Tackling Linguistic Difficulty through Serious Games

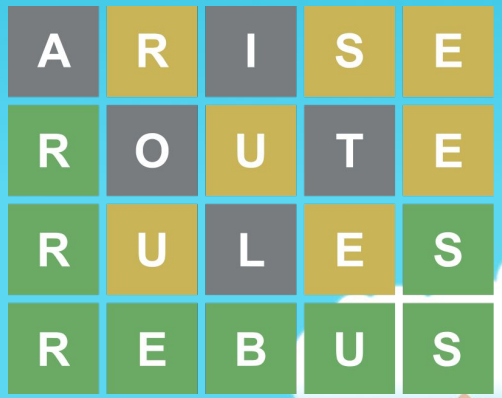
Matthew Pattemore
University of Barcelona

Universitat de Barcelona
Grup de Recerca en Adquisició de Llengües
Language Acquisition Research Group
iREAD personalised reading
JALTCALL 17-19 June 2022

Digital Game-Based Language Learning

Games can be great for language learning:

- Inherently motivational
- Input rich
- Personalised support



Cornillie, F., Thorne, S. L., & Desmet, P. (2012). ReCALL special issue: Digital games for language learning: challenges and opportunities: Editorial Digital games for language learning: From hype to insight?. ReCALL, 24(3), 243-256.

c.f. Young et al. (2012), Yudintseva (2015), Poole & Clarke-Midura (2020) for research syntheses on DGBL and Language Learning

Linguistic Difficulty

Linguistic Difficulty

- Language is undoubtedly difficult to acquire
  - and some language features are more difficult than others

a language feature is more **difficult** than another if its processing and learning requires more time and/or more mental activity from a particular language learner in a particular learning context

Housen & Simoens, 2016

Linguistic Complexity

- Difficulty is a different construct than complexity
  - (but sometimes the words get interchanged)
- Complexity has to do with the number of elements in a language system / text, and their relational patterns

(Pallotti, 2015)

- Sometimes linguistic structures that are structurally more complex may actually be easier to learn and produce

Linguistic Difficulty

- Subjective, Learner-Related Difficulty
  - individual differences in cognitive abilities
  - knowledge of previously learned languages
  - overall L2 proficiency and stage of L2 development
  - socioaffective and personality factors (e.g., motivation, extraversion, and anxiety)
- Objective, Feature-Related Difficulty
  - elaborateness (the number of concepts, steps, subrules, or criteria they contain)
  - conceptual clarity (amount and degree of technicality of metalanguage used)
  - scope (with general rules vs. specific rules)
  - reliability and truth value (number of exceptions) of the rules

Housen & Simoens, 2016

Linguistic Difficulty

- Subjective, Learner-Related Difficulty
  - individual differences in cognitive abilities
  - knowledge of previously learned languages
  - overall L2 proficiency and stage of L2 development
  - socioaffective and personality factors (e.g., motivation, extraversion, and anxiety)

- Objective, Feature-Related Difficulty
  - elaborateness (the number of concepts, steps, subrules, or criteria they contain)
  - conceptual clarity (amount and degree of technicality of metalanguage used)
  - scope (with general rules vs. specific rules)
  - reliability and truth value (number of exceptions) of the rules

INTERACTIONS

Housen & Simoens, 2016

Why does difficulty matter?

- Research
  - do particular instructional approaches work better for easier/harder language features?
  - control for difficulty when investigating other topics
- Teaching
  - should teachers focus on simple or difficult features?

How do we measure linguistic difficulty?

- Until now some of the best indicators of feature difficulty are the judgments of experts (e.g. teachers) and perceptions of students. (Housen, 2014).

- Some few studies that use reaction time measures or brain or eye-tracking methods to approximate cognitive difficulty
  - (e.g. Godfroid, 2016, Morgan-Short et al., 2010, Godfroid & Uggen, 2013)
  - but they necessarily have to focus on a comparison of a small number of features

Context:
The iRead Project

iRead innovation project

- English
- English for dyslexia
- EFL
- German
- Greek
- Spanish

Thousands of children across Europe.

A 4-year EC Funded project (Jan 2017-Dec 2020). This project has received funding from the European Union's Horizon 2020 innovation programme under grant agreement No 731724.


least
X


b u l


The
frogs
made so
bigger
noise

Challenges in interpreting game data

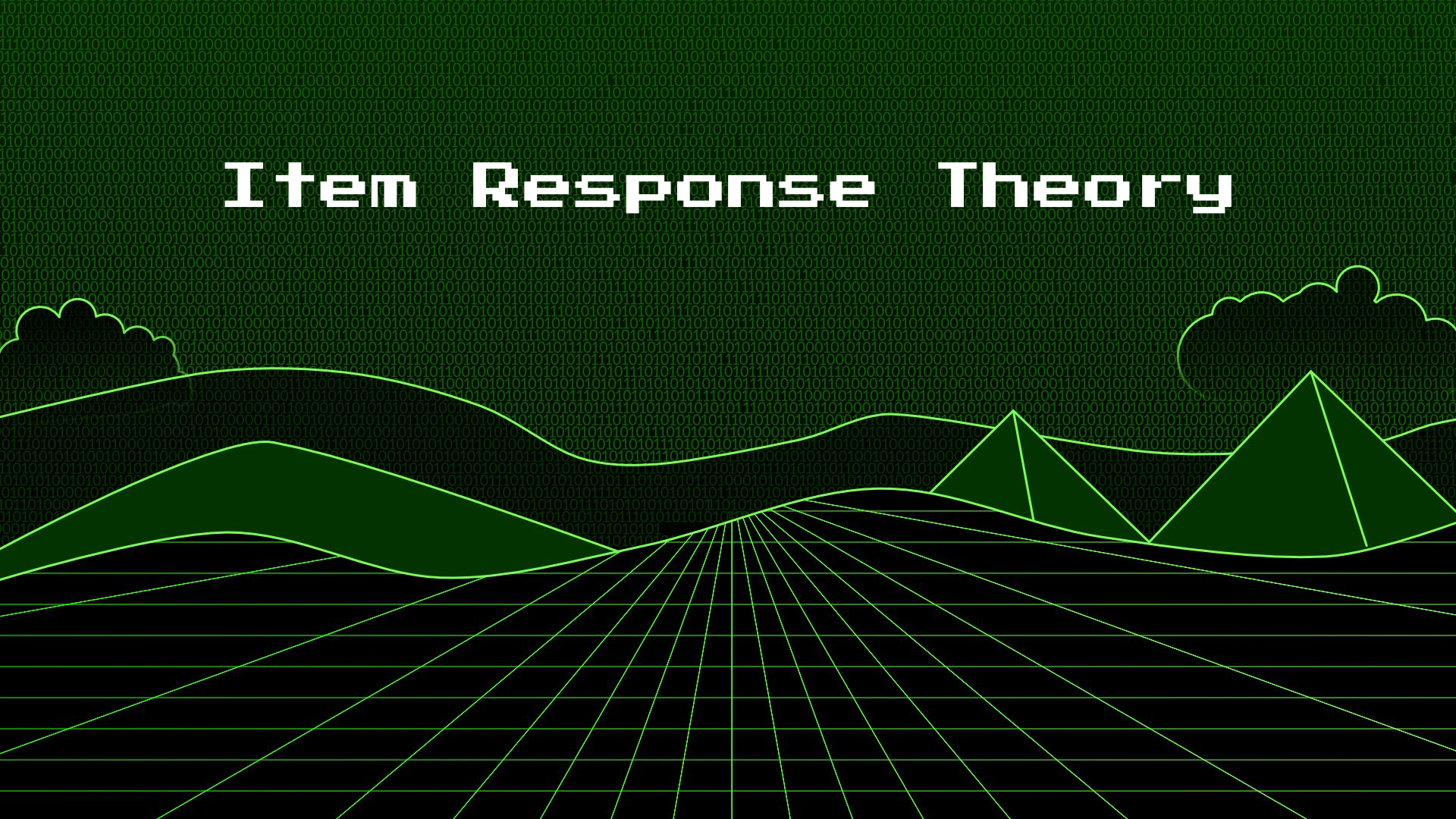
When a child makes a mistake, how do we disentangle whether the mistake is due to low proficiency, the game being challenging, or the feature itself being challenging?

One of the big advantages of digital games is...

DATA

From the games, we have tens of thousands of observations of hundreds of Spanish children playing minigames based on more than 200 features.

Item Response Theory

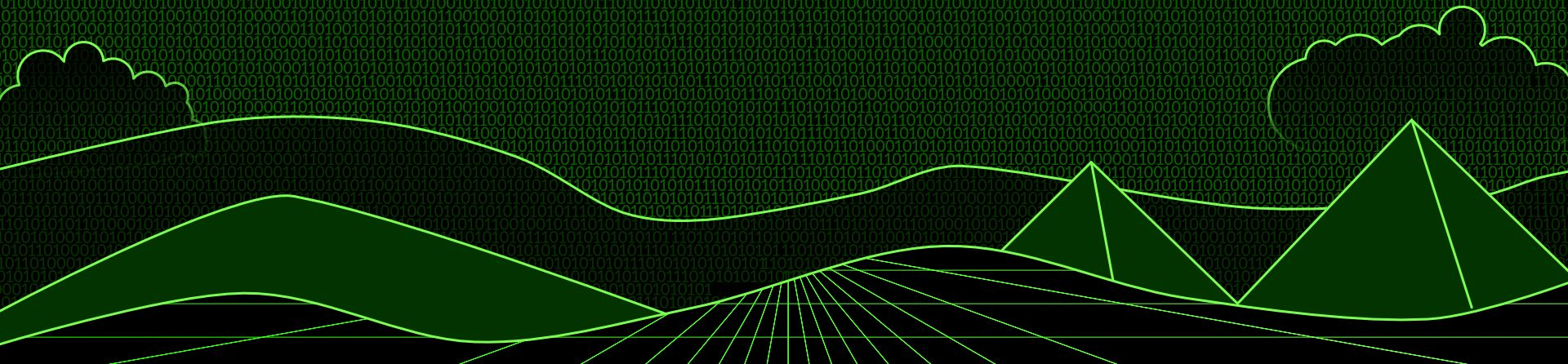


Classical test theory

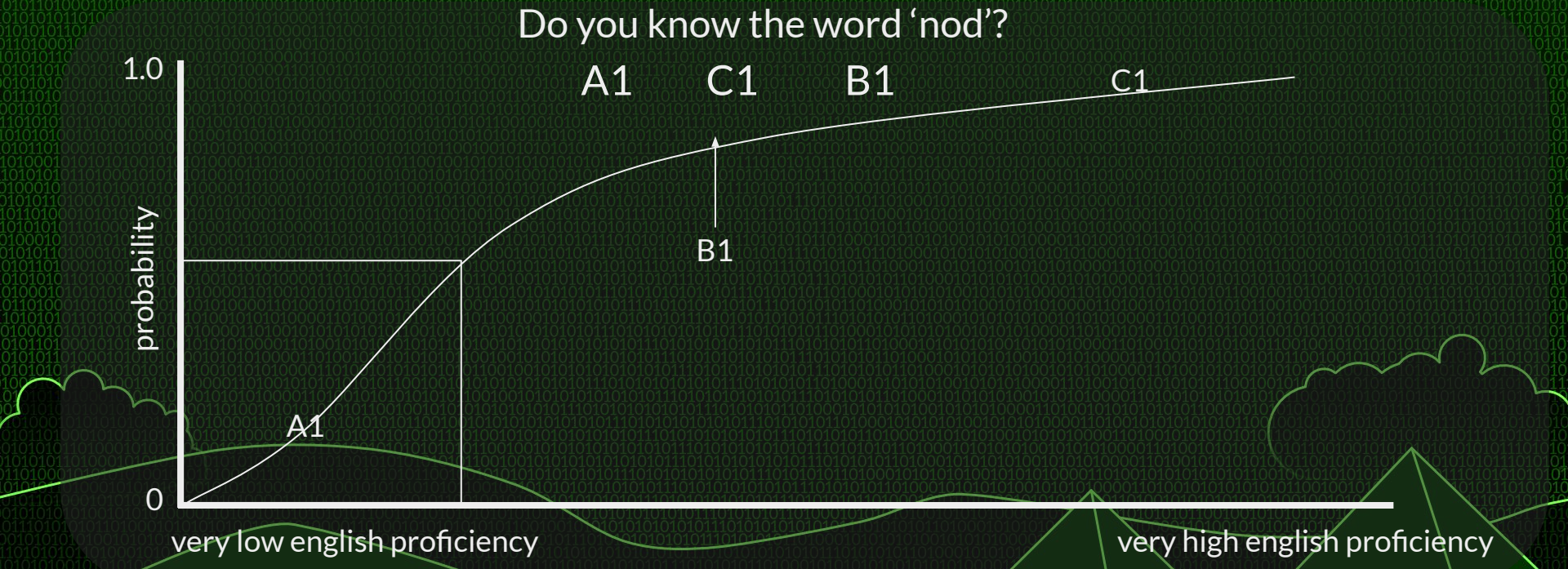
Measured ability = true ability + error

Item response theory

An IRT model estimates the likelihood of different responses to items by people with different levels of the trait being measured.

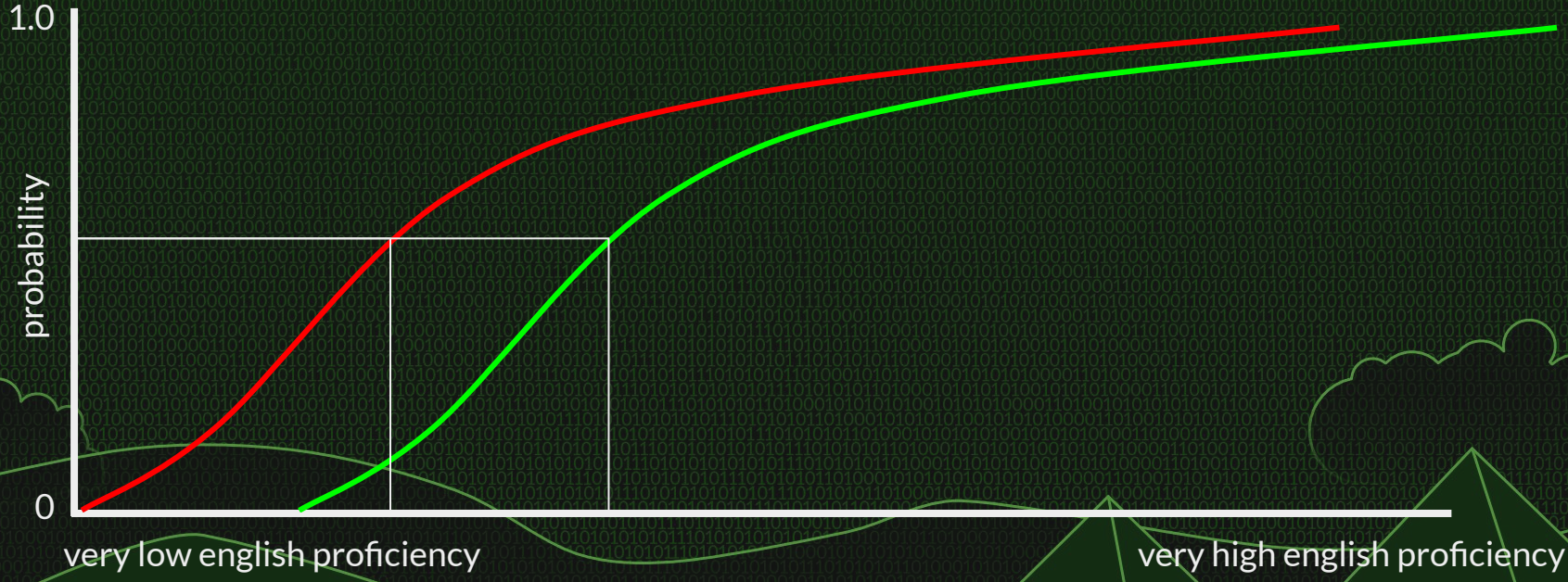


Item response theory



based on Karon F. Cook’s youtube series:
A conceptual introduction to Item Response Theory

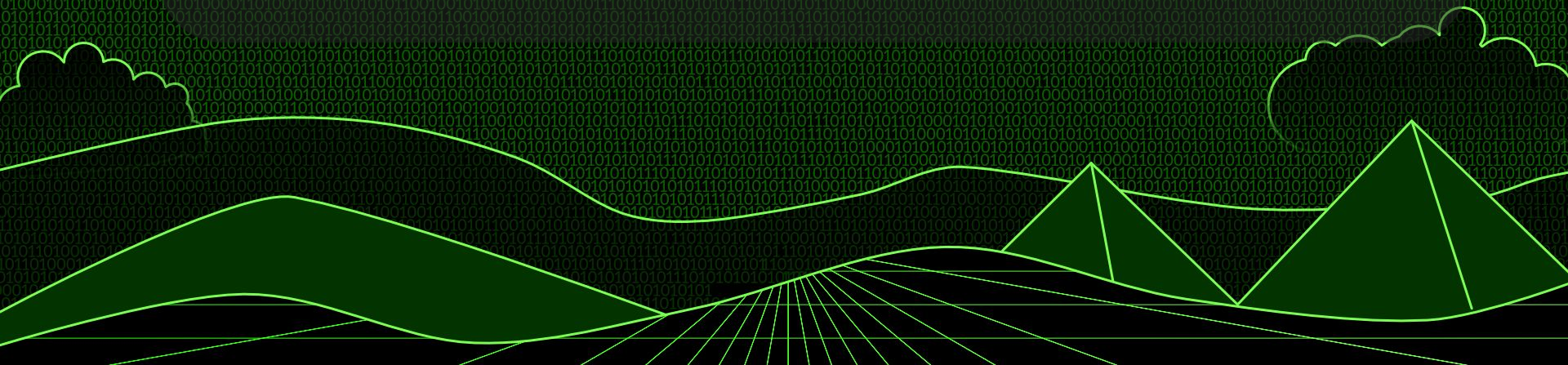
Item response theory – item difficulty



based on Karon F. Cook's youtube series:
A conceptual introduction to Item Response Theory

Advantages of IRT

- IRT models can work with missing data
- IRT models can make data comparable between different instruments
- IRT models can give you information about individual questions
  - e.g. how well they discriminate and at what levels they discriminate best



Item response theory

In practice, we can construct (some types of) IRT models using a binomial form of a GLMM.

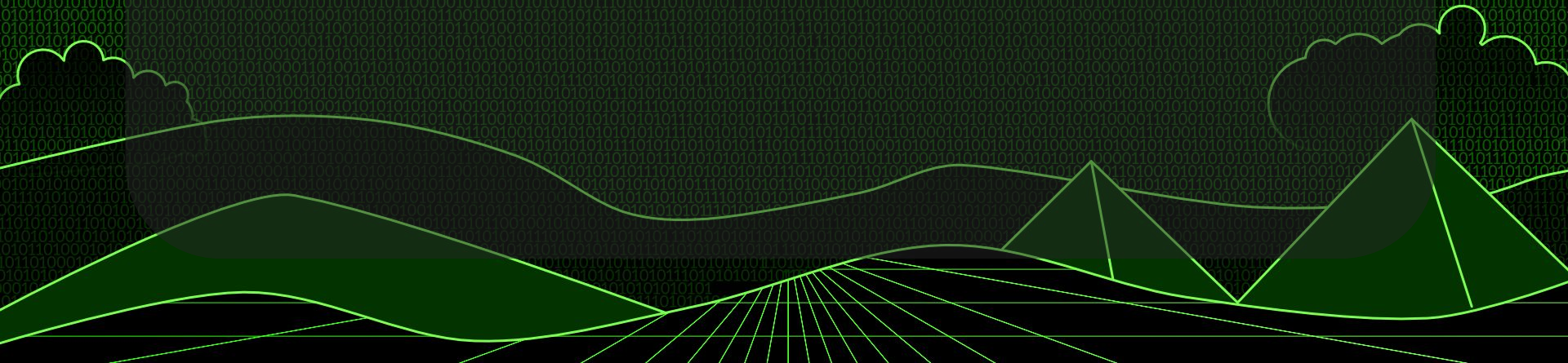
De Boeck et al., 2011

```
IRTmodel = glmer(response ~ -1 + fixedeffect +
(1|randomeffect) + (1|randomeffect), family=binomial("logit"),
data=datasource
```

Applying this to games

Item Response Theory

- “the probability of a correct response to an item is a mathematical function of person and item parameters”
- Each question posed by a minigame is an item




triumph
batch
utograph
sketch
paragraph
fantastic

Item Response Theory

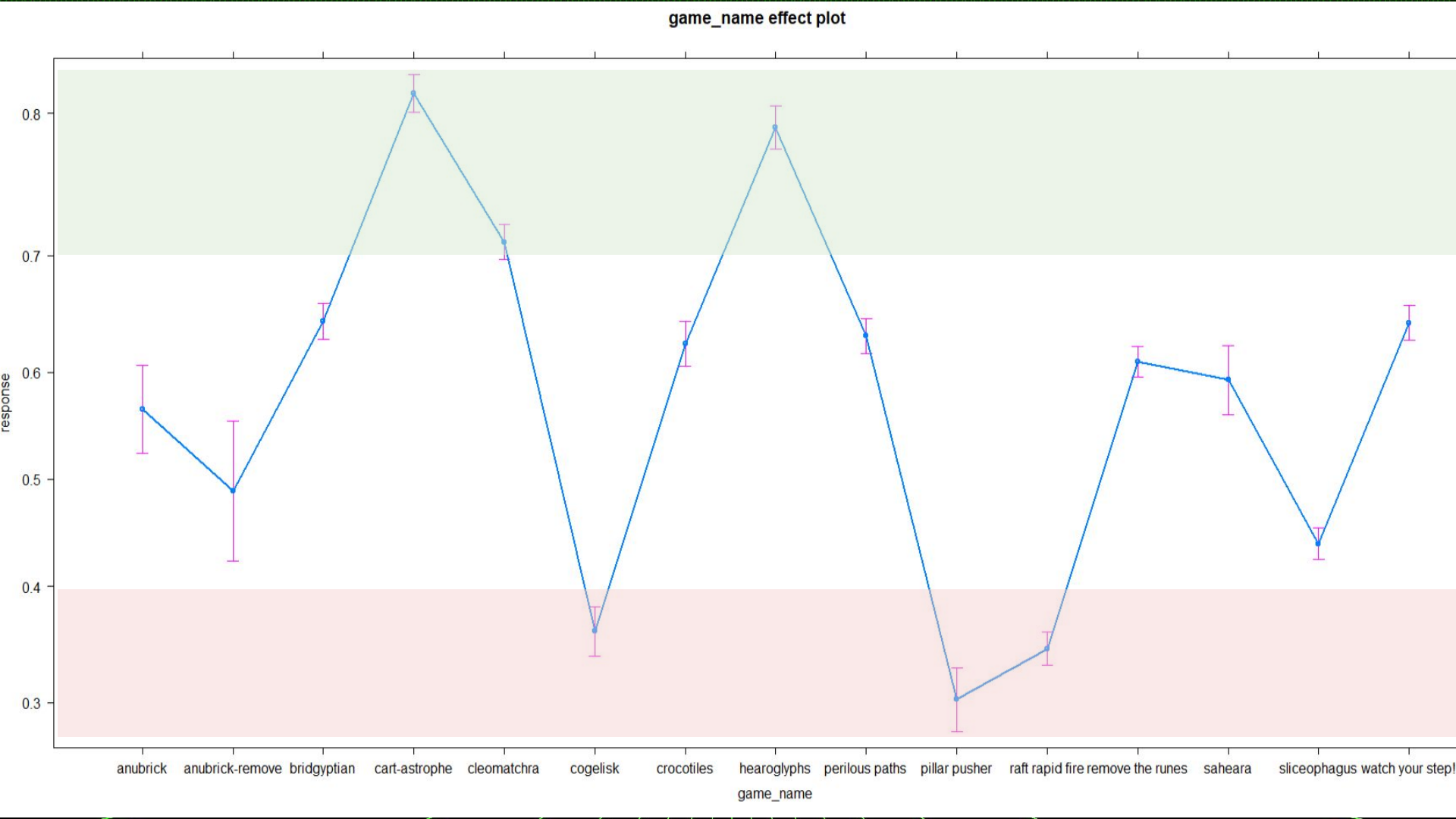
- “the probability of a correct response to an item is a mathematical function of person and item parameters”
- Each question posed by a minigame is an item
- Item parameters include minigame type and feature

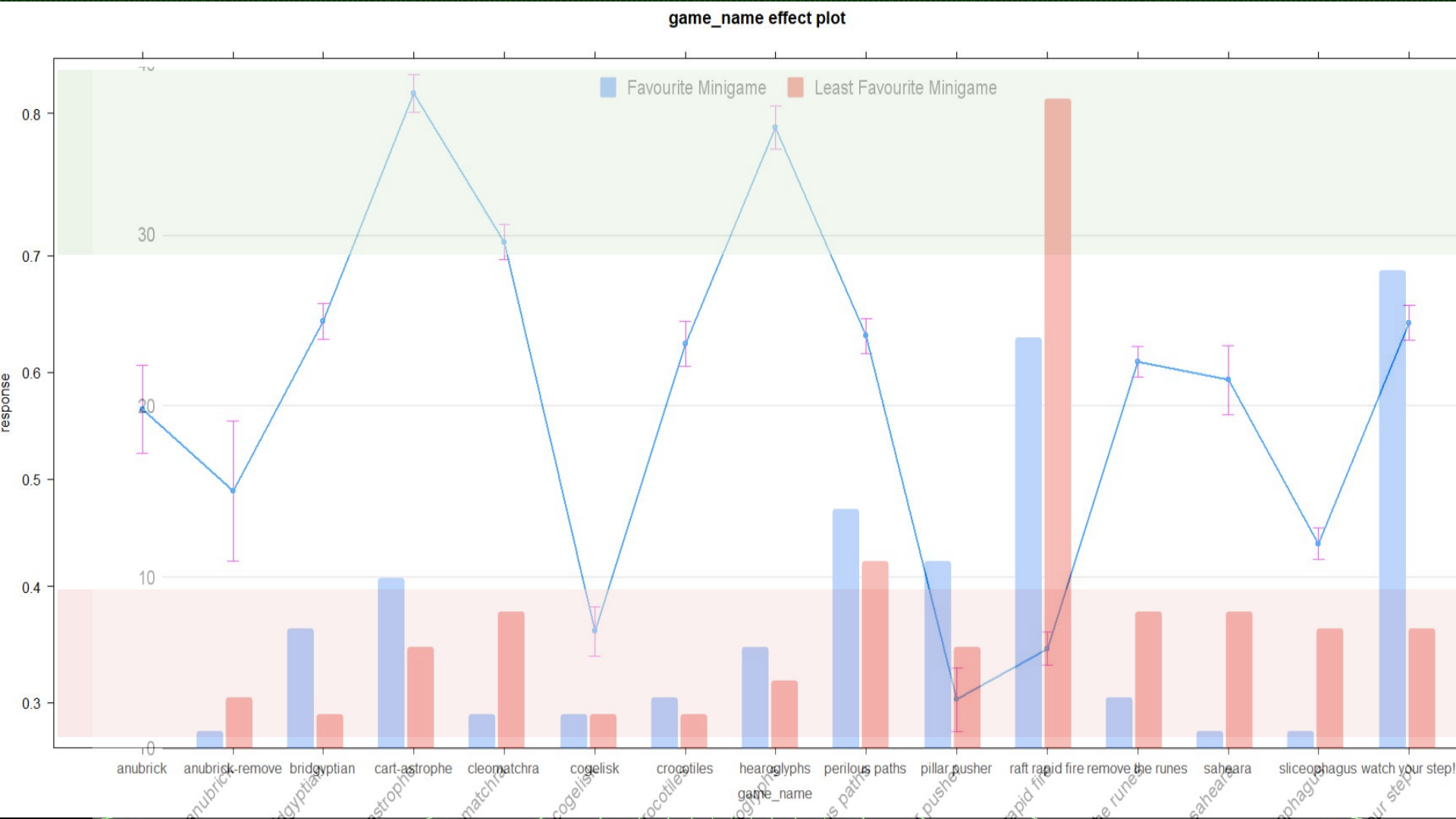
IRT - difficulty model

- **ability + minigame difficulty + feature difficulty**
- This allows us to disentangle feature and minigame difficulty
- Actually utilises the large dataset - ~432000 item responses
- Can create an objective list of relative difficulties of various features for EFL learners

IRT - difficulty model

ability + minigame difficulty + feature difficulty


game_name effect plot
response
0.8
0.7
0.6
0.5
0.4
0.3
anubrick
anubrick-remove
bridgyptian
cart-astrophe
cleomatchra
cogelisk
crocotiles
hearoglyphs
perilous paths
pillar pusher
raft rapid fire remove the runes
saheara
sliceophagus watch your step!
game_name


game_name effect plot
Favourite Minigame
Least Favourite Minigame
response
0.8
0.7
0.6
0.5
0.4
0.3
30
20
10
anubrick
anubrick-remove
bridgyptian
cart-astrophe
cleomatchra
cogelisk
crocotiles
hearoglyphs
perilous paths
pillar pusher
raft rapid fire remove the runes
saheara
sliceophagus
watch your step!
game_name

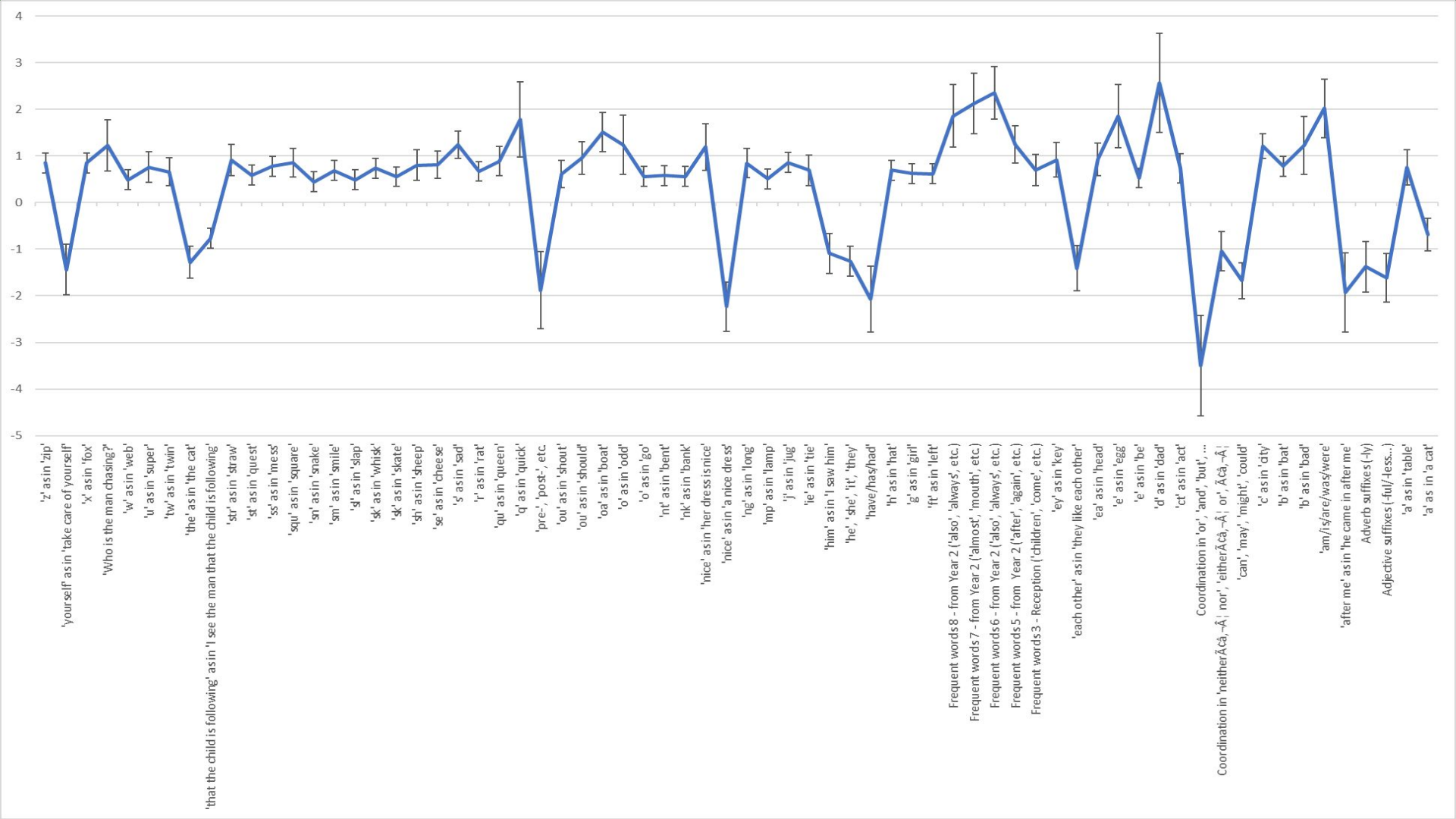
IRT - difficulty model

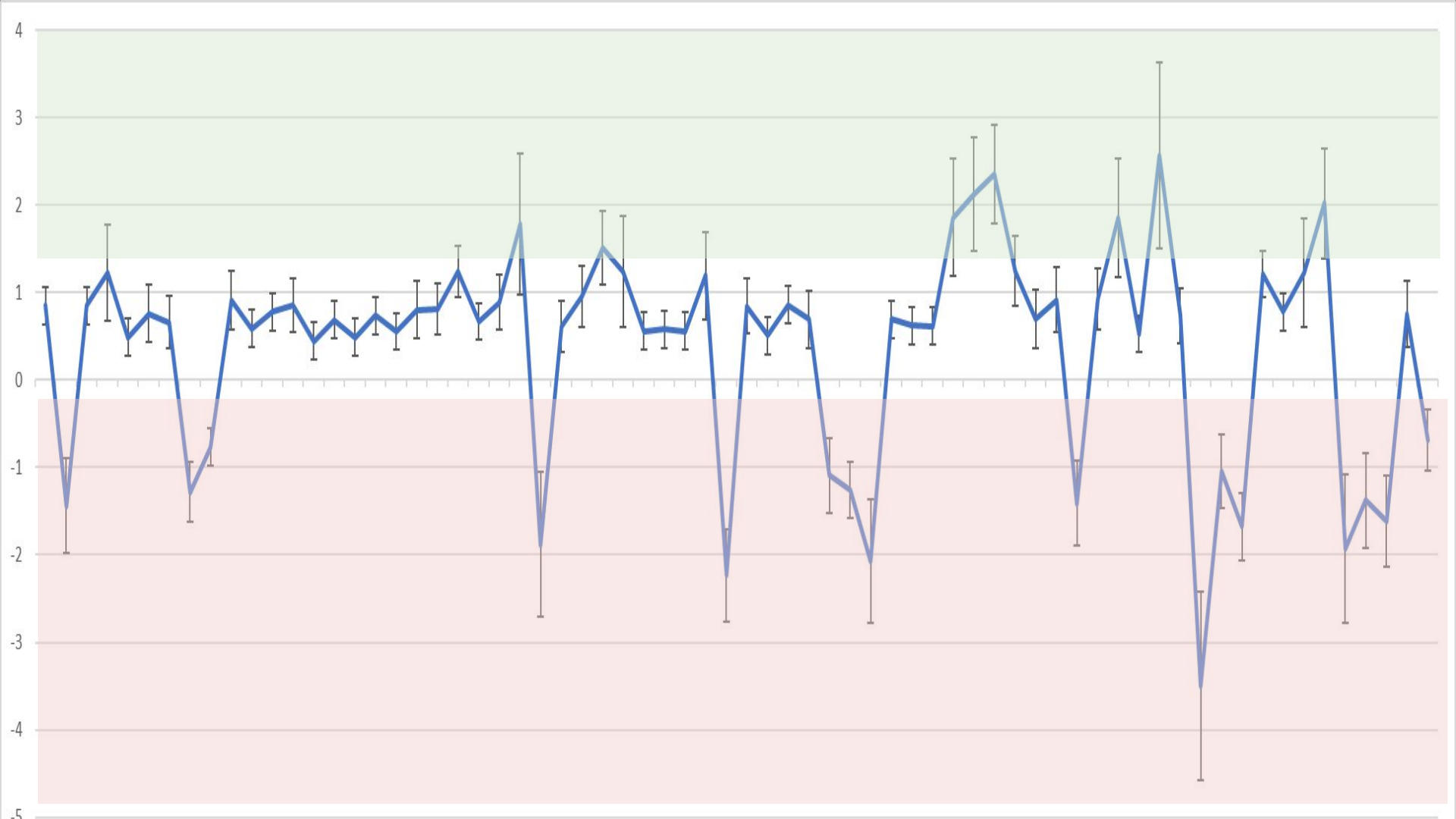
ability + minigame difficulty + feature difficulty

IRT - difficulty model

ability + minigame difficulty + feature difficulty

	A	B	C	D	E
1		Estimate	Std. Error	z value	Pr(>|z|)
2	description(CC)+S+V+DO+OI(CC)	0.6185341326	0.335053338	1.846076616	0.0648810816
3	description/a/-a	1.272109561	0.185686308	6.850852769	0.0000000000
4	description/aa/-aa	0.1474663877	0.210519334	0.7004885698	0.4836222409
5	description/ae/-ae	0.269436068	0.187413734	1.437653801	0.1505323076
6	description/ai/-aÃ	0.3433298742	0.187181510	1.834208265	0.0666230851
7	description/aj/-ai	0.2837153053	0.185918709	1.526018048	0.1270053699
8	description/aj/-ay	0.2170620969	0.184154865	1.178693253	0.2385203433
9	description/ao/-ao	-0.018903032	0.221220198	-0.085448943	0.9319044839
10	description/au/-aÃº	0.3907264532	0.187493342	2.083948413	0.0371648600
11	description/aw/-au	0.3181904956	0.185990995	1.710784415	0.0871209176
12	description/b/-b	0.5442931531	0.187686552	2.900011455	0.0037314902
13	description/b/-v	0.6372239257	0.191081174	3.334833633	0.0008535045
14	description/d/-d	1.033677336	0.191290039	5.403717522	0.0000000653
15	description/e/-e	1.253558949	0.187229133	6.695319937	0.0000000000
16	description/Ê§/-ch	0.6934631719	0.188049770	3.687657634	0.0002263279
17	description/Ê□/-ll	0.6798365367	0.186156271	3.65196687	0.0002602395
18	description/Ê□/-y	0.5233539483	0.189097285	2.767643901	0.0056463112
19	description/É¾/-r	-0.026412240	0.185074730	-0.142711217	0.8865182573


4
3
2
1
0
-1
-2
-3
-4
-5
'z' as in 'zip'
'yourself' as in 'take care of yourself'
'x' as in 'fox'
'Who is the man chasing?'
'w' as in 'web'
'u' as in 'super'
'tw' as in 'twin'
'the' as in 'the cat'
'that the child is following' as in 'I see the man that the child is following'
'str' as in 'straw'
'st' as in 'quest'
'ss' as in 'mess'
'squ' as in 'square'
'sn' as in 'snake'
'sm' as in 'smile'
'sl' as in 'slap'
'sk' as in 'whisk'
'sk' as in 'skate'
'sh' as in 'sheep'
'se' as in 'cheese'
's' as in 'sad'
'r' as in 'rat'
'qu' as in 'queen'
'q' as in 'quick'
'pre-', 'post-', etc.
'ou' as in 'shout'
'ou' as in 'should'
'oa' as in 'boat'
'o' as in 'odd'
'o' as in 'go'
'nt' as in 'bent'
'nk' as in 'bank'
'nice' as in 'her dress is nice'
'nice' as in 'a nice dress'
'ng' as in 'long'
'mp' as in 'lamp'
'j' as in 'jug'
'ie' as in 'tie'
'him' as in 'I saw him'
'he', 'she', 'it', 'they'
'have/has/had'
'h' as in 'hat'
'g' as in 'girl'
'ft' as in 'left'
Frequent words 8 - from Year 2 ('also', 'always', etc.)
Frequent words 7 - from Year 2 ('almost', 'mouth', etc.)
Frequent words 6 - from Year 2 ('also', 'always', etc.)
Frequent words 5 - from Year 2 ('after', 'again', etc.)
Frequent words 3 - Reception ('children', 'come', etc.)
'ey' as in 'key'
'each other' as in 'they like each other'
'ea' as in 'head'
'e' as in 'egg'
'e' as in 'be'
'd' as in 'dad'
'ct' as in 'act'
Coordination in 'or', 'and', 'but', ...
Coordination in 'neitherÃ¢â‚¬Â¦ nor', 'eitherÃ¢â‚¬Â¦ or', Ã¢â‚¬Â¦
'can', 'may', 'might', 'could'
'c' as in 'city'
'b' as in 'bat'
'b' as in 'bad'
'am/is/are/was/were'
'after me' as in 'he came in after me'
Adverb suffixes (-ly)
Adjective suffixes (-ful/-less...)
'a' as in 'table'
'a' as in 'a cat'


4
3
2
1
0
-1
-2
-3
-4
-5

Easier features

feature	Estimate
'd' as in 'dad'	2.566385
Frequent words 6 ('also', 'always', etc.)	2.354535
Frequent words 7 ('almost', 'mouth', etc.)	2.129937
'am/is/are/was/were'	2.021187
Frequent words 8 ('also', 'always', etc.)	1.862003
'e' as in 'egg'	1.854427
'q' as in 'quick'	1.780006
'oa' as in 'boat'	1.506819

- phonology that matches with L1
- common words
- the verb to be
- low frequency consonants

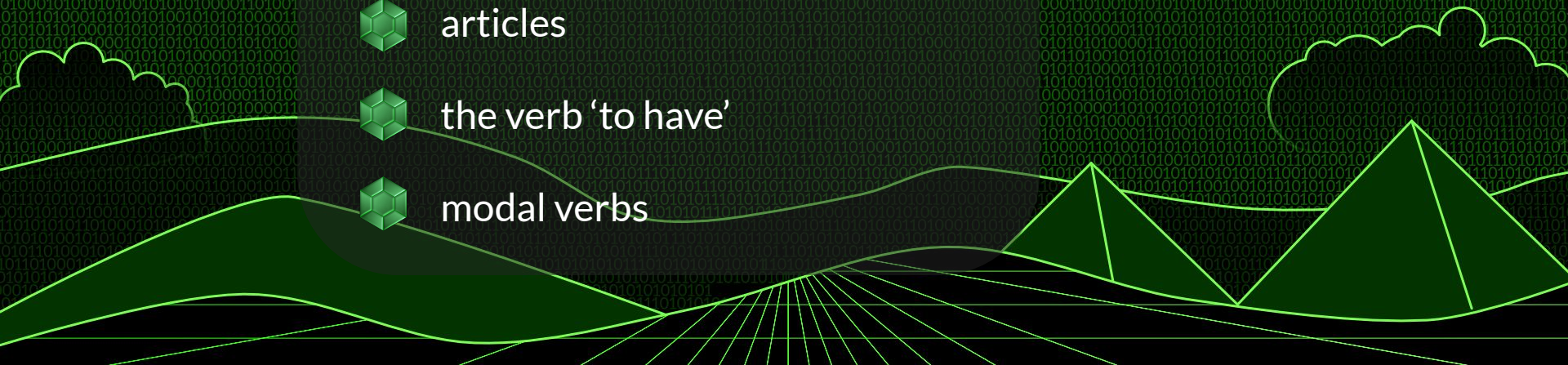
Harder features

feature	Estimate
Coordination in 'or', 'and', 'but', ...	-3.50077
'nice' as in 'a nice dress'	-2.23587
'have/has/had'	-2.07429
'after me' as in 'he came in after me'	-1.93248
'pre-', 'post-', etc.	-1.87932
'can', 'may', 'might', 'could'	-1.6747
Adjective suffixes (-ful/-less...)	-1.61517
'yourself' as in 'take care of yourself'	-1.4372

feature	Estimate
'each other' as in 'they like each other'	-1.4078
Adverb suffixes (-ly)	-1.37697
'the' as in 'the cat'	-1.28361
'he', 'she', 'it', 'they'	-1.25597
'him' as in 'I saw him'	-1.09317
Coordination in 'neither/nor', 'either/or', ...	-1.04798
'that the child is following' as in 'I see the man that the child is following'	-0.7685
'a' as in 'a cat'	-0.68278

Harder features

- conjunctions and coordinators
- anaphora and pronouns
- prefixes / suffixes
- articles
- the verb ‘to have’
- modal verbs

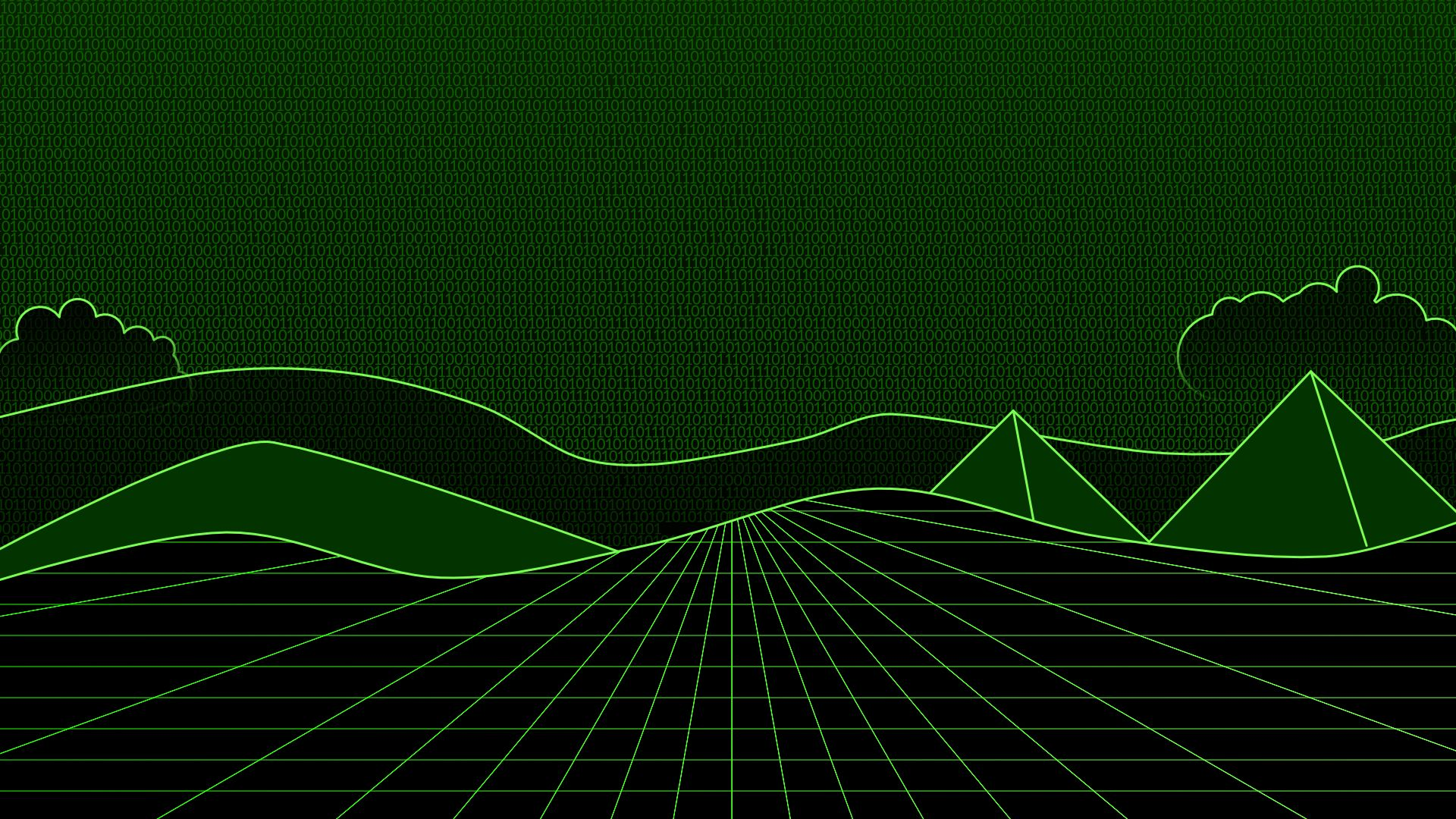


Comparison - L1 transfer

feature	estimate
'nice' as in 'her dress is nice'	1.192034
'nice' as in 'a nice dress'	-2.23587

Her dress is nice - su vestido es bonito

A nice dress - un vestido bonito



Conclusions and Discussion

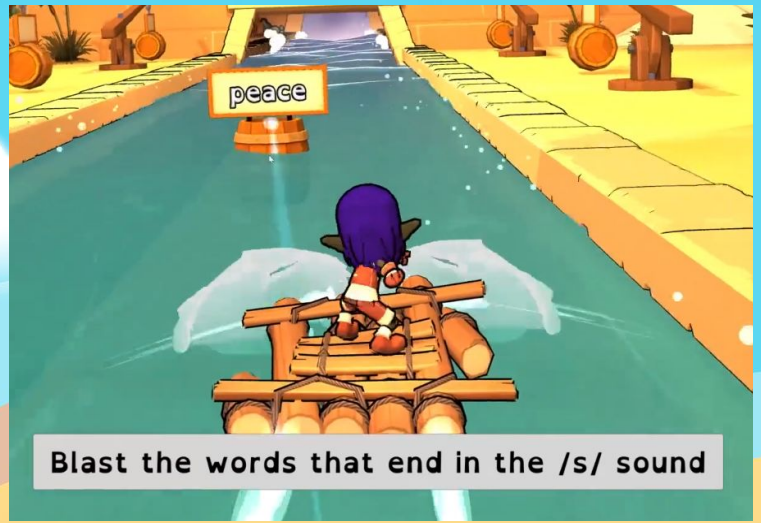
- Using the vast amounts of data from the Navigo games
- and an IRT model
- we can show that some linguistic features are demonstrably easier or harder than others
- for this group of learners (relatively low proficiency, Spanish/Catalan 11-year olds)

Conclusions and Discussion

- Clearly, there would be differences with other populations.
- The way the features are tested may still make a difference.
- We have only limited explanations as to *why* these features are difficult.
- But, conceptually, this seems to be an objective method for investigating linguistic feature difficulty.

- Any other thoughts or comments?

Thank you!



matthew.pattemore@ub.edu

UNIVERSITAT DE BARCELONA

Grup de Recerca en Adquisició de Llengües
Language Acquisition Research Group

iREAD
personalised reading

JALTCALL
17-19 June 2022