AN ANALYSIS OF THE REGULARITY OF "SIGHT WORDS"

Katharine Pace Miles, Ph.D.

Brooklyn College, CUNY

Devin Kearns, Ph.D.

University of Connecticut

Terms

- ■ Sight words
- ■ Irregular words
- ■ High frequency words
- ■ Vocabulary words

Ways Words are Read

- **By Decoding**
  - *D O G --> /d/ /a/ /g/ --> "dog"*
  - *CH E CK (5 letters, 3 graphemes) - > /č/ /ɛ/ /k/ - > "check"*
  - *EXCELLENT -> /ex/ /cel/ /lent/ (syllables)*
  - *UPHOLDING -> /up/ /hold/ /ing/ (prefix, root, suffix)*

- **By Analogy**
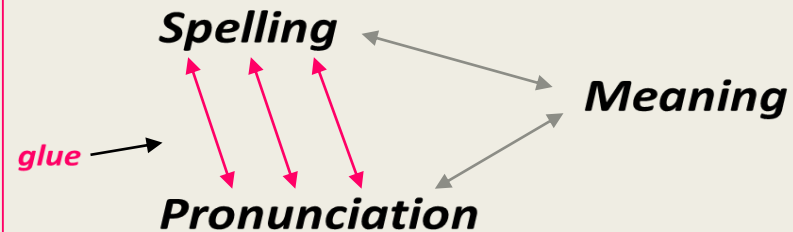 - *j ump → d ump*

- Prediction: context clues and partial

- **By Memory/Sight**

How Words are Stored as Sight Words

- Process of forming connections or mappings

Spelling

Meaning

glue →

Pronunciation

**Knowledge of the grapheme-phoneme
system provides the *glue* connecting
spellings to pronunciations in memory**

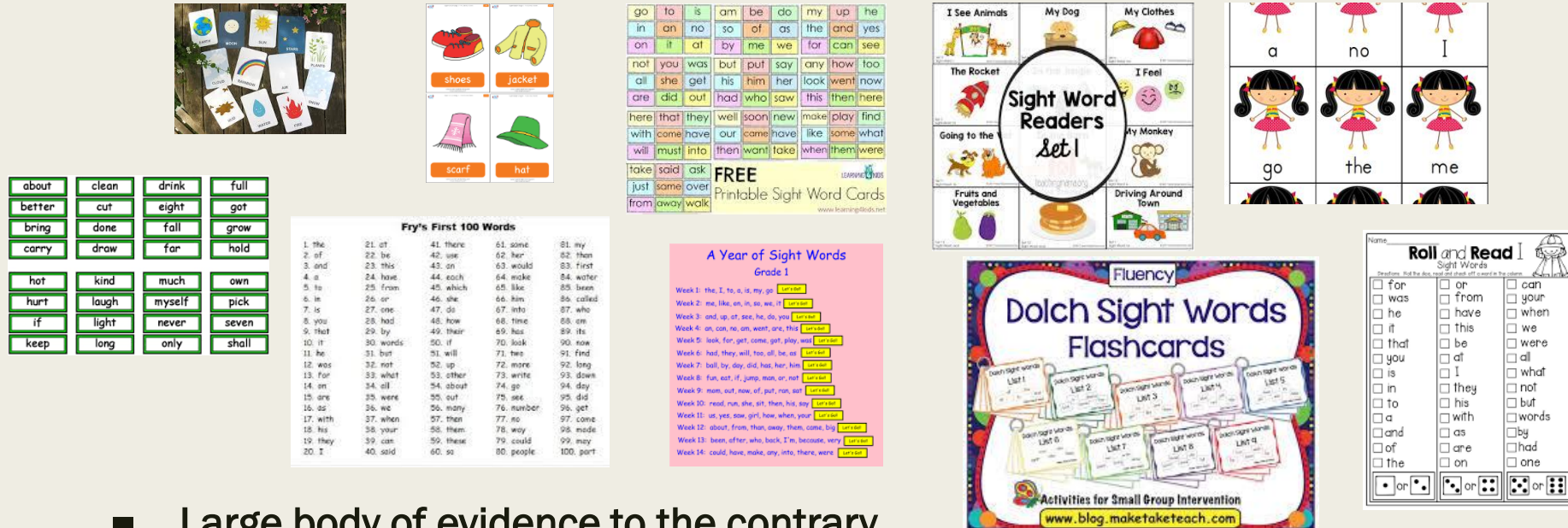# Contrast this with "Sight Word" Learning

■ Research

■ Practice

Disconnect between Research and Practice

■ Large body of evidence to the contrary

– *Arra & Aaron, 2001; Boyer & Ehri, 2011; Castles, Rastle, & Nation, 2018; Ehri, Satlow, & Gaskins, 2009; Miles, Rubin, Gonzalez-Frey; 2017; Ouellette & Senechal, 2008; Shahar-Yames & Share, 2008; Stuart, Masterson, and Dixon, 2000; Uhry & Shepherd, 1997;*

Regularity of English Spellings

- **English spellings are more regular than often perceived**
    - Carreker, 2011; Ehri, 1997; Joshi, Treiman, Carreker, & Moats, 2005, 2008, 2009; Miles, Rubin, Gonzalez-Frey, 2017; Trieman & Kessler, 2013

- **Teacher linguistic knowledge**
    - Moats, 1994, 2002, 2009a, 2009b, 2011;

    Moats & Foorman, 2003; Puliatte & Ehri, 2017

    Spear-Swerling, 2010;

Types of Words on "Sight Word" Lists

- **Regularly Spelled**
  - *Follow most common g-p relations*

- **Temporarily Irregularly Spelled**
  - *G-p relations/patterns students have not yet learned*
  - *Once learned can apply to multiple words that contain the spelling pattern*

- **Permanently Irregularly Spelled**
  - *G-p relations are idiosyncratic to that word or only a few others*
  - *Violations of typical g-p relations or spelling rules*
  - *Silent letters*
  - *Although often times, several letters in the word may still map onto reliable g-p relations*

Teacher Survey

- What is your definition of the term *sight words*?

- What methods of instruction do you use to teach sight words?

the
of
and
a
to
in
is
you
that
it
he
was
for
on
are
as
with
his
they
I
at
be
this
have
from

or
one
had
by
word
but
not
what
all
were
we
when
your
can
said
there
use
an
each
which
she
do
how
their
if

Teacher Survey Results

- Definition of Sight Words
  - *69% of teachers said these words should not/cannot be decoded*

- Methods used to teach sight words
  - *7% said analyze letter-sounds*

- When given letter-sound as an option
  - *Only 45% checked the box*

- Accuracy in categorizing words
  - *Average accuracy score of 67% (range 48-82%)*
  - *The highest frequency words intended for K and 1st grade*

Survey Take-Aways

- Comprehensive execution of linguistic knowledge

- Teachers need linguistic training
    - *their, from*

- Curriculum needs to be responsible for this linguistic knowledge

- **The difficulty in doing this!**

A "Computational" Approach to Examining Word Regularities

- One kind of computational model
  - *A simulation of human behavior using a program*
  - *The simulation is conducted by feeding the computer words one at a time and seeing it how performs*
- Another kind of computation
  - *Analyzing the characteristics of words themselves*
  - *Learning about the nature of the words without setting up the rules ahead of time*

A Roughly Computational Approach to Reading a Word

world

The "Computational" Model Here

- Analysis of the regularity of sound-spelling patterns in English words
  - *uses information about words' letters and sounds*
  - *makes no assumptions about patterns are regular*
  - *uses a set of iterative computations to determine whether words are regular*
- Militates against bias in decision-making about what to teach

world   *OR = /ɝ/*

Is *OR = /ɝ/* a regular pattern?
Is *world* a regular word?

Analysis 1: Kindergarten

- Expert coding—What do experts think are regular words for Kindergarteners?

- Program coding—What does the computational approach think are regular words for Kindergarteners?

Expert Analysis of K Word List

- Purpose
  - *determine how many of the words on commonly used sight word/high frequency lists have reliable grapheme-phoneme relations*

- Materials
  - *Combined list of words from both Dolch and Fry*
  - *Moats 44 Phoneme-Grapheme Chart (most frequent spellings of phonemes)*
  - *CCSS, ELA, Foundational Skills, K, 3. Know and apply grade level phonics and word analysis skills in decoding words*
    - (A) Demonstrate basic knowledge of one-to-one letter-sound correspondences by producing the *primary sound or many of the most frequent sounds for each consonant*
    - (B) Associate the *long and short sounds* with common spellings (graphemes) for *the five major vowels*

Expert Analysis of K Word List Coding Rules and Examples

K Regular	K Temp Irr	K Perm Irr
Rule	Rule	Rule
Single Letter	GPC on Moats List	GPC not on Moats list
GPC on Moats list	GPC with 4+ occurrences	
Examples	Examples	Examples
Short /a/ spelled a	Short /e/ spelled ea	Short /u/ spelled o_e
Short /e/ spelled e	Short /i/ spelled y	Short /e/ spelled ai
Short /i/ spelled i	Short /o/ spelled wa, al	Silent w
Short /o/ spelled o	Short /u/ spelled o, oo, ou	
Short /u/ spelled u	/ā/ spelled a_e, ai, ay, ea, -y, eigh, ei, ey	
/ā/ spelled a	/ē/ spelled ee, e_e, ea, ey, -y, ie, ei	
/ē/ spelled e	/ī/ spelled i_e, ie, -y, igh,	
/ī/ spelled i	/ō/ spelled o_e, oa, oe, ow	
/ō/ spelled o		

Expert Analysis of K Word List
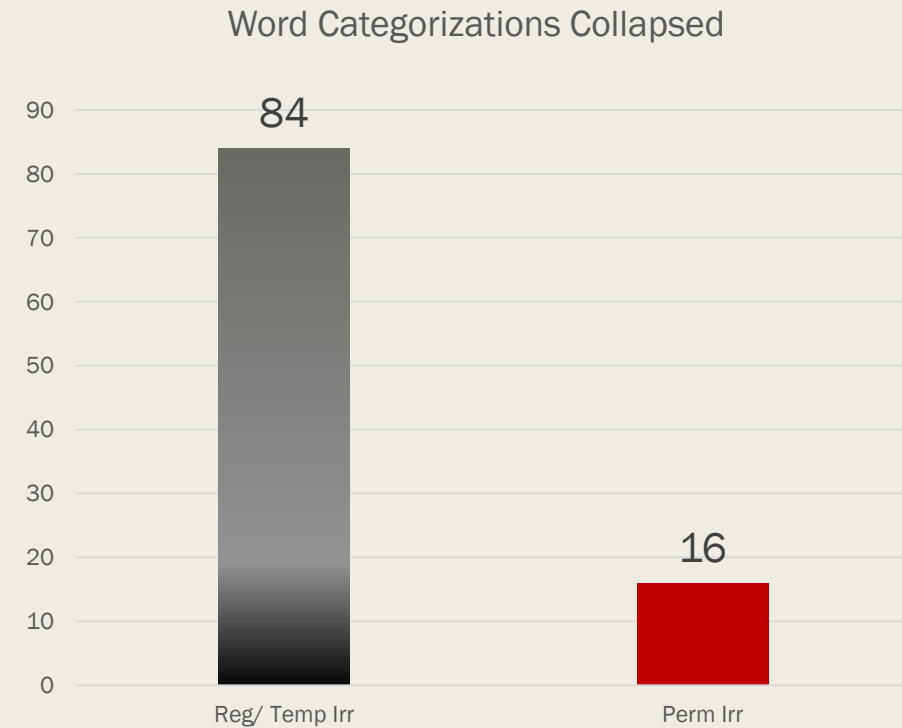
- Procedures
  - *Coded all words on Dolch and Fry lists (N = 419)*
    - Regularly Spelled (1)
    - Temporarily Irregularly Spelled (2)
    - Permanently Irregularly Spelled (3)
  - *Followed rules previously described*
- Coding
  - *3 Stage Coding Process*
    - 1/3 of the list coded together for training; establish rules
    - 1/3 coded independently; resolved issues; refined rules
    - 1/3 coded independently; resolved issues
  - *Cohen's kappa: 0.90, resolved issues*

Expert Analysis of K Word List

Word Categorizations

```
70
60                        63
50
40
30
    21                              16
20
10
 0
    Reg        Temp Irr      Perm Irr
```

Word Categorizations Collapsed

```
90
80          84
70
60
50
40
30
20                        16
10
 0
  Reg/ Temp Irr        Perm Irr
```
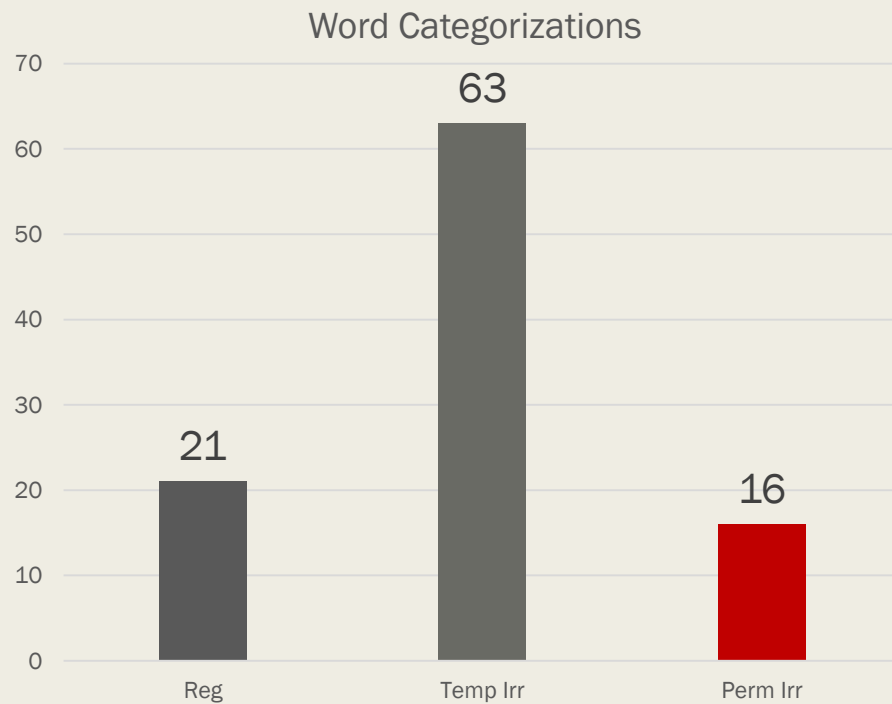
- 353 Words Regularly or Temporarily Irregularly
- 66 Words Permanently Irregularly Spelled

Program Analysis of K Word List

- Purpose
  - *use iterative computations to explore regularity*
  - *replicate or improve upon the expert coding*
- Materials
  - *Same list of Dolch and Fry words*
  - *Words coded by GPC*
  - *Program parameters*

Words coded by GPC:
Letter String and Pronunciation Database

directly	d@rEkt5i
directness	d@rEktn@s
director	d@rEkt@Xr
director's	d@rEkt@Xrz
directorate	d@rEkt@Xr@t
directors	d@rEkt@Xrz
directorship	d@rEkt@XrSIp
directory	d@rEkt@Xri
directs	d@rEkts
direful	daIrfl=
dirge	d3XrdZ
dirt	d3Xrt
dirty	d3Xr4i
disabilities	dIs@bI5@4iz
disability	dIs@bI5@4i
disable	dIsebl=
disabled	dIsebl=d
disabling	dIsebl=IN
disabuse	dIs@bjuz

- List of words from Fitt (2001)
  - *Source for English Lexicon Project (Balota et al., 2007)*
  - *American English spellings*
  - *General American pronunciations*
  - *List of GPCs*
  - *List of words coded grapheme-by-grapheme*

Words coded by GPC:
List of Possible GPCs

```
g = dZ g   Z
e = @  e  E  i
t = t  tS
```

e = /@/ → given

e = /e/ → *mesa, Rodeo*

e = /E/ → *get*
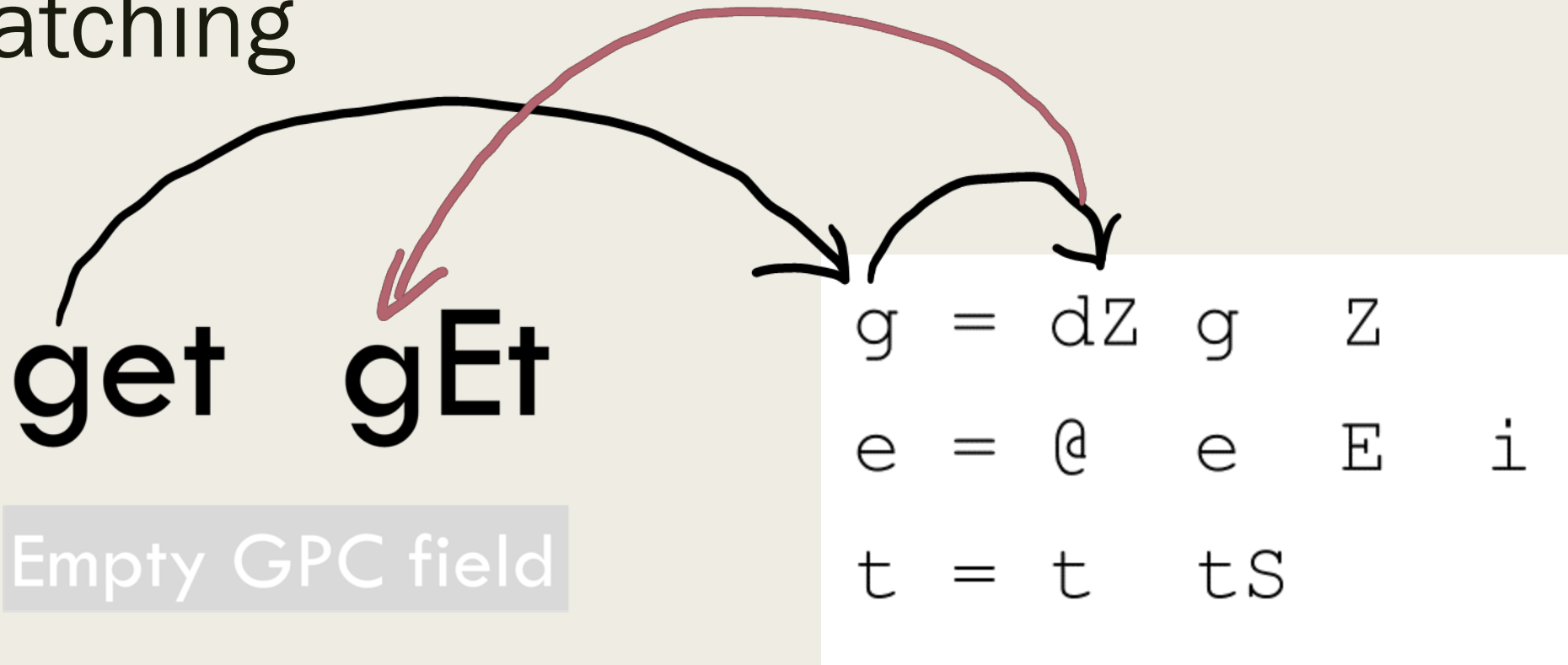
e = /i/ → *she*

Words coded by GPC: Matching

get gEt

Empty GPC field

g = dZ g Z

e = @ e E i

t = t tS

Letter String and Pronunciation Database

List of Possible GPCs

Words coded by GPC: Matching

get gEt

Empty GPC field

g = dZ g Z

e = @ e E i

t = t tS

Letter String and
Pronunciation Database

List of Possible GPCs

Words coded by GPC: Matching

get   gEt

g = g

g	=	dZ	g	Z
e	=	@	e	E	i
t	=	t	tS

Letter String and Pronunciation Database

List of Possible GPCs

Data available: https://phinder.devinkearns.org

Program parameters

- **Regular** word category
  - *Only one-letter graphemes were considered*
  - *Vowel letters each had 2 pronunciations*
  - *GPCs with schwa were not permitted*

- **Temporarily irregular** word category
  - *Multiple levels of **consistency***
  - *Multiple levels of **frequency***

- **Permanently irregular** words
  - *Anything that did not fit after applying regular and temp. irr. rules*

Temporarily Irregular Word Category: Consistency

■ In what percentage of words is this grapheme pronounced with this phoneme?

Graph.	Phon.	Cons. (%)	Freq.
e	/ē/	5.6	217
e	/ĕ/	36.0	1388
e	/-/	10.3	397
e	/ĭ/	5.6	217
e	/ŏ/	< 0.1	1
e	/ā/	0.3	13
e	/ə/	42.1	1626

E = /ĕ/ in 35% of words with the grapheme *E*

Temporarily Irregular Word Category: Frequency

- In how many words is this grapheme pronounced with this phoneme?

- Word database for frequency counts
  - *based on the Educator's Word Frequency Guide (Zeno et al., 1995)*
  - *includes words that occur in EWFG Grade 1, 2, or 3 data*
  - N = *12,080*

Graph.	Phon.	Cons. (%)	Freq.
e	/ē/	5.6	217
e	/ĕ/	36.0	1388
e	/-/	10.3	397
e	/ĭ/	5.6	217
e	/ŏ/	< 0.1	1
e	/ā/	0.3	13
e	/ə/	42.1	1626

E = /ĕ/ in 1,388 words

Multiple Levels of Frequency and Consistency

Consistency Level	Frequency Level
0%	0
0%	10
0%	20 … to 100
1%	0
1%	10
1%	20 … to 100
2% … to 9%	0
2% … to 9%	10
2% … to 9%	20 … to 100

Level	Interpretation
Consistency = 0% Frequency = 0 words	Any GPC occurring once or more
Consistency = 1% Frequency = 0 words	A GPC where grapheme has this sound at least 1% of the time
Consistency = 0% Frequency = 10 words	Any GPC occurring more than 10 times
Consistency = 1% Frequency = 10 words	A GPC where grapheme has this sound at least 1% of the time and that occurs 10 or more times

Regularity calculation for each word

■ Regular and Temporarily Irregular GPCs meet the frequency and consistency requirements for a given analysis

Word Category	Criteria to Code a Word to This Category
Regular	
	All GPCs are Regular
Temporarily Irregular	
	All GPCs are Temporarily Irregular Some GPCs are Temp.Irr. and some are Regular
Irregular	
	At least one GPC is Irregular

Program Analysis: Calculating

Key Question

- What is the total number of sight words that can be read with these sound-spellings?

- How many sound-spellings need to be taught to maximize accuracy?

Expert Comparison

- 109 GPCs created

- 353 words decoded

Consistency and Frequency Cut-Points
for Determining Regularity of High-Frequency Words (Kindergarten-Level) vs. Expert Coding

Words Considered Regular

GPCs Required ... eve This Regularity

135 GPCs required
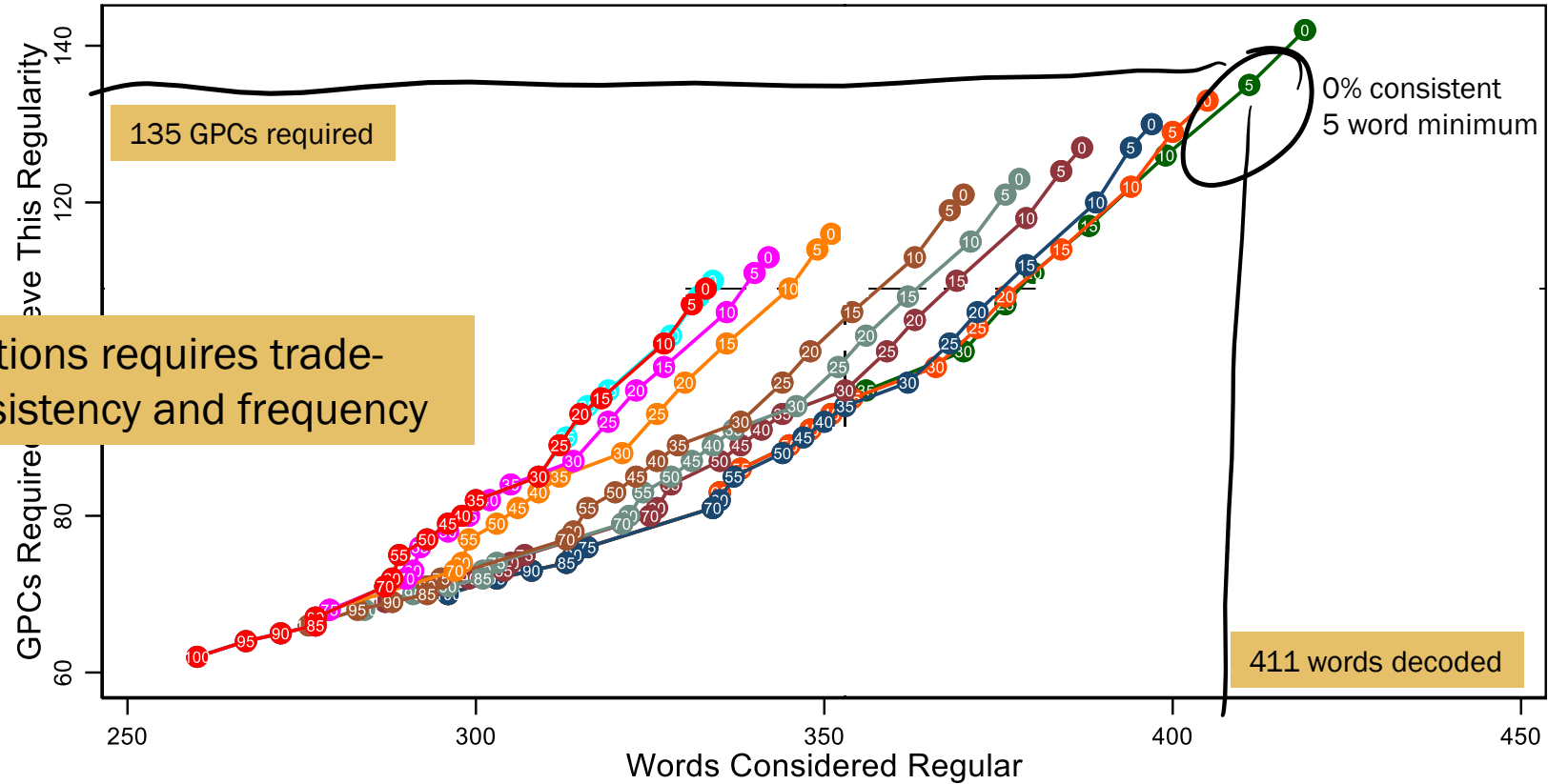
Regularity calculations requires trade-offs between consistency and frequency
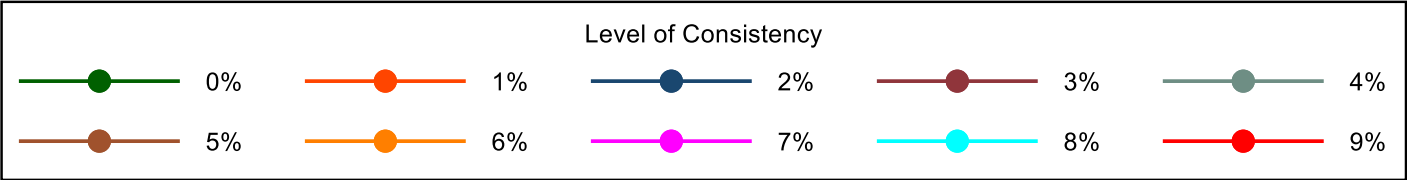
0% consistent
5 word minimum

411 words decoded

Level of Consistency

- 0%
- 1%
- 2%
- 3%
- 4%
- 5%
- 6%
- 7%
- 8%
- 9%

Consistency and Frequency Cut-Points

for Determining Regularity of High-Frequency Words (Kindergarten-Level) vs. Expert Coding

Expert GPCs required = 109
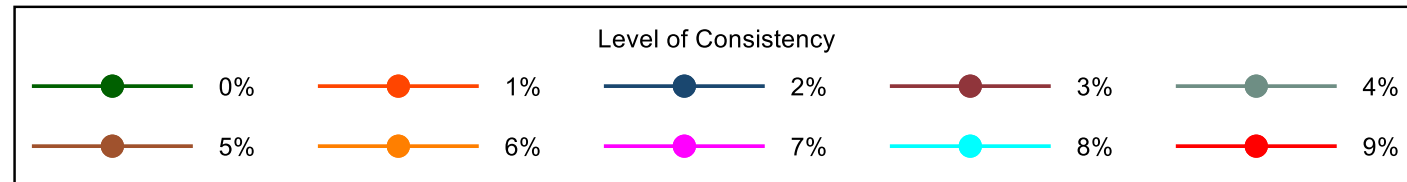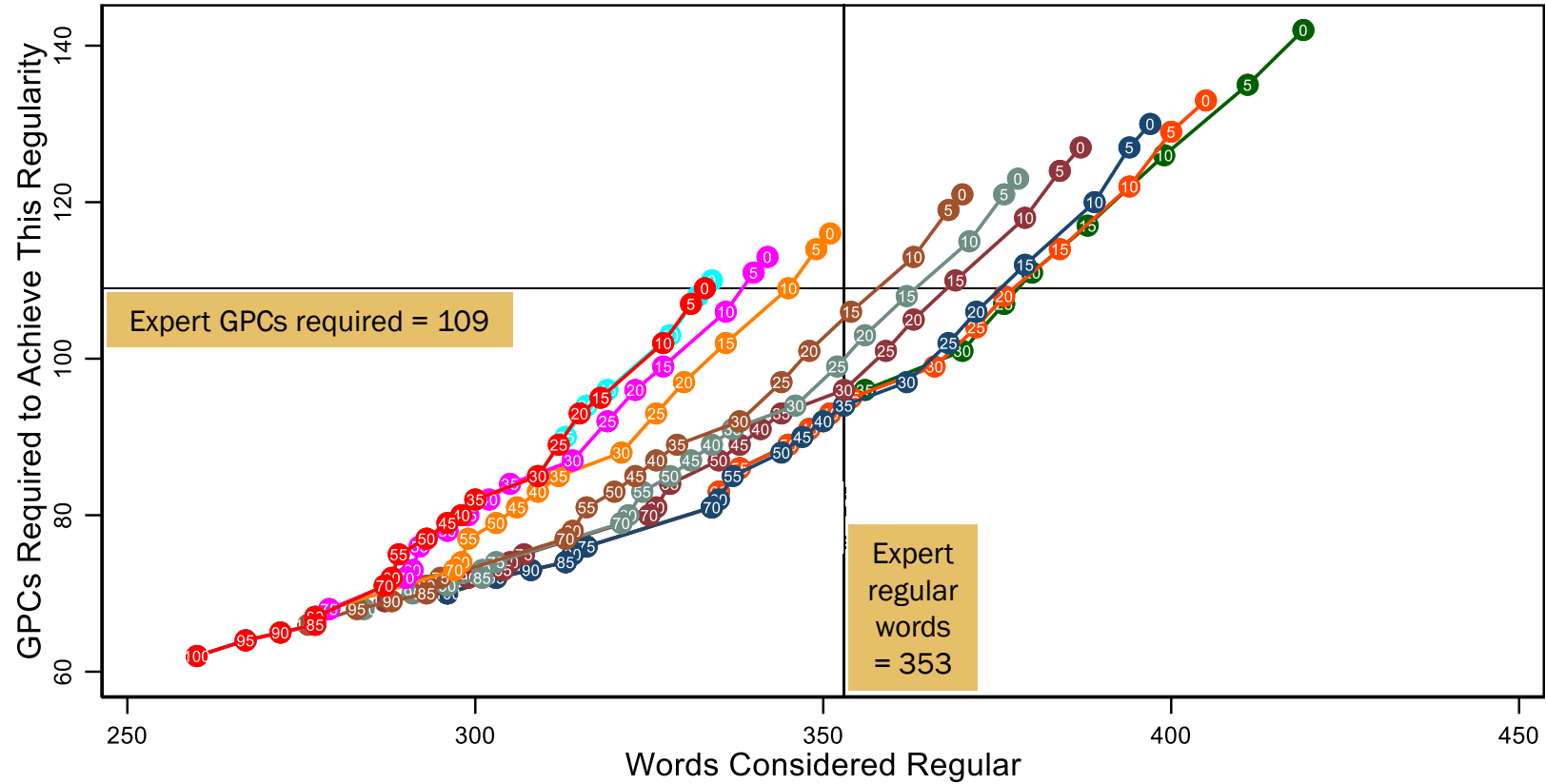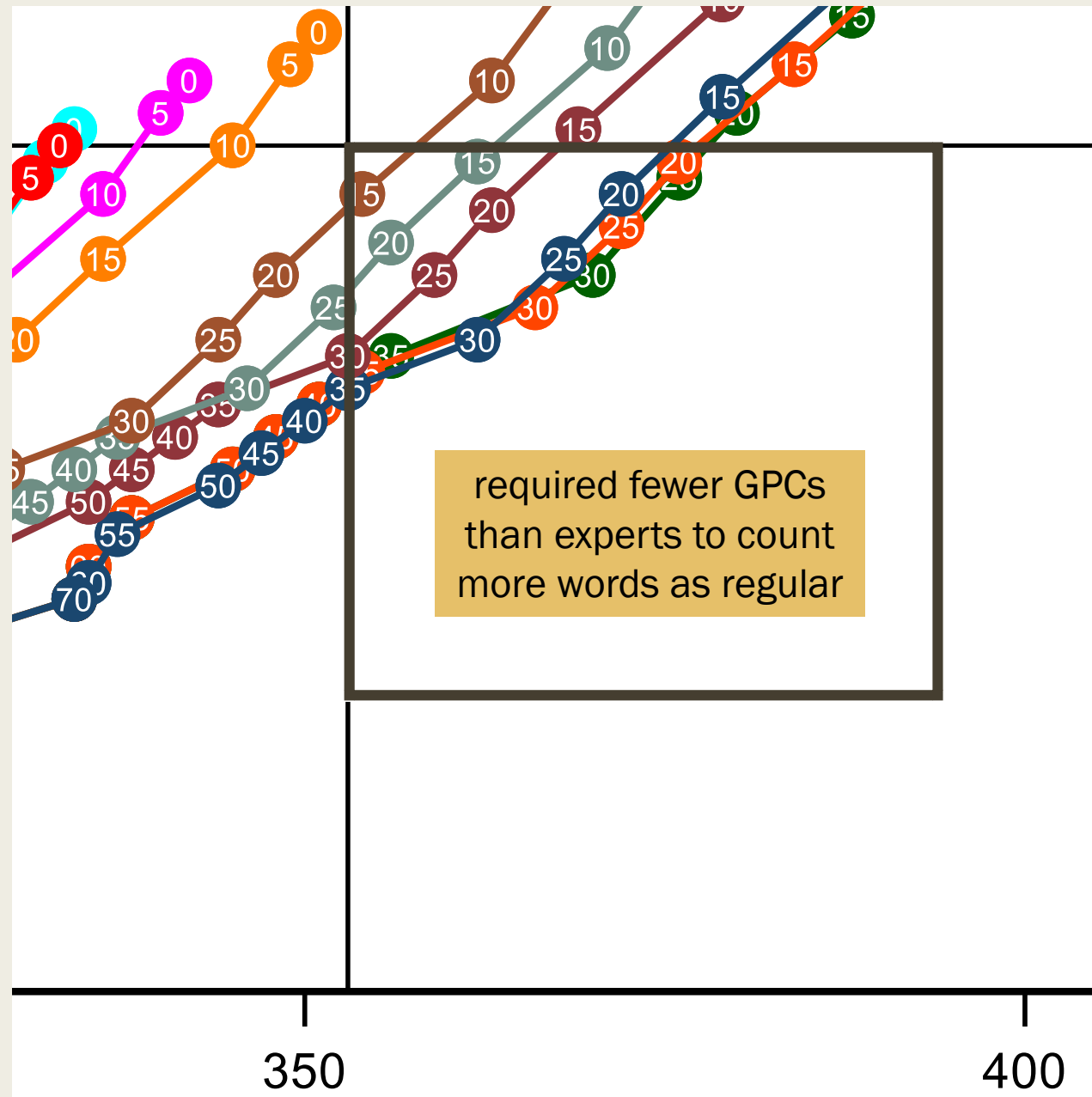
Expert regular words = 353

GPCs Required to Achieve This Regularity

Words Considered Regular

Level of Consistency

0% | 1% | 2% | 3% | 4%
5% | 6% | 7% | 8% | 9%

required fewer GPCs
than experts to count
more words as regular

Possible "Best" GPC Sets

| Min. Cons.	Min. count	Reg/Temp Irr.	Total GPCs
0%	25	376	107
0%	30	370	101
0%	35	356	96
1%	20	376	108
1%	25	372	104
1%	30	366	99
1%	35	354	95
2%	20	372	106
2%	25	368	102
2%	30	362	97
2%	35	353	94
3%	20	363	105
3%	25	359	101
3%	30	353	96
4%	15	362	108
4%	20	356	103
5%	15	354	106

- Balancing number of words considered regular or temporarily irregular against the number of GPCs required to achieve this level of accuracy.

- 2% Consistency with 30 words may be best
    - *362 words*
    - *97 GPCs*

Program Analysis using K Criteria
With 2% Consistency and 30 Word Minimums

WORD CATEGORIZATIONS

■ Reg ■ Temp Irr ■ Perm Irr

64.2

22.2

13.6

WORD CATEGORIZATIONS

■ Reg/Temp Irr ■ Perm Irr

86.4

13.6
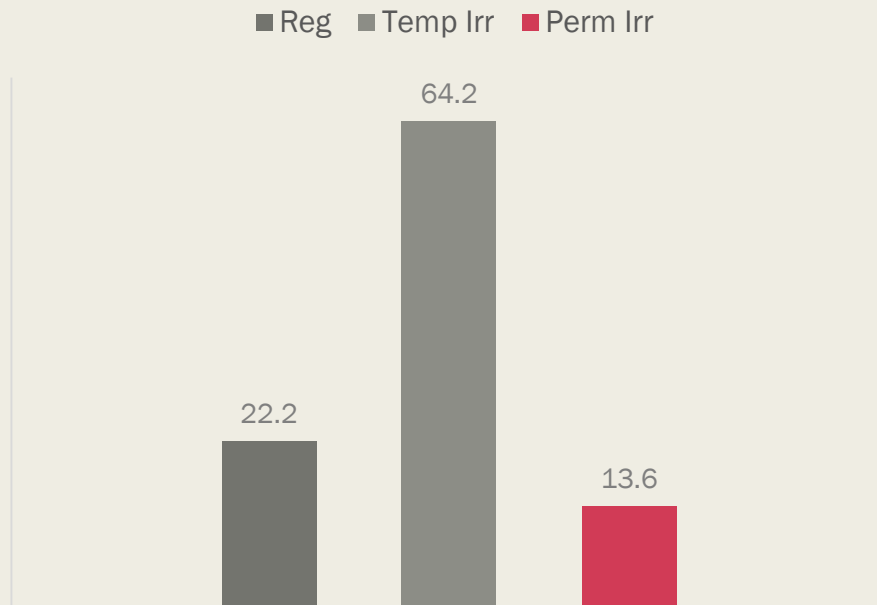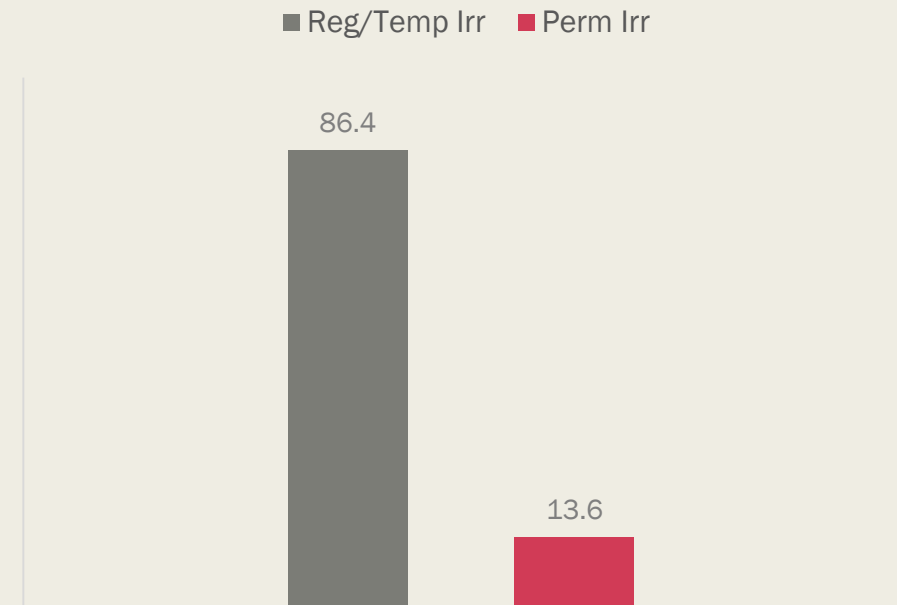
- 362 Words Regularly or Temporarily Irregularly
- 57 Words Permanently Irregularly Spelled

Comparison

Expert

- 84% regular
- 353 words
- 109 GPCs

Program

- 86% regular
- 362 words
- 97 GPCs

ANALYSIS 2: FIRST GRADE

Expert Coding for 1ˢᵗ Grade Word List

- **Common Core Standards regarding phonics knowledge**
  - *Assumes mastery from K regarding primary or many of the most frequent sounds for consonants*
  - *Long and short vowel sounds*
  - *Additionally, common consonant digraphs, final e, common vowel teams, inflectional endings*
- **Moats's List of 44 Phoneme-Grapheme Relations**
- **Fundations Curriculum: Objectives and Scope and Sequence for 1ˢᵗ grade**
  - *R-controlled vowels, vowel digraphs and diphthongs, vowel-consonant-e, double letters, etc.*

Expert Coding for 1st Grade Word List

1st Regular	1st Temp Irr	1st Perm Irr
Rule	Rule	Rule
GPC on Common Core for 1st	GPC on Moats list	GPC not on Moats list
GPC on Moats list	GPC with 4+ occurrences	
GPC in Fundations for 1st Grade		
Examples	Examples	Examples
Short /a/ spelled a		
Short /e/ spelled e	Short /e/ spelled ea	
Short /i/ spelled i	Short /i/ spelled y	
Short /o/ spelled o	Short /o/ spelled wa, al	
Short /u/ spelled u	Short /u/ spelled o, oo, ou	
/ā/ spelled a, a_e, ai, ay, ea,	/ā/ spelled y, eigh, ei, ey	All other spellings of sounds not listed in the 44 P-G List
/ē/ spelled e, ee, e_e, ea, ey	/ē/ spelled y, ie, ei	If 4 or more words with the spelling then coded as temp irr not perm irr
/ī/ spelled i, i_e,	/ī/ spelled ie, -y, igh,	
/ō/ spelled o, o_e, oa, oe, ow		

Expert vs. Program

- The results were the same as for Kindergarten

- It was possible to collapse Regular and Temporarily Irregular and determine level of *matching*

- Level of match was very high

Category	Matches	% of All Words
2% with 30 words	368	87.8
2% with 35 words	375	89.4

Analysis of 1st Grade Word List
Program Temp Irregular vs Expert Regular

Based on program output with most matches (N = 336, 2% consistent with 35 words)

- As, began, begin, don't, even, find, go, has, he, his, hold, I, is, kind, me, most, no, often, old, open, robin, seven, so, we, yes

- Program only used one pronunciation of vowels for regular coding (15 cases)

- Program counted s = /z/ as temporarily irregular because not the most frequent (4 cases)

- Program recognized a schwa sound (7 cases)

Best-matched analysis differences: 375 words using 2% Cons. and 35 word

Program Regular & Experts Irregular:

- Other Useful GPCs?
  - *reduced vowels (schwa): a America idea Indian the*
  - *silent E: give horse house goodbye leave live were*
  - *OR = /er/: word work world*
  - *Other?*
    - E = /I/: pretty (5% consistent, 217 words)
    - A = /A/: father (3% consistent, 108 words)
    - AR = /Or/: warm (7% consistent, 40 words)
- GPC Coding differences: carry (A = /a/) very (e = /e/) does (o = /u/ and e = silent)

Experts Regular & Program Irregular:

- Not useful enough?
  - *Not enough words: bear because blue few goes great group know enough often our they thought you eight these use*
  - *Too inconsistent: talk walk*
- GPC coding differences: example (le = /ul/) here (ere = /eer/) their (eir = /air/)

CONCLUSION

Key Ideas

- *Disconnect between theory/research and practice*

- *Orthographic regularity of words on "Sight Word" lists is common*

- *Consider student knowledge of grapheme-phoneme relations*

- *Program vs. Expert Coding...*

What does a computational approach add (if anything)?

- Something?
  - *Importance of frequency vs. consistency/ regularity*
  - *Suggesting new units to teach*
  - *Drawing awareness to tricky cases (carry)*
  - *Supporting the development of new sequences of sound-spelling instruction*
  - *Lends credibility to the expert coding (with caveats)*
- Nothing?
  - *Much ado*