. quantitative	33 *(0.6%)*
6. qualitative	33 *(0.6%)*
7. how	30 *(0.6%)*
8. big	29 *(0.5%)*
9. analysis	28 *(0.5%)*
10. deep	25 *(0.5%)*

630

Coding Process

1. Identify common theories or ideas about the topic
1. Read and get to know the data
2. Describe each line

How do you study / what is the best way for you to study?	Description
It depends on what I'm trying to learn	
Quizlet	
Flashcards, highlight notes	
It depends on the subject	
Have someone quiz me on the material	
I like using the Quizlet when I need note memorization and vocabulary. For more complex topics, I always do well reading from a textbook and taking handwritten notes as well as completing or working through practice problems.	

Coding Process

1. Identify common theories or ideas about the topic
1. Read and get to know the data
2. Describe each line
3. Identify ideas or codes (quizzing, context-based, processing info)

Coding Process

1. Identify common theories or ideas about the topic
1. Read and get to know the data
2. Describe each line
3. Identify ideas or codes
4. Building wordlists (we'll get to this later)
5. Building categories of codes

What should I code?

- What's interesting?
- Why is it interesting?
- Why am I interested in that? (Richards, 2009)

From Hatch 2002:

- Similarity (things happen the same way)
- Difference (they happen in predictably different ways)
- Frequency (they happen often or seldom)
- Sequence (they happen in a certain order)
- Correspondence (they happen in relation to other activities or events)
- Causation (one appears to cause another) (p.155)

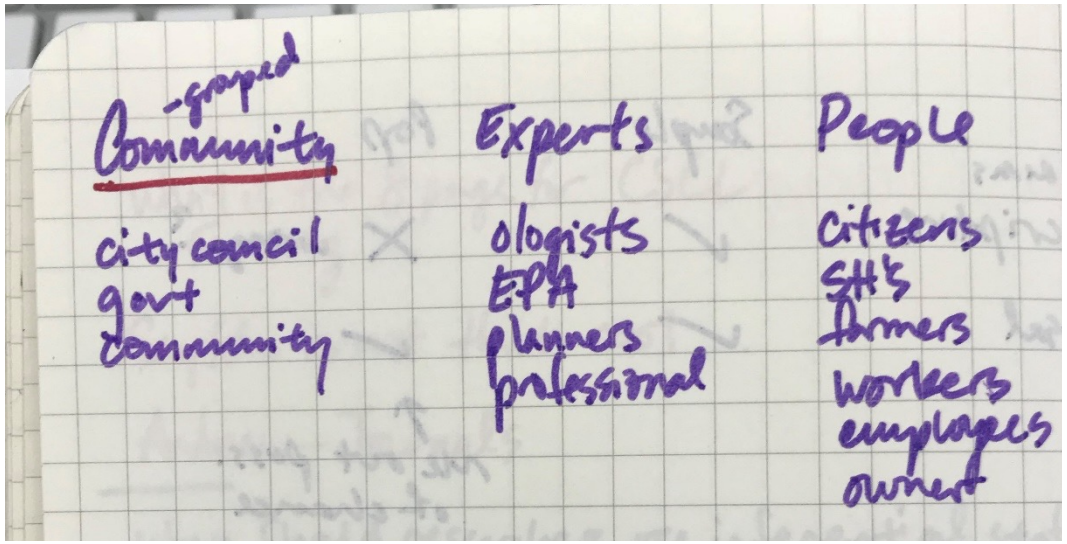
(Amanda keeps a cheat sheet and multiple books of ideas to help inspire ideas)

Second (to nth) cycle of coding

- Recode the data because more accurate words and phrases were discovered for the original codes
- Merge together similar codes
- Separate codes that are too large
- Infrequent codes will be assessed for their utility (then kept or dropped)

Memos

‘(A memo is) the theorizing write-up of ideas about codes and their relationships as they strike the analyst while coding... it can be a sentence, a paragraph or a few pages... it exhausts the analyst’s momentary ideation based on data with perhaps a little conceptual elaboration’ (Glaser, 1978: 83)



	SPRAWL	INFILL / SMART GROWTH
Density	Lower	Higher
Activities / services / Goods	Dispersed, regional requires driving Cotger (Costco)	Clustered local, smaller Trader Joe's H: Fight
Growth	Greenfield	Brownfield Greyfield
	"urban periphery"	
Transport	Autos. Poorly suited for walking or biking + transit	Multimodal supports transit options
Connectivity	Hierarchical road network w/ many unconnected roadways	Highly connected, allows direct travel
Planning	Unplanned. little coordination btw SH & jurisdictions	Planned
Public Space	Emphasis on Private realms (yards, malls, gated communities, clubs)	Public Realms (Shopping streets, parks)
	Irregular Settlement discontinuous, multiple centers	Vacant parcels OR redevelopment
Buildings	Low height, homogenous, single use	Mixed use

Memo

Memos are ways of summarizing where you are at during your analysis and potential interpretations you may have about your data.

- Codes, categories, and their relationships
- Initial thoughts on data analysis
- pulling together incidents that appear to have commonalities
- proposals for a specific new pattern code
- when the analyst does not have a clear concept in mind but is struggling to clarify one

Role of Researcher

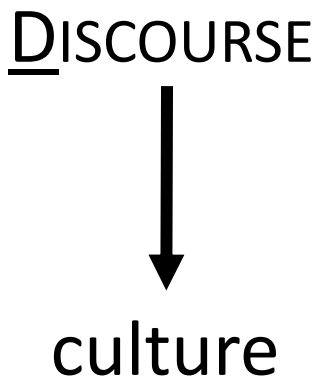
YOU are the data collection and analysis instrument

- You take notes and decide what topics to record
- What questions do you ask or not ask?
- What do you deem important?
- What are your implicit/explicit theories?
- What do you value?

"A DISCOURSE is a socially accepted association among ways of using language, of thinking, feeling, believing, valuing, and of acting that can be used to identify oneself as a member of a socially meaningful group… or to signal (that one is playing) a socially meaningful role."

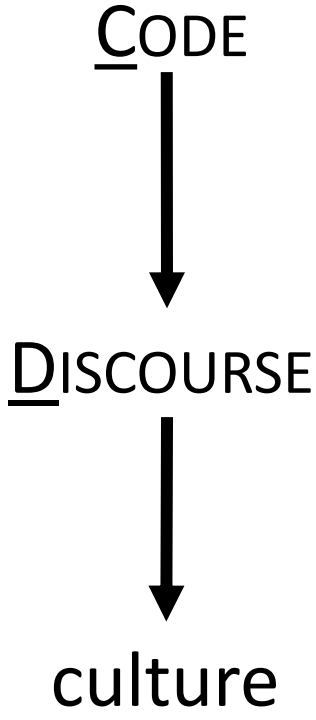
- Jim Gee

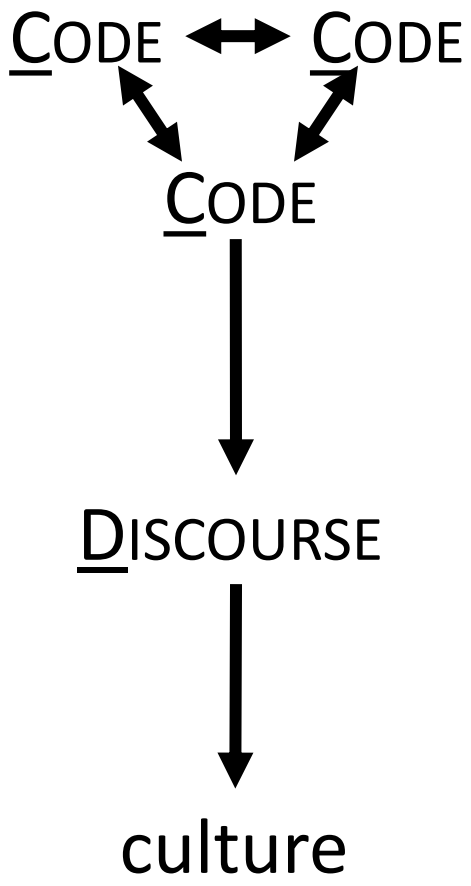
Learning is a process of Enculturation



CODES

Culturally-relevant and meaningful aspects of a DISCOURSE




CODE
CODE
CODE
DISCOURSE
culture

CODES

Culturally-relevant and meaningful aspects of a DISCOURSE

codes

Things that count as evidence or *warrants* for CODES

Codebooks

Names	Definition	Examples
Performance Metrics	Discussion of one or more criteria for device functionality: agility, payload, cost, recharge interval, and/or safety.	***I thought that safety near the maximum was not very good (close to 225 - one had 218 RPN), but other than that I was fine with the safety as long as it was around 200 or lower.***

Codebooks

Names	Definition	Examples
Performance Metrics	Discussion of one or more criteria for device functionality: agility, payload, cost, recharge interval, and/or safety.	*I thought that safety near the maximum was not very good (close to 225 - one had 218 RPN), but other than that I was fine with the safety as long as it was around 200 or lower.*
CODES	CODES	codes

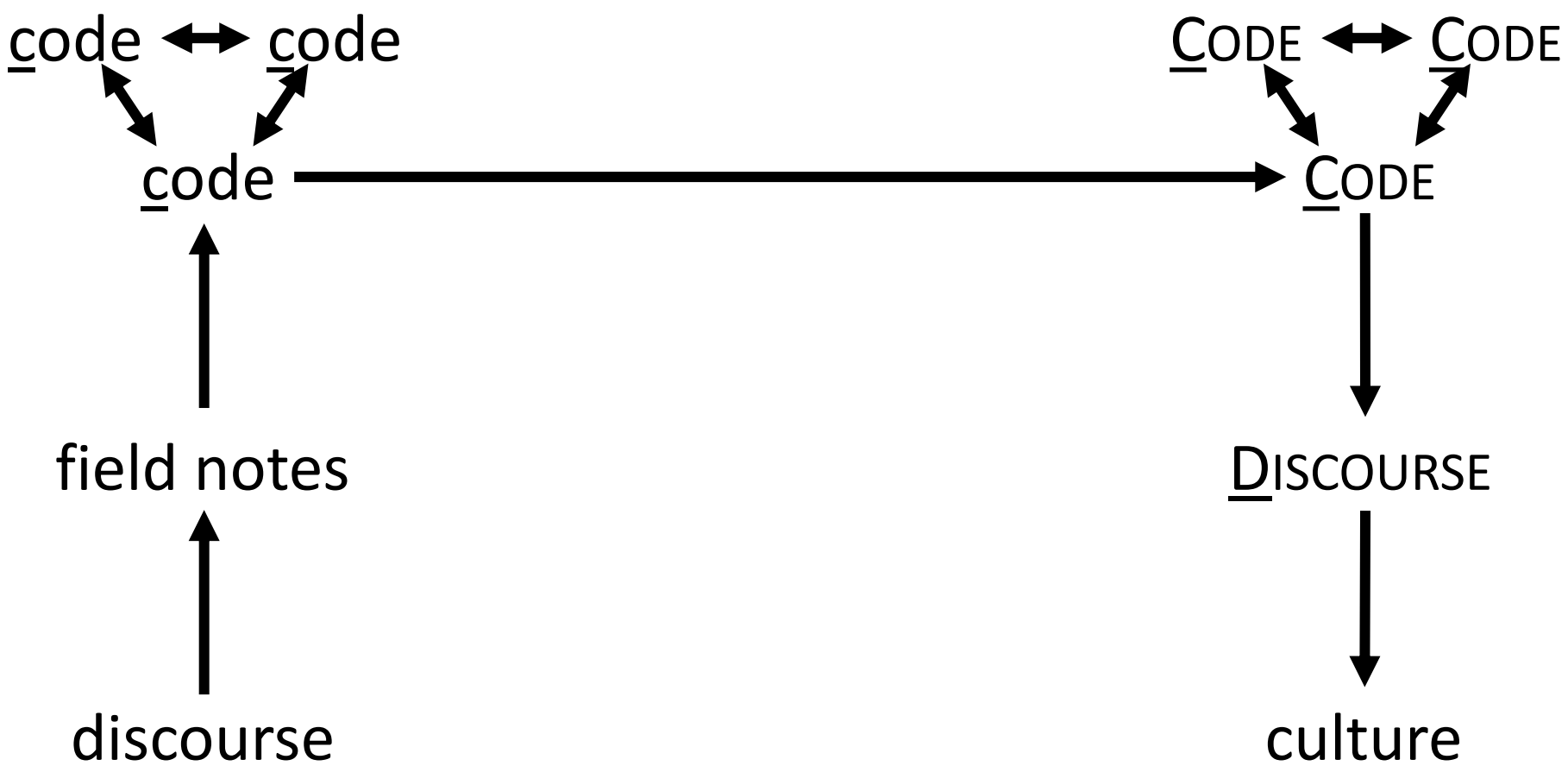
CODES

Culturally-relevant and meaningful aspects of a DISCOURSE

Grip

codes

Things that count as evidence or *warrants* for CODES


code ↔ code
code
CODE ↔ CODE
CODE
field notes
DISCOURSE
discourse
culture

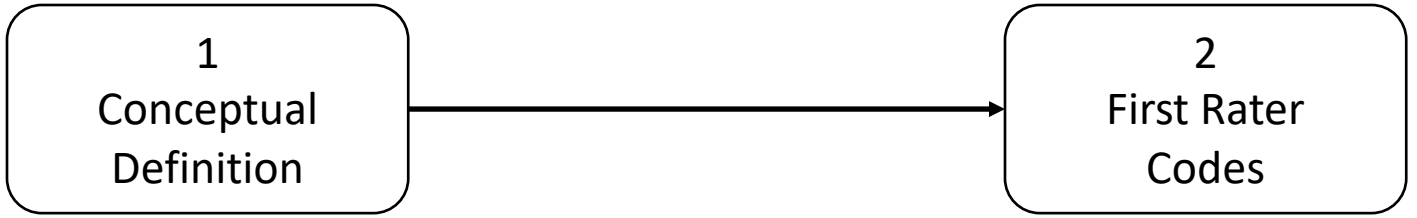

code ↔ code
code
CODE ↔ CODE
CODE
Thick Description
field notes
discourse
DISCOURSE
culture

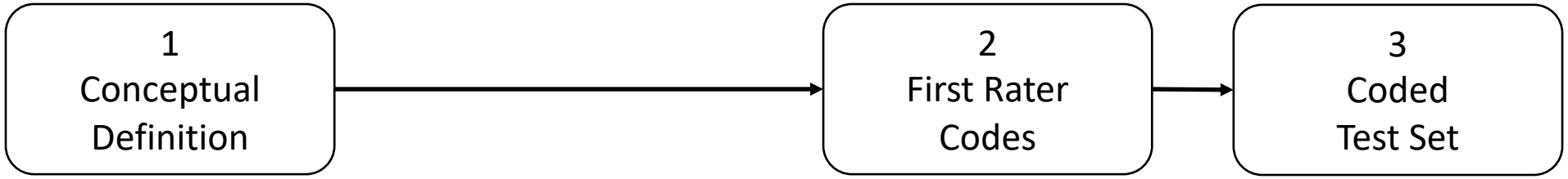

code ↔ code
code
CODE ↔ CODE
CODE
Thick Description
field notes
DISCOURSE
discourse
culture

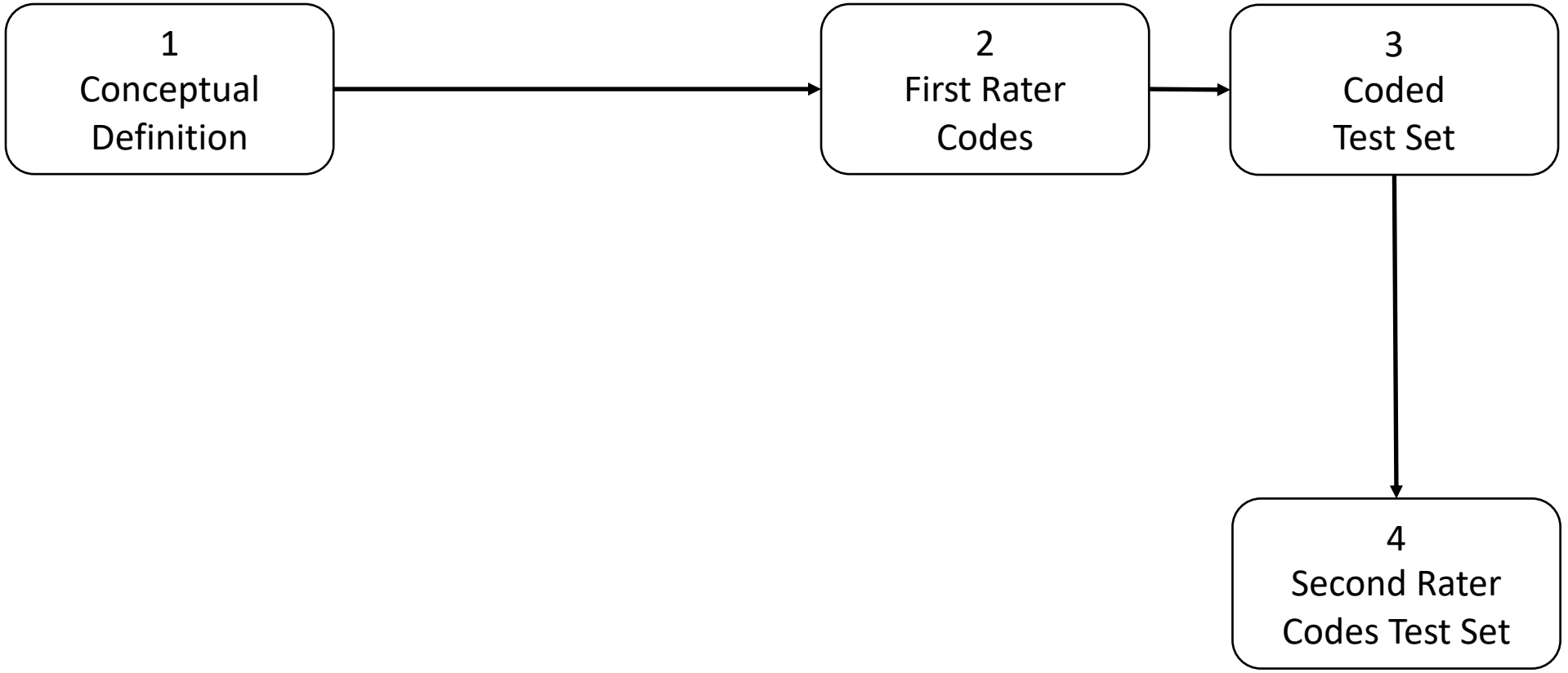

code ↔ code
CODE ↔ CODE
code
CODE
Thick Description
field notes
DISCOURSE
discourse
culture

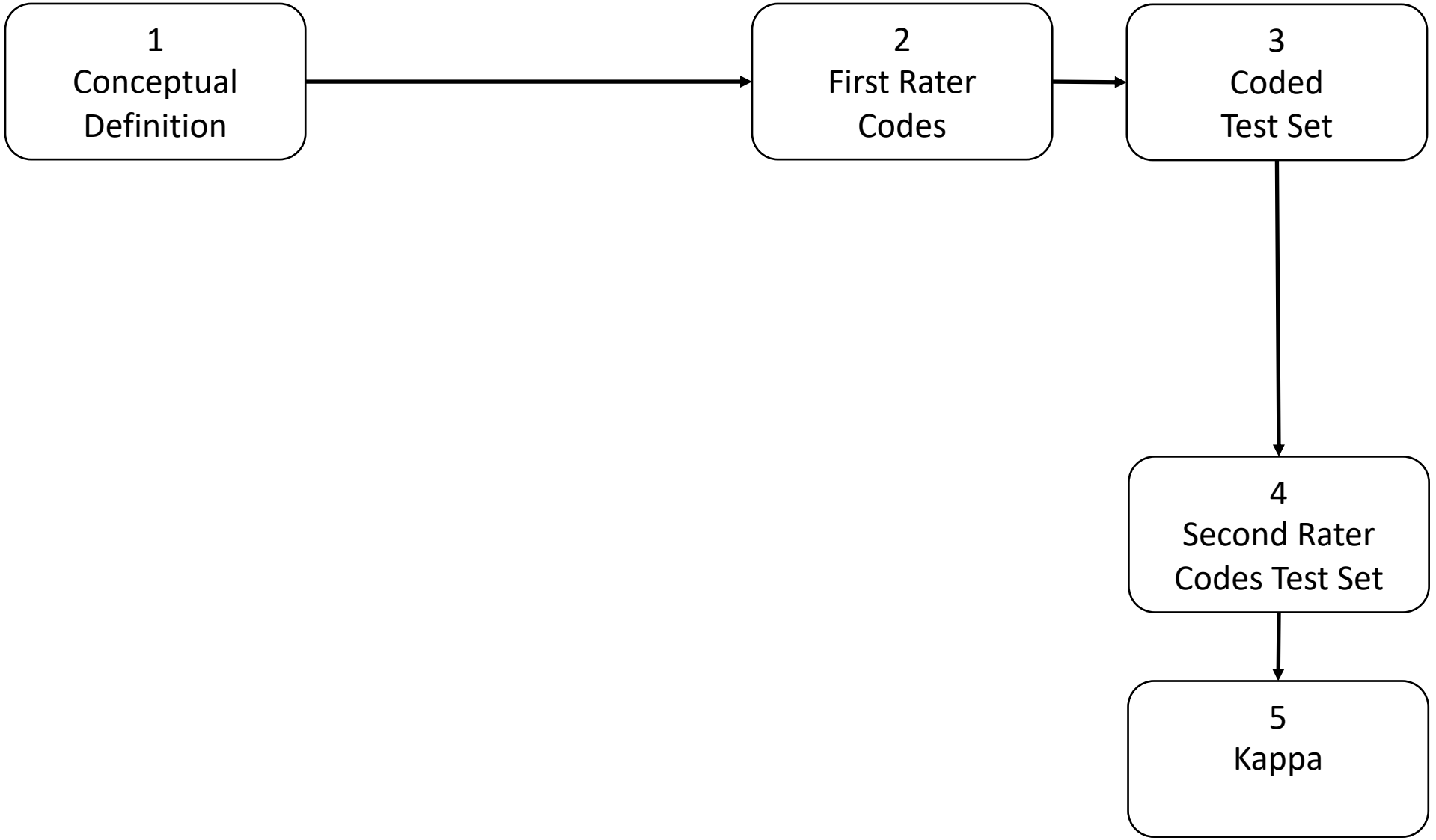

code ↔ code
CODE ↔ CODE
code
CODE
Thick Description
field notes
DISCOURSE
discourse
culture

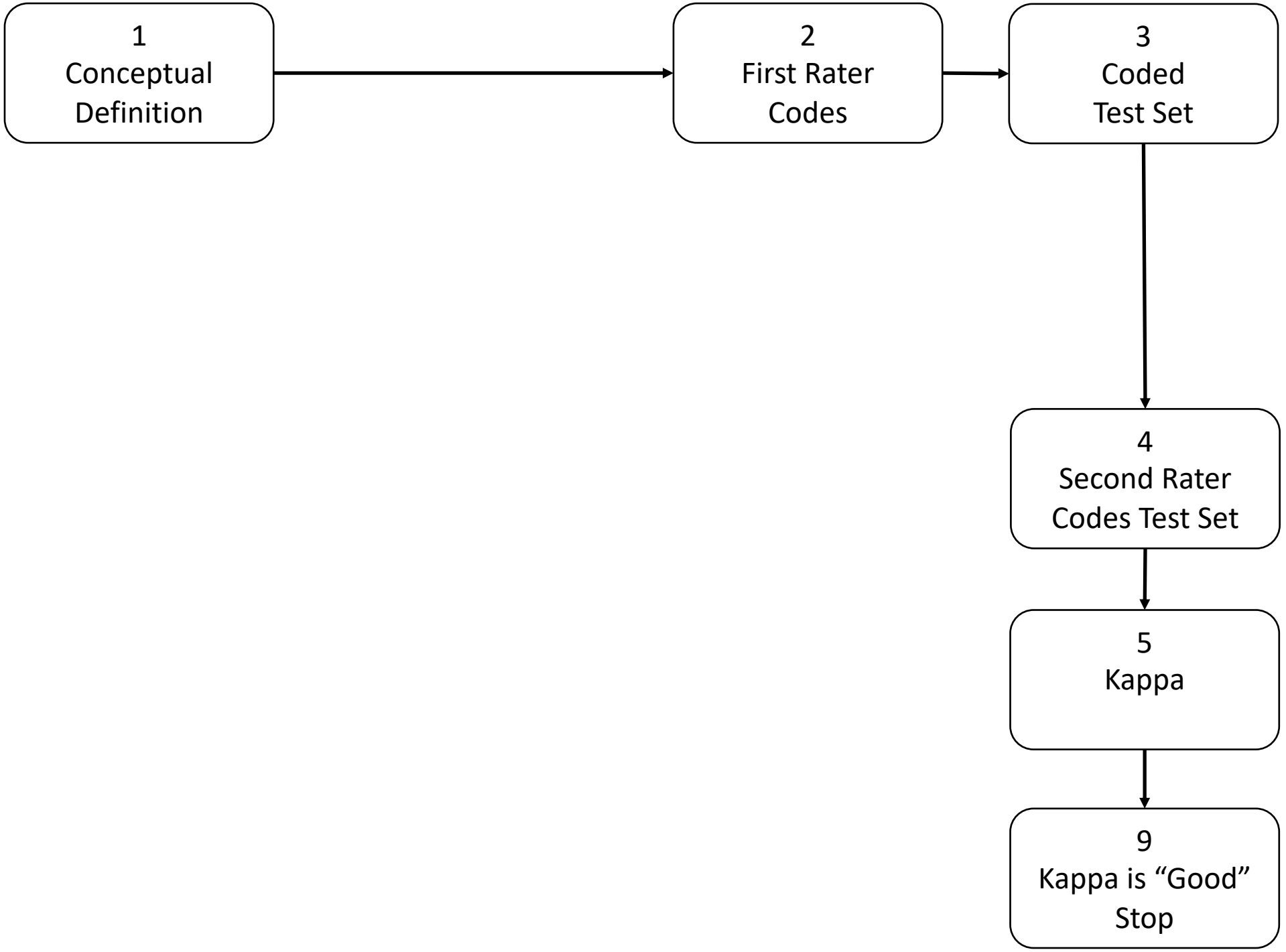
1
Conceptual
Definition

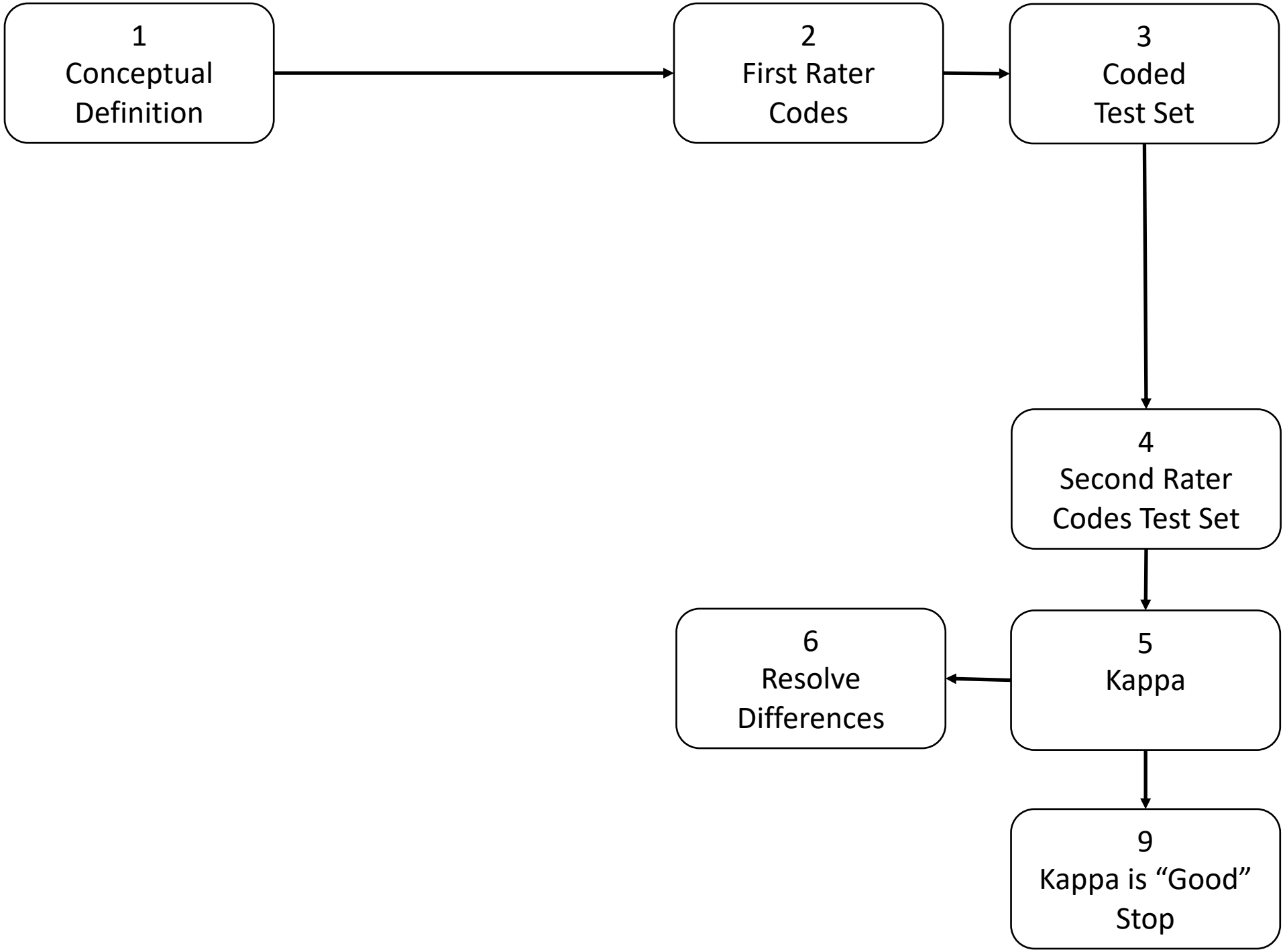

1
Conceptual
Definition
2
First Rater
Codes

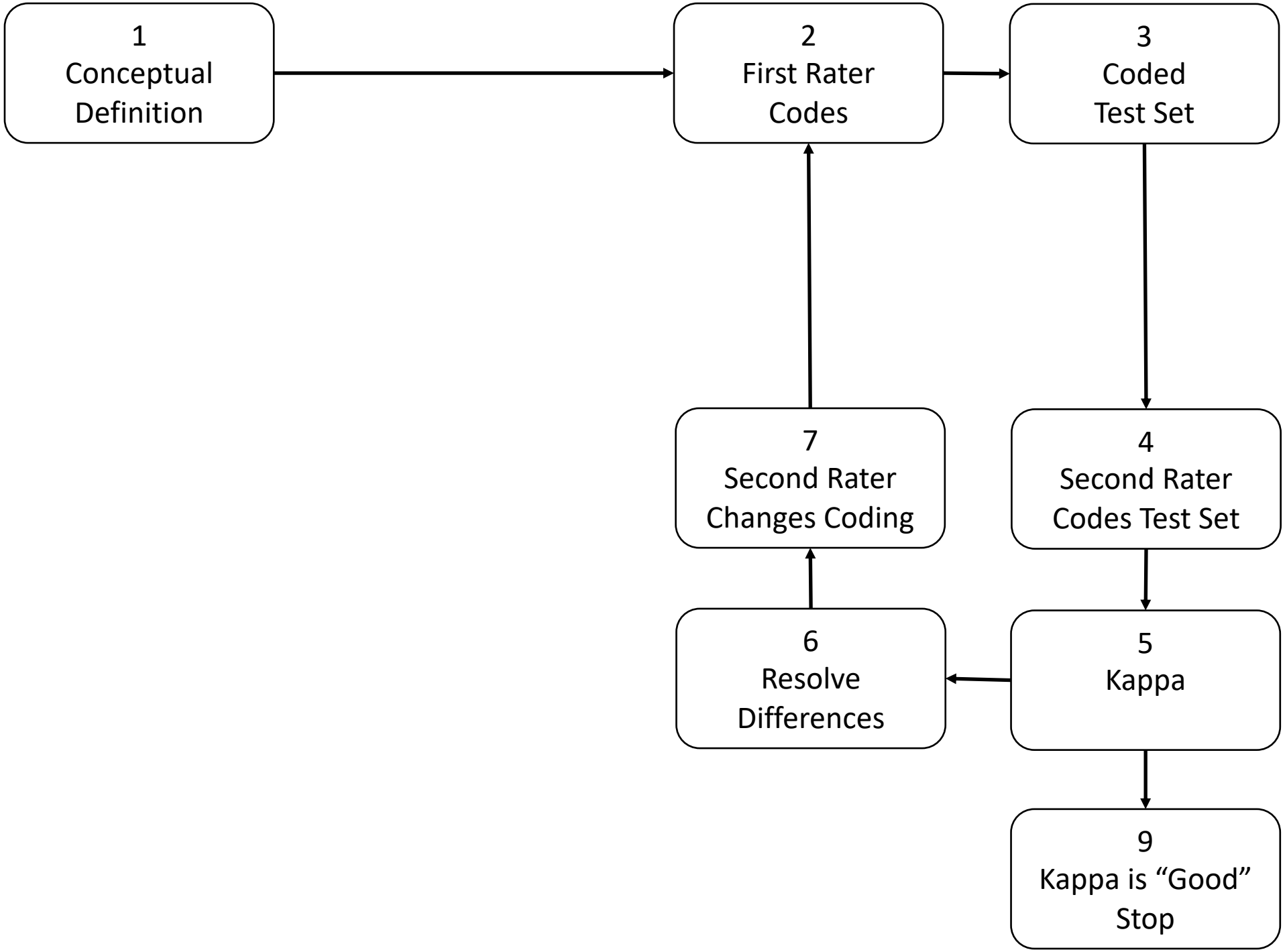

1
Conceptual
Definition
2
First Rater
Codes
3
Coded
Test Set

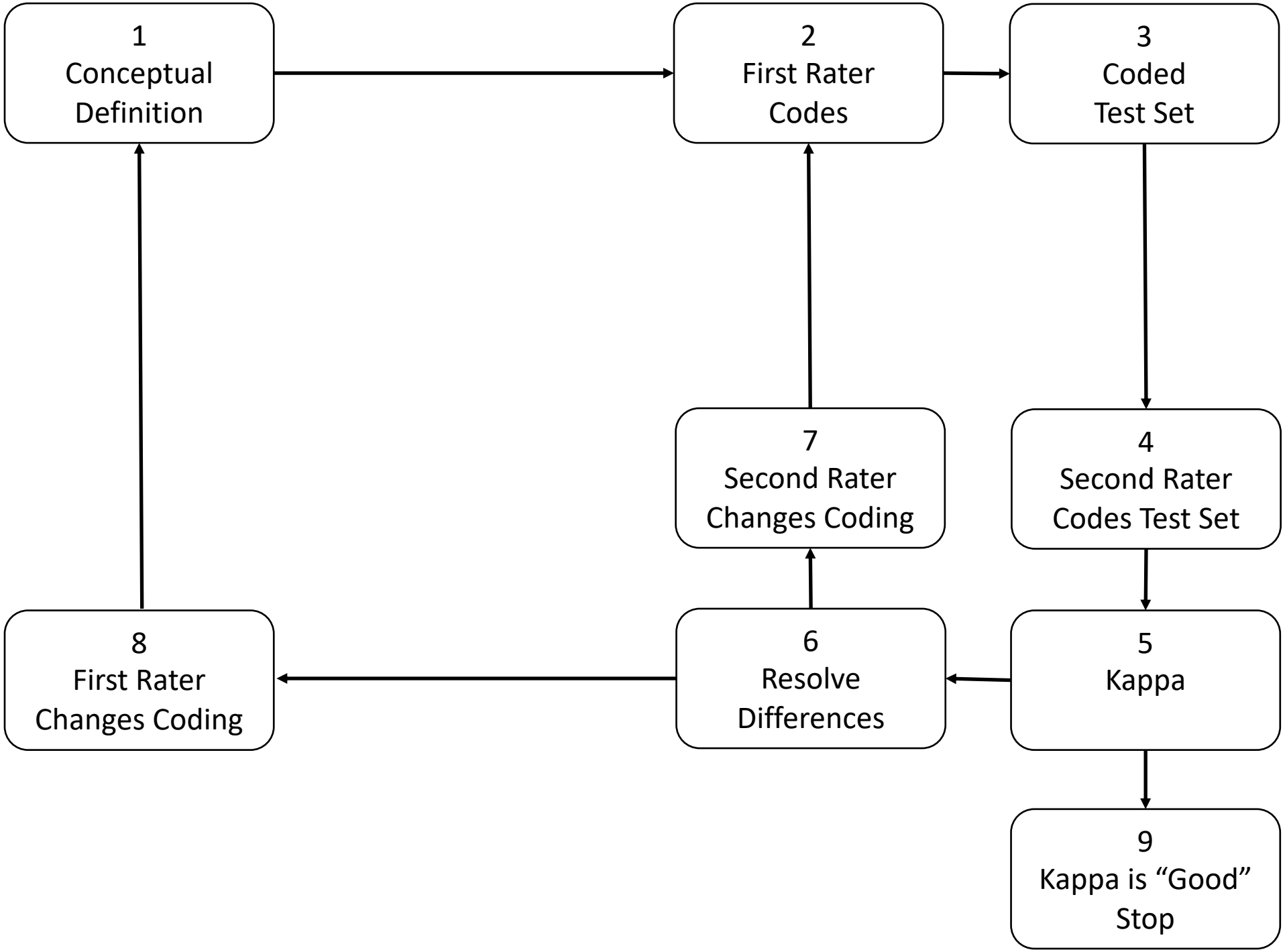

1
Conceptual
Definition
2
First Rater
Codes
3
Coded
Test Set
4
Second Rater
Codes Test Set

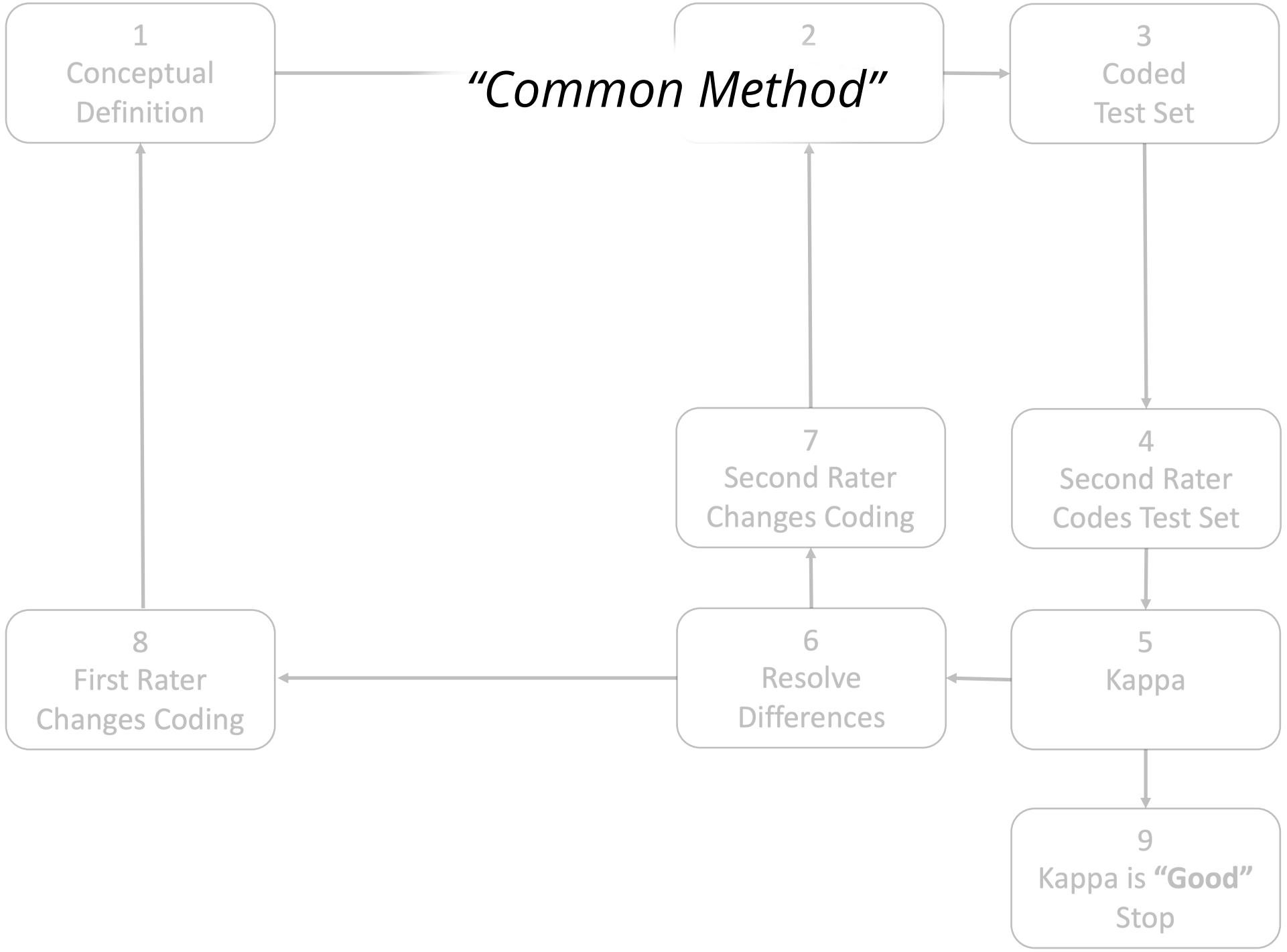

1
Conceptual
Definition
2
First Rater
Codes
3
Coded
Test Set
4
Second Rater
Codes Test Set
5
Kappa

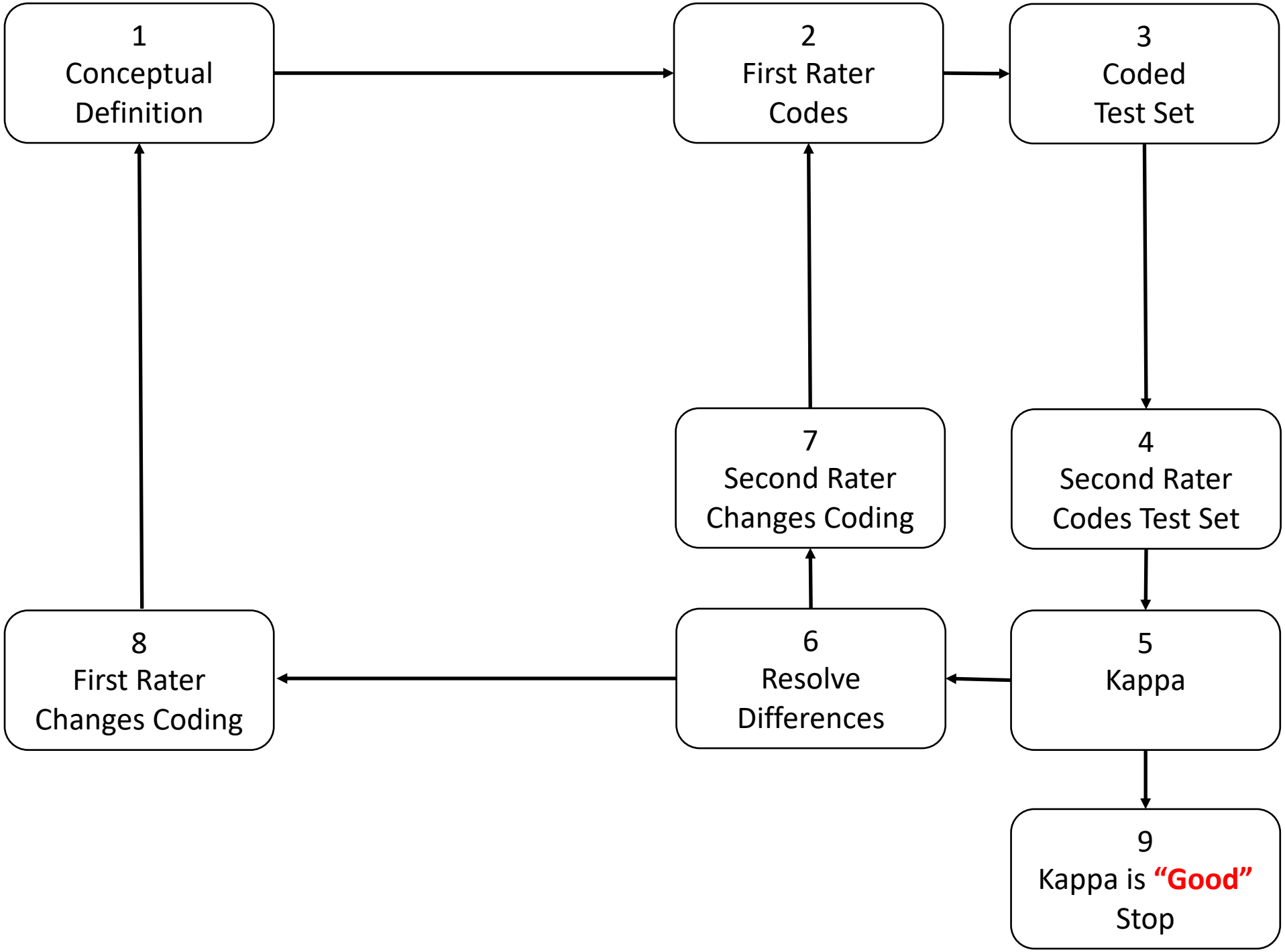

1
Conceptual
Definition
2
First Rater
Codes
3
Coded
Test Set
4
Second Rater
Codes Test Set
5
Kappa
9
Kappa is “Good”
Stop


1
Conceptual
Definition
2
First Rater
Codes
3
Coded
Test Set
4
Second Rater
Codes Test Set
5
Kappa
6
Resolve
Differences
9
Kappa is “Good”
Stop


1
Conceptual
Definition
2
First Rater
Codes
3
Coded
Test Set
4
Second Rater
Codes Test Set
5
Kappa
6
Resolve
Differences
7
Second Rater
Changes Coding
9
Kappa is “Good”
Stop


1
Conceptual
Definition
2
First Rater
Codes
3
Coded
Test Set
4
Second Rater
Codes Test Set
5
Kappa
6
Resolve
Differences
7
Second Rater
Changes Coding
8
First Rater
Changes Coding
9
Kappa is “Good”
Stop


1
Conceptual
Definition
2
3
Coded
Test Set
"Common Method"
4
Second Rater
Codes Test Set
5
Kappa
6
Resolve
Differences
7
Second Rater
Changes Coding
8
First Rater
Changes Coding
9
Kappa is "Good"
Stop


1
Conceptual
Definition
2
First Rater
Codes
3
Coded
Test Set
4
Second Rater
Codes Test Set
5
Kappa
6
Resolve
Differences
7
Second Rater
Changes Coding
8
First Rater
Changes Coding
9
Kappa is "Good"
Stop

Percent positive agreement (>70%)

	20	40	80	160	200	400	600	800	900	1000
0.01	0.723	0.638	0.517	0.339	0.284	0.167	0.124	0.0933	0.0925	0.0877
0.05	0.46	0.292	0.179	0.11	0.0867	0.0609	0.0491	0.0377	0.0382	0.0318
0.1	0.308	0.189	0.114	0.0684	0.0608	0.0471	0.0353	0.0274	0.0273	0.0239
0.2	0.194	0.129	0.0851	0.057	0.0512	0.0329	0.0256	0.0226	0.0221	0.0206
0.3	0.169	0.116	0.0782	0.0539	0.0464	0.0316	0.0272	0.023	0.0211	0.0214
0.5	0.183	0.144	0.0976	0.0658	0.0605	0.0448	0.0318	0.0311	0.0255	0.0232

Recall (>0.65)

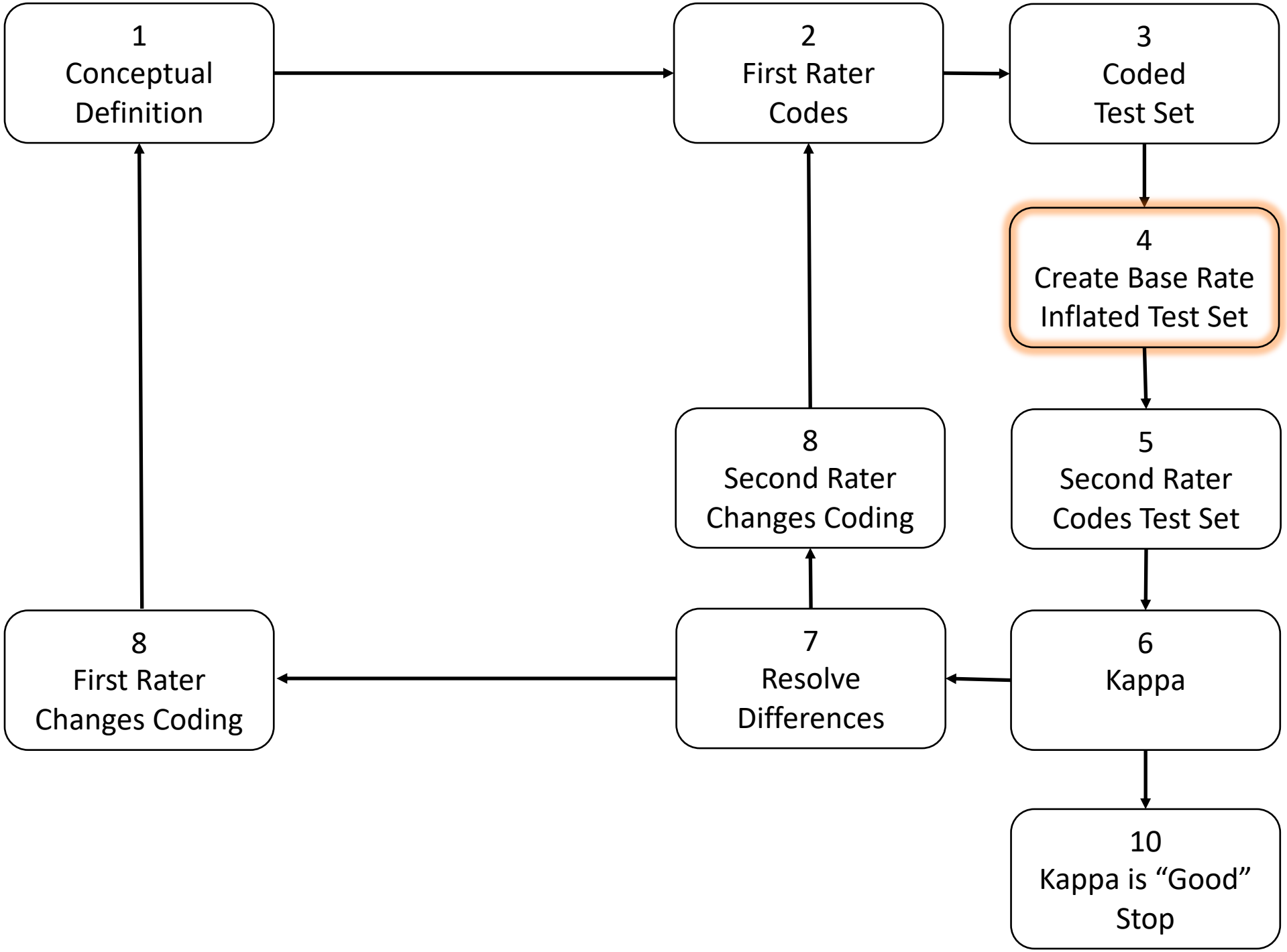
	20	40	80	160	200	400	600	800	900	1000
0.01	0.73	0.661	0.561	0.419	0.374	0.227	0.175	0.142	0.119	0.115
0.05	0.519	0.383	0.25	0.147	0.12	0.0734	0.0613	0.0549	0.0499	0.0441
0.1	0.396	0.271	0.15	0.0926	0.0788	0.0574	0.041	0.039	0.0354	0.0329
0.2	0.289	0.179	0.104	0.0721	0.0695	0.0428	0.0369	0.0293	0.0278	0.0268
0.3	0.228	0.141	0.101	0.0692	0.0624	0.0422	0.0348	0.0308	0.0302	0.0257
0.5	0.232	0.166	0.128	0.0882	0.0784	0.0536	0.0415	0.0374	0.0387	0.0328

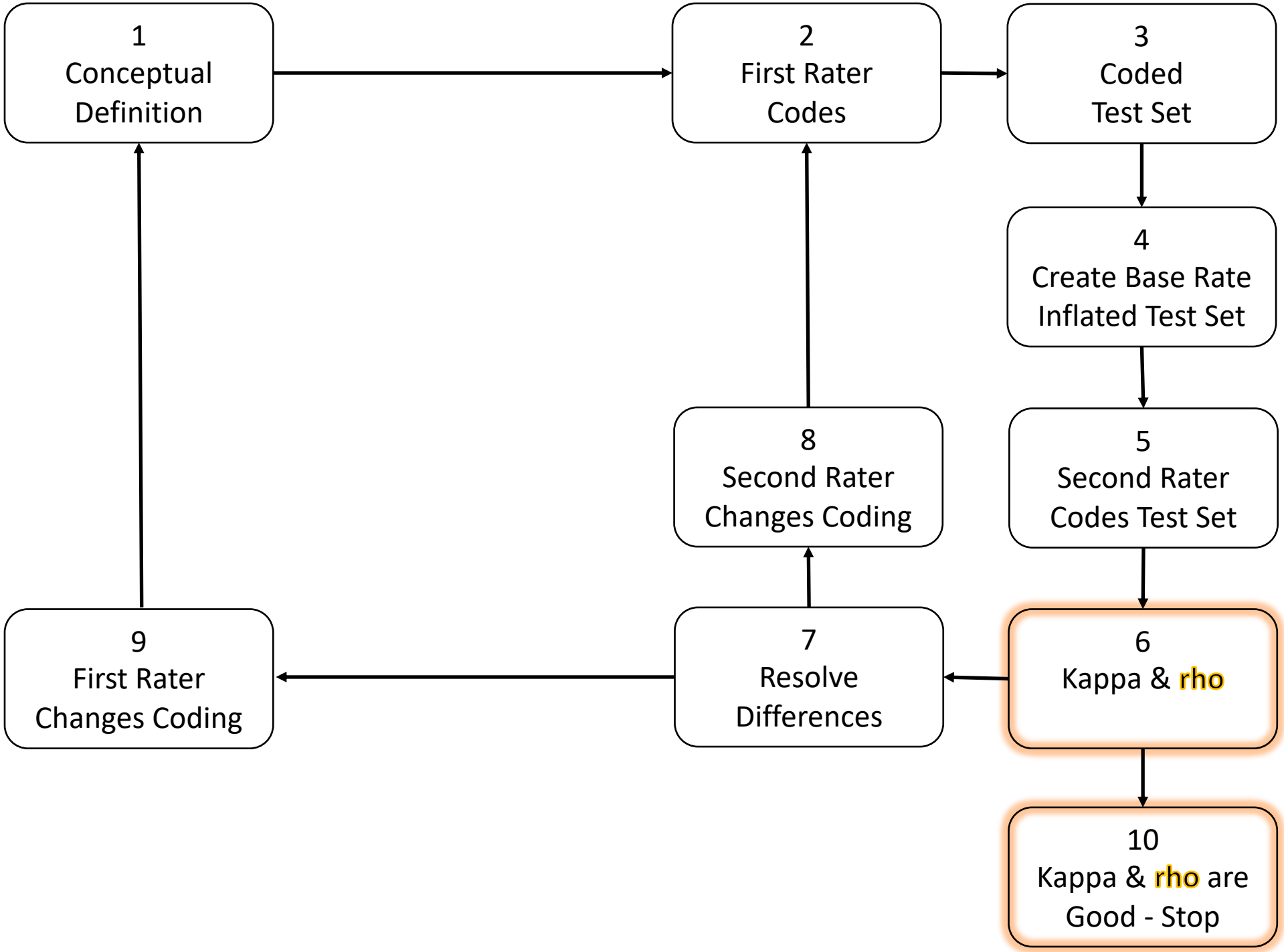
Precision (>0.65)

	20	40	80	160	200	400	600	800	900	1000
0.01	0.609	0.609	0.569	0.544	0.576	0.496	0.521	0.48	0.472	0.456
0.05	0.565	0.558	0.544	0.501	0.463	0.422	0.422	0.387	0.395	0.376
0.1	0.57	0.508	0.48	0.46	0.432	0.391	0.339	0.324	0.313	0.338
0.2	0.53	0.466	0.431	0.417	0.392	0.318	0.306	0.273	0.267	0.24
0.3	0.509	0.417	0.401	0.393	0.389	0.305	0.271	0.229	0.212	0.229
0.5	0.464	0.339	0.384	0.338	0.333	0.258	0.246	0.231	0.226	0.248

F statistic (>0.65)

	20	40	80	160	200	400	600	800	900	1000
0.01	0.8	0.789	0.75	0.611	0.563	0.362	0.263	0.215	0.196	0.18
0.05	0.722	0.578	0.377	0.219	0.195	0.12	0.0962	0.0846	0.0817	0.0799
0.1	0.581	0.372	0.229	0.142	0.126	0.0912	0.0741	0.0625	0.0587	0.0545
0.2	0.4	0.253	0.166	0.121	0.103	0.0736	0.0561	0.0501	0.0544	0.0466
0.3	0.339	0.227	0.158	0.11	0.114	0.0709	0.0585	0.0521	0.0466	0.0475
0.5	0.349	0.264	0.235	0.168	0.159	0.113	0.0841	0.0728	0.0684	0.0672


1
Conceptual
Definition
2
First Rater
Codes
3
Coded
Test Set
4
Create Base Rate
Inflated Test Set
5
Second Rater
Codes Test Set
6
Kappa
7
Resolve
Differences
8
Second Rater
Changes Coding
8
First Rater
Changes Coding
10
Kappa is “Good”
Stop


1
Conceptual
Definition
2
First Rater
Codes
3
Coded
Test Set
4
Create Base Rate
Inflated Test Set
5
Second Rater
Codes Test Set
6
Kappa & rho
7
Resolve
Differences
8
Second Rater
Changes Coding
9
First Rater
Changes Coding
10
Kappa & rho are
Good - Stop

Type I error rate rho, using kappa (threshold = 0.65) base rate inflation

	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90
0.01	0.028	0.015	0.02	0.028	0.018	0.031	0.025	0.03	0.029	0.028	0.031	0.027	0.031	0.032	0.032
0.05	0.024	0.027	0.034	0.029	0.034	0.031	0.035	0.034	0.035	0.032	0.035	0.032	0.035	0.034	0.037
0.1	0.03	0.033	0.033	0.035	0.033	0.035	0.037	0.035	0.035	0.035	0.034	0.037	0.035	0.034	0.038
0.2	0.032	0.035	0.036	0.034	0.036	0.037	0.035	0.036	0.036	0.036	0.037	0.036	0.035	0.035	0.037
0.3	0.032	0.035	0.036	0.035	0.036	0.034	0.035	0.036	0.037	0.034	0.033	0.036	0.035	0.038	0.037
0.5	0.033	0.035	0.035	0.035	0.035	0.036	0.032	0.037	0.035	0.035	0.036	0.035	0.034	0.037	0.037

Type I error rate rho, using kappa (threshold = 0.65) base rate inflation

	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90
0.01	0.028	0.015	0.02	0.028	0.018	0.031	0.025	0.03	0.029	0.028	0.031	0.027	0.031	0.032	0.032
0.05	0.024	0.027	0.034	0.029	0.034	0.031	0.035	0.034	0.035	0.032	0.035	0.032	0.035	0.034	0.037
0.1	0.03	0.033	0.033	0.035	0.033	0.035	0.037	0.035	0.035	0.035	0.034	0.037	0.035	0.034	0.038
0.2	0.032	0.035	0.036	0.034	0.036	0.037	0.035	0.036	0.036	0.036	0.037	0.036	0.035	0.035	0.037
0.3	0.032	0.035	0.036	0.035	0.036	0.034	0.035	0.036	0.037	0.034	0.033	0.036	0.035	0.038	0.037
0.5	0.033	0.035	0.035	0.035	0.035	0.036	0.032	0.037	0.035	0.035	0.036	0.035	0.034	0.037	0.037

Type II error rate rho, using kappa (threshold = 0.65) base rate inflation

	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90
0.01	0.495	0.54	0.481	0.376	0.41	0.314	0.321	0.284	0.268	0.267	0.234	0.24	0.215	0.217	0.198
0.05	0.525	0.459	0.389	0.372	0.33	0.307	0.291	0.271	0.261	0.247	0.234	0.224	0.206	0.206	0.198
0.1	0.488	0.435	0.385	0.346	0.321	0.296	0.273	0.257	0.246	0.237	0.224	0.21	0.202	0.192	0.188
0.2	0.481	0.402	0.363	0.339	0.311	0.282	0.263	0.249	0.239	0.227	0.212	0.2	0.195	0.187	0.181
0.3	0.471	0.409	0.368	0.335	0.304	0.282	0.266	0.255	0.238	0.23	0.217	0.213	0.201	0.19	0.19
0.5	0.483	0.412	0.365	0.337	0.312	0.289	0.269	0.252	0.238	0.223	0.212	0.205	0.2	0.188	0.184

Type I error rate rho, using kappa (threshold = 0.65) base rate inflation

	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90
0.01	0.028	0.015	0.02	0.028	0.018	0.031	0.025	0.03	0.029	0.028	0.031	0.027	0.031	0.032	0.032
0.05	0.024	0.027	0.034	0.029	0.034	0.031	0.035	0.034	0.035	0.032	0.035	0.032	0.035	0.034	0.037
0.1	0.03	0.033	0.033	0.035	0.033	0.035	0.037	0.035	0.035	0.035	0.034	0.037	0.035	0.034	0.038
0.2	0.032	0.035	0.036	0.034	0.036	0.037	0.035	0.036	0.036	0.036	0.037	0.036	0.035	0.035	0.037
0.3	0.032	0.035	0.036	0.035	0.036	0.034	0.035	0.036	0.037	0.034	0.033	0.036	0.035	0.038	0.037
0.5	0.033	0.035	0.035	0.035	0.035	0.036	0.032	0.037	0.035	0.035	0.036	0.035	0.034	0.037	0.037

Type II error rate rho, using kappa (threshold = 0.65) base rate inflation

	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90
0.01	0.495	0.54	0.481	0.376	0.41	0.314	0.321	0.284	0.268	0.267	0.234	0.24	0.215	0.217	0.198
0.05	0.525	0.459	0.389	0.372	0.33	0.307	0.291	0.271	0.261	0.247	0.234	0.224	0.206	0.206	0.198
0.1	0.488	0.435	0.385	0.346	0.321	0.296	0.273	0.257	0.246	0.237	0.224	0.21	0.202	0.192	0.188
0.2	0.481	0.402	0.363	0.339	0.311	0.282	0.263	0.249	0.239	0.227	0.212	0.2	0.195	0.187	0.181
0.3	0.471	0.409	0.368	0.335	0.304	0.282	0.266	0.255	0.238	0.23	0.217	0.213	0.201	0.19	0.19
0.5	0.483	0.412	0.365	0.337	0.312	0.289	0.269	0.252	0.238	0.223	0.212	0.205	0.2	0.188	0.184

Type I error rate rho, using kappa (threshold = 0.9) base rate inflation

	20	40	80	160	200	400	600	800
0.01	0	0	0.0421	0.0381	0.037	0.0392	0.0461	0.0432
0.05	0	0.00655	0.0394	0.0415	0.0409	0.0428	0.0389	0.044
0.1	0	0.0356	0.0453	0.0441	0.0441	0.0446	0.0451	0.0436
0.2	0	0.0367	0.044	0.0417	0.0425	0.0432	0.0411	0.0444
0.3	0.000468	0.0398	0.0447	0.0403	0.0391	0.0464	0.0418	0.0417
0.5	0.00446	0.0425	0.0401	0.0417	0.0438	0.038	0.0429	0.0397

Type II error rate rho, using kappa (threshold = 0.9) base rate inflation

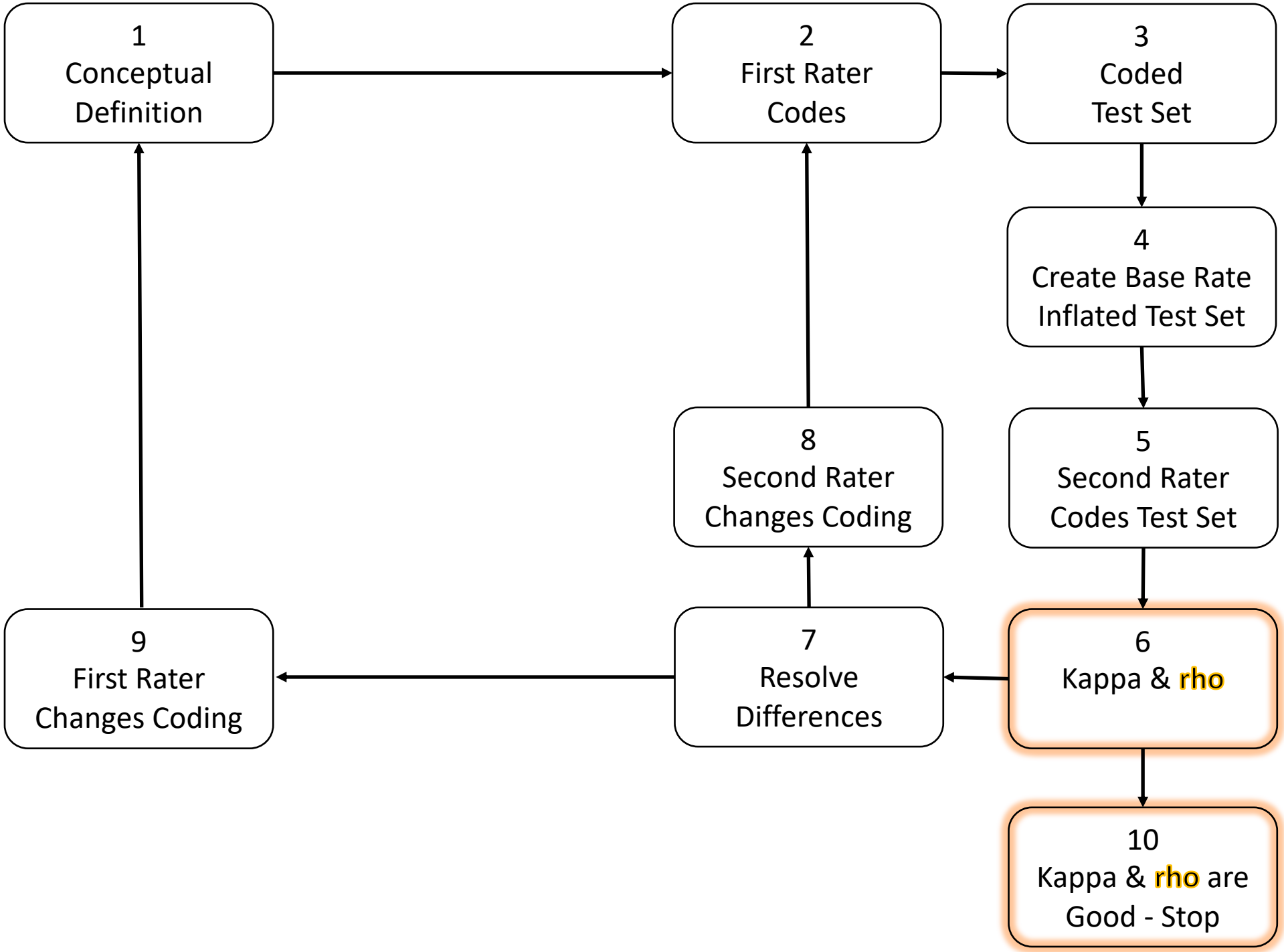
	20	40	80	160	200	400	600	800
0.01	1	1	0.479	0.324	0.323	0.262	0.251	0.204
0.05	1	0.899	0.345	0.224	0.204	0.16	0.163	0.123
0.1	1	0.439	0.277	0.171	0.129	0.0914	0.0855	0.0571
0.2	1	0.404	0.213	0.101	0.0865	0.0405	0.0336	0.0322
0.3	0.993	0.332	0.181	0.075	0.0766	0.0272	0.0209	0.0137
0.5	0.969	0.309	0.14	0.0727	0.0524	0.0253	0.00729	0.013

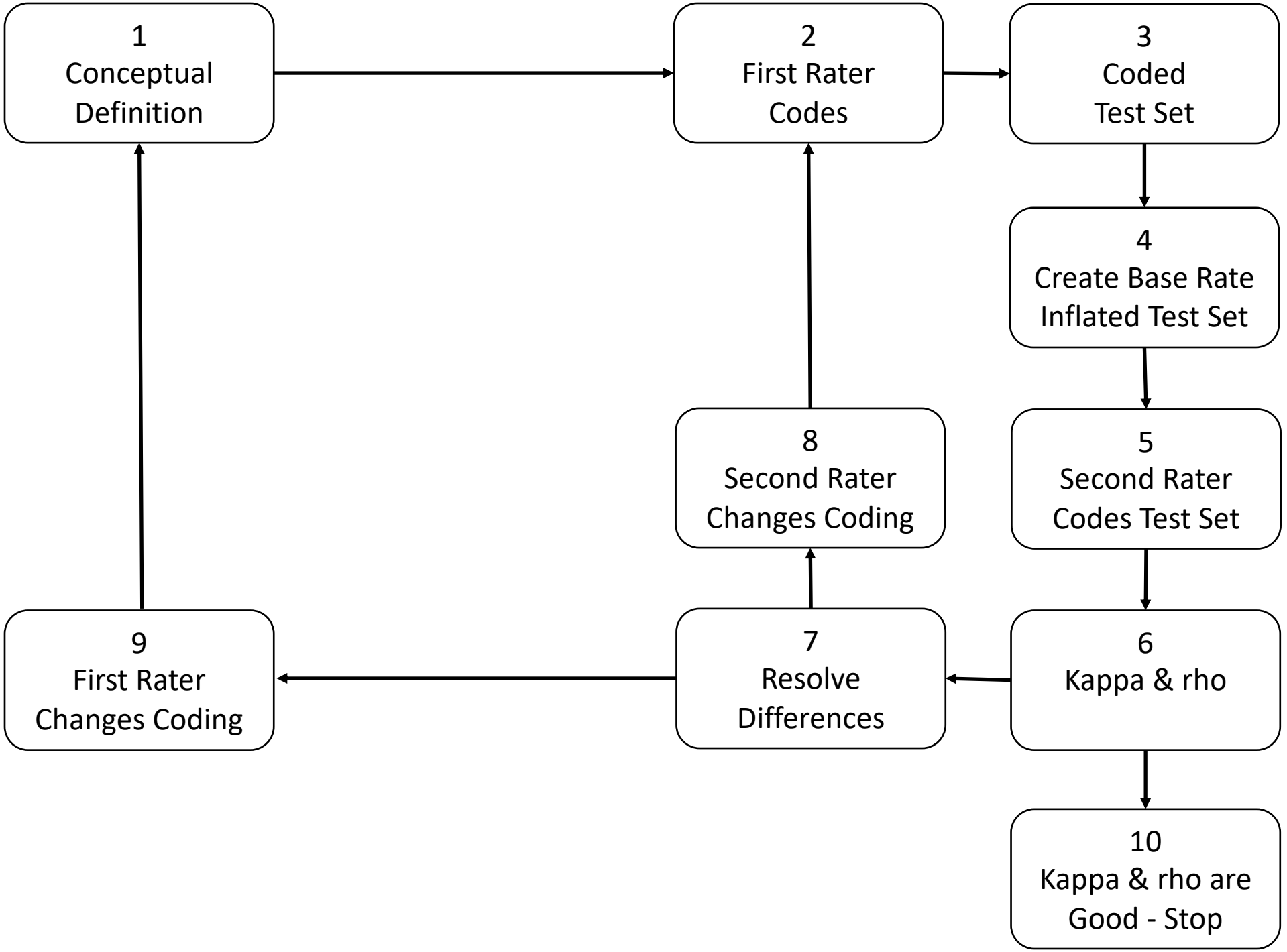
Type I error rate rho, using kappa (threshold = 0.9) base rate inflation

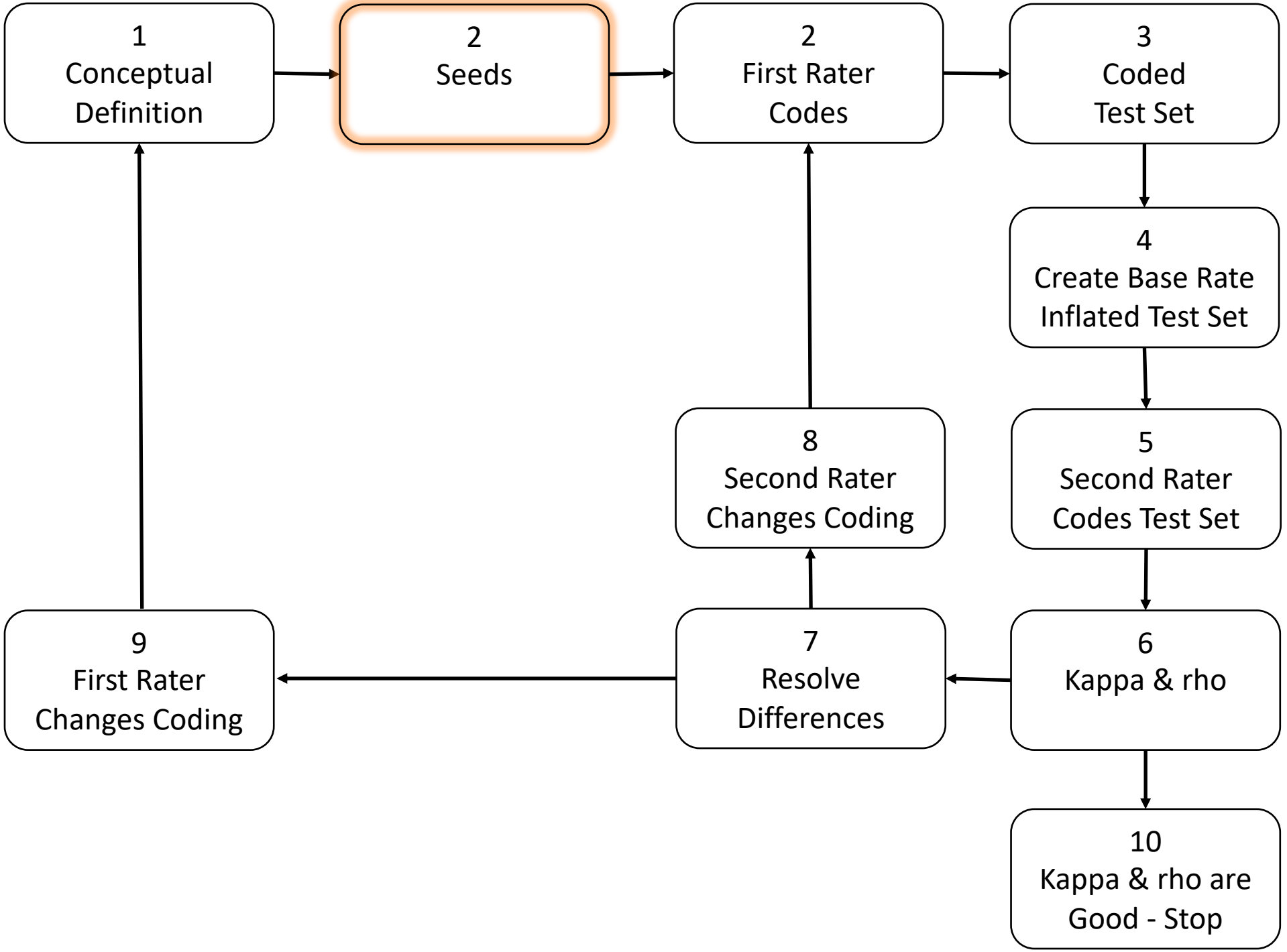
	20	40	80	160	200	400	600	800
0.01	0	0	0.0421	0.0381	0.037	0.0392	0.0461	0.0432
0.05	0	0.00655	0.0394	0.0415	0.0409	0.0428	0.0389	0.044
0.1	0	0.0356	0.0453	0.0441	0.0441	0.0446	0.0451	0.0436
0.2	0	0.0367	0.044	0.0417	0.0425	0.0432	0.0411	0.0444
0.3	0.000468	0.0398	0.0447	0.0403	0.0391	0.0464	0.0418	0.0417
0.5	0.00446	0.0425	0.0401	0.0417	0.0438	0.038	0.0429	0.0397

Type II error rate rho, using kappa (threshold = 0.9) base rate inflation

	20	40	80	160	200	400	600	800
0.01	1	1	0.479	0.324	0.323	0.262	0.251	0.204
0.05	1	0.899	0.345	0.224	0.204	0.16	0.163	0.123
0.1	1	0.439	0.277	0.171	0.129	0.0914	0.0855	0.0571
0.2	1	0.404	0.213	0.101	0.0865	0.0405	0.0336	0.0322
0.3	0.993	0.332	0.181	0.075	0.0766	0.0272	0.0209	0.0137
0.5	0.969	0.309	0.14	0.0727	0.0524	0.0253	0.00729	0.013


1
Conceptual
Definition
2
First Rater
Codes
3
Coded
Test Set
4
Create Base Rate
Inflated Test Set
5
Second Rater
Codes Test Set
6
Kappa & rho
7
Resolve
Differences
8
Second Rater
Changes Coding
9
First Rater
Changes Coding
10
Kappa & rho are
Good - Stop


1
Conceptual
Definition
2
First Rater
Codes
3
Coded
Test Set
4
Create Base Rate
Inflated Test Set
5
Second Rater
Codes Test Set
6
Kappa & rho
7
Resolve
Differences
8
Second Rater
Changes Coding
9
First Rater
Changes Coding
10
Kappa & rho are
Good - Stop


1
Conceptual
Definition
2
Seeds
2
First Rater
Codes
3
Coded
Test Set
4
Create Base Rate
Inflated Test Set
5
Second Rater
Codes Test Set
6
Kappa & rho
7
Resolve
Differences
8
Second Rater
Changes Coding
9
First Rater
Changes Coding
10
Kappa & rho are
Good - Stop

Imagine that you have a special instrument that allows you to see what makes up odor.

The large circle in the drawing represents a spot that is magnified many times, so you can see it up close.

Create a model of what you would see if you could focus on one tiny spot in the area between the jar and your nose.

What is this about?

Imagine that you have a special instrument that allows you to see what makes up odor.
The large circle in the drawing represents a spot that is magnified many times, so you can see it up close.
Create a model of what you would see if you could focus on one tiny spot in the area between the jar and your nose.

What is this about?

Segmentation: how you divide your data into meaningful parts

Imagine that you have a special instrument that allows you to see what makes up odor.
The large circle in the drawing represents a spot that is magnified many times, so you can see it up close.
Create a model of what you would see if you could focus on one tiny spot in the area between the jar and your nose.

What is this about?

Segmentation: how you divide your data into meaningful parts

How have you segmented your data?

Imagine that you have a special instrument that allows you to see what makes up odor.
The large circle in the drawing represents a spot that is magnified many times, so you can see it up close.
Create a model of what you would see if you could focus on one tiny spot in the area between the jar and your nose.

What is this about?

Modeling:

Imagine that you have a special instrument that allows you to see what makes up odor.
The large circle in the drawing represents a spot that is magnified many times, so you can see it up close.
Create a model of what you would see if you could focus on one tiny spot in the area between the jar and your nose.

What is this about?

Modeling:

Draw
Represent
Circle

Imagine that you have a special instrument that allows you to see what makes up odor.
The large circle in the drawing represents a spot that is magnified many times, so you can see it up close.
Create a model of what you would see if you could focus on one tiny spot in the area between the jar and your nose.

What is this about?

Modeling:

Draw
Represent
Circle
Model

Imagine that you have a special instrument that allows you to see what makes up odor.
The large circle in the drawing represents a spot that is magnified many times, so you can see it up close.
Create a model of what you would see if you could focus on one tiny spot in the area between the jar and your nose.

What is this about?

Modeling:

Draw
Represent
Circle
Model

Imagine – create a mental representation? hypothesize? thought experiment?

THIS COMIC MADE POSSIBLE THANKS TO ADAM LINGELBACH

MRLOVENSTEIN.COM

The large circle in the drawing represents a spot that is magnified many times, so you can see it up close.
Let’s sit in a circle.

Modeling:

Circle

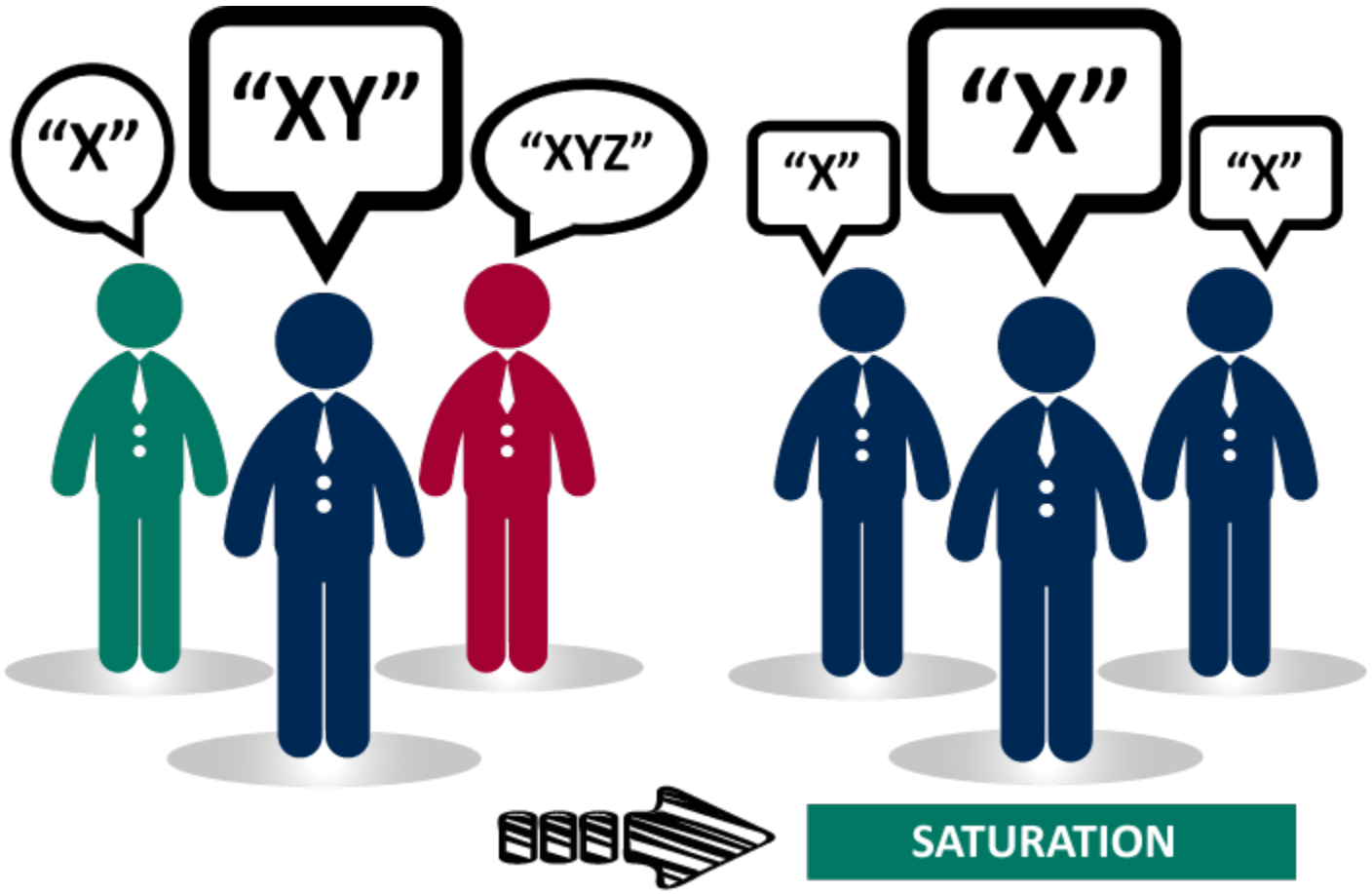
5-minute break

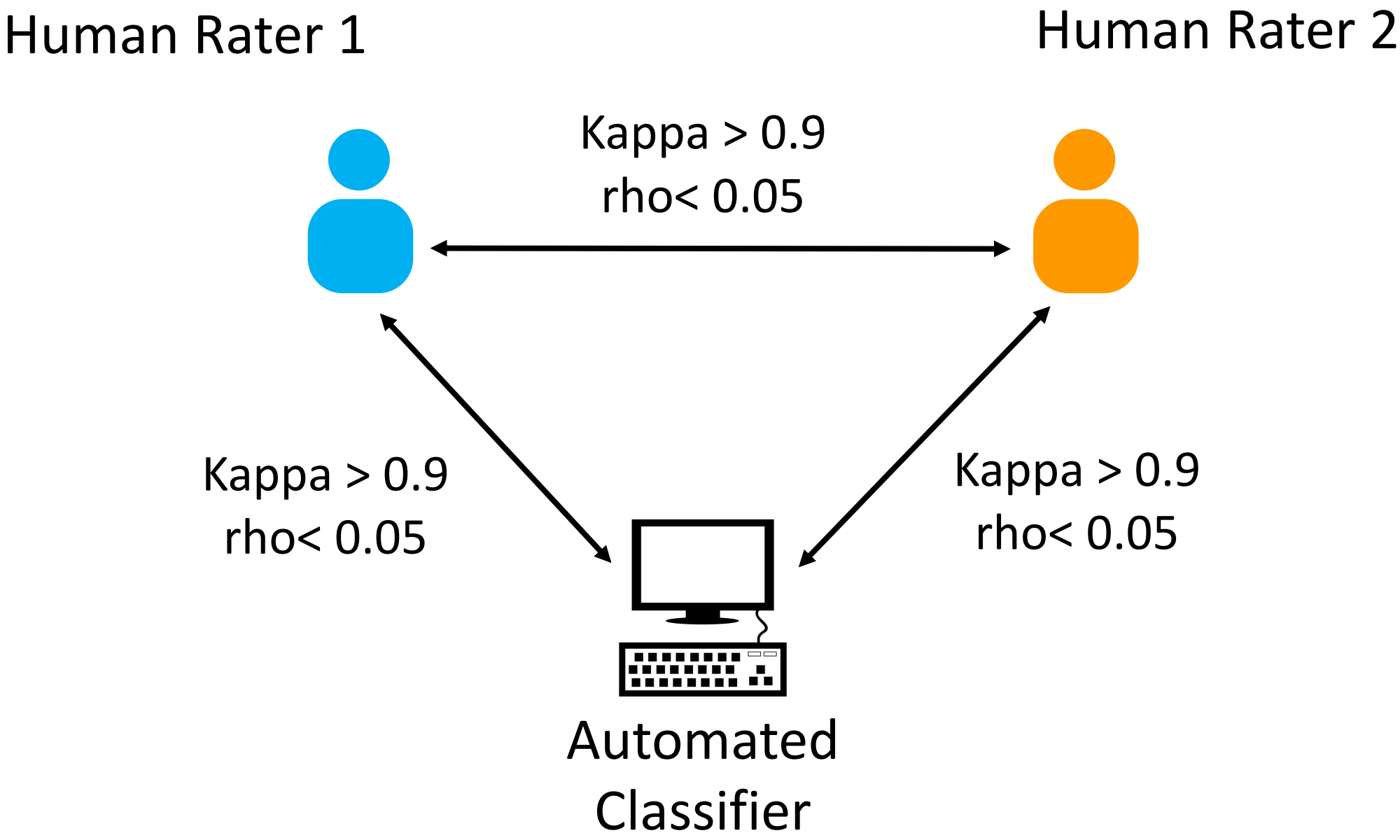
Hydration, snack, chatting, etc.

https://app.n-coder.org/

https://go.wisc.edu/0nxl70

When have I coded enough?




Human Rater 1
Human Rater 2
Kappa > 0.9
rho< 0.05
Kappa > 0.9
rho< 0.05
Kappa > 0.9
rho< 0.05
Automated
Classifier

ICQE20

Workshop 2A: Introduction to nCoder

An Applied Example

Seung B. Lee
Pepperdine University

Introduction

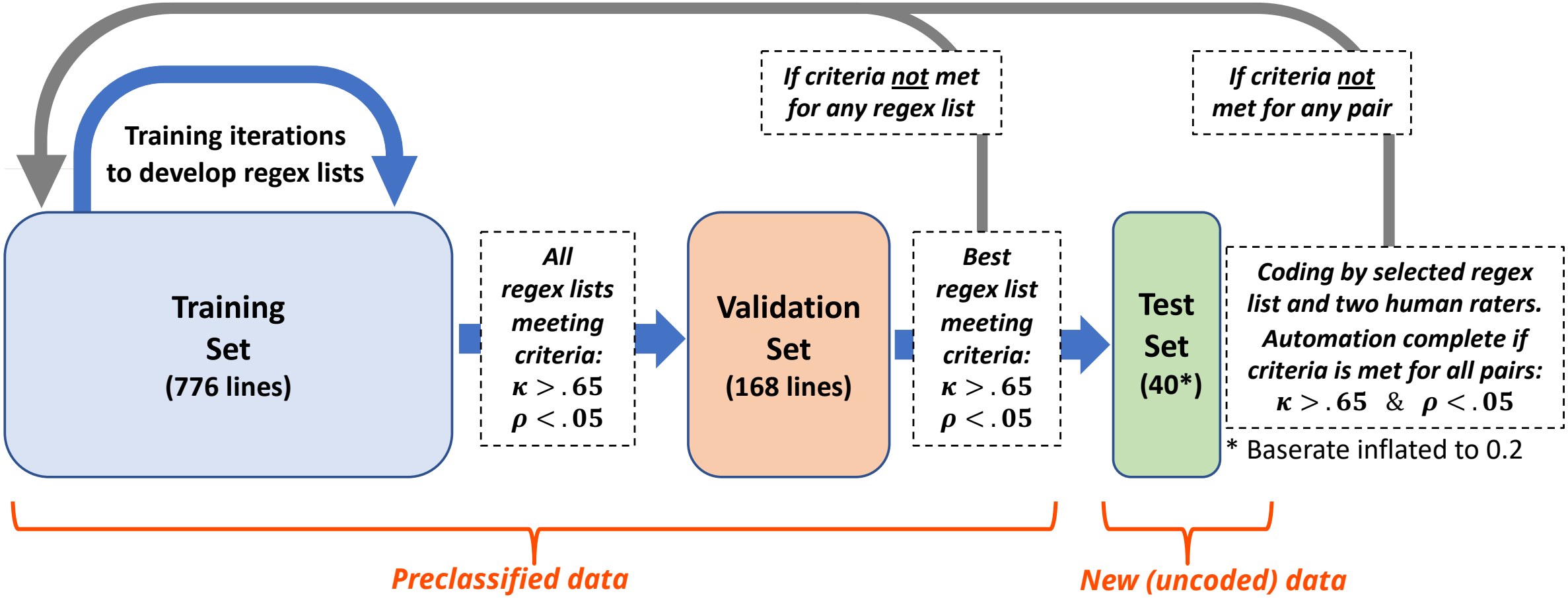
The International Community for Collaborative Content Creation (IC4)

- Examines **cross-boundary online collaboration** of students in informal STEM learning
- Students aged 12-17 from **six countries** across four continents develop digital media artifacts (videos, presentations) about STEM subjects
- Communicate **synchronously** through online global meet-ups (video conference calls) and **asynchronously** (email, Slack)

Data & Analysis

- The largest portion of our data consists of meet-up transcripts – which we have mainly analyzed through **hand coding** (20+ constructs)
- But… there is **a lot of data** (on avg. 1-3 meet-ups each week, each about 1 hr)
- Our approach to automated coding:
  - Utilize data that have already been coded (preclassified data)
  - Pilot test on two codes: social disposition, inclusive disposition

Process


Training iterations to develop regex lists
If criteria not met for any regex list
If criteria not met for any pair
Training Set (776 lines)
All regex lists meeting criteria: $\kappa > .65$ $\rho < .05$
Validation Set (168 lines)
Best regex list meeting criteria: $\kappa > .65$ $\rho < .05$
Test Set (40*)
Coding by selected regex list and two human raters. Automation complete if criteria is met for all pairs: $\kappa > .65$ & $\rho < .05$
* Baserate inflated to 0.2
Preclassified data
New (uncoded) data

Results (after one cycle)

- Training stage
  - Inclusive disposition: 7th iteration (kappa = 0.72; rho = 0.0038)
  - Social disposition: 4th iteration (kappa = 0.82; rho = 0)
- Validation stage
  - Inclusive disposition: kappa = 0.41; rho = 1 (Overfitting)
  - Social disposition: kappa = 0.73; rho = 0.02
- Testing stage
  - Social disposition (40 lines, baserate-inflated to 0.2)
    - Computer vs. Human 1: kappa = 0.81; rho = 0.04
    - Computer vs. Human 2: kappa = 0.88; rho = 0.01
    - Human 1 vs. Human 2: kappa = 0.94; rho = 0.001

Use of Preclassified Data in Automated Coding

Advantages

- Greater familiarity of the data
  - Completed initial process of identifying, defining and applying the codes
  - Common understanding between human raters: Concept validity
- Can increase efficiency in certain contexts
  - When coded data already exists
  - Researchers can refine regular expressions without additional coding
    - Helpful for deriving regex lists for more complex codes
  - If multiple codes used, training set can be used for automating multiple codes

Use of Preclassified Data in Automated Coding

Limitations

- Need to consider size of full data set
  - Not efficient for data sets that are smaller
- Possibility of regex list overfitting the training set
- Question of whether some codes can ultimately be automated
  - When should the iterative loop be stopped? Need to consider complexity of individual codes

Thank you

Questions?

Email us at:
info@ic4.site
seung.lee@pepperdine.edu

This project, supporting students who collaborate in digital makerspaces in five continents, is funded by the Advancing Informal STEM Learning (AISL) program of the US National Science Foundation (NSF) Award #1612824.

nCoder

https://app.n-coder.org/

Amanda Siebert-Evenstone
alevenstone@wisc.edu

Seung B. Lee
seung.lee@pepperdine.edu

Brendan Eagan
beagan@wisc.edu