# Machine Learning as a Wind Tunnel for Human Learning

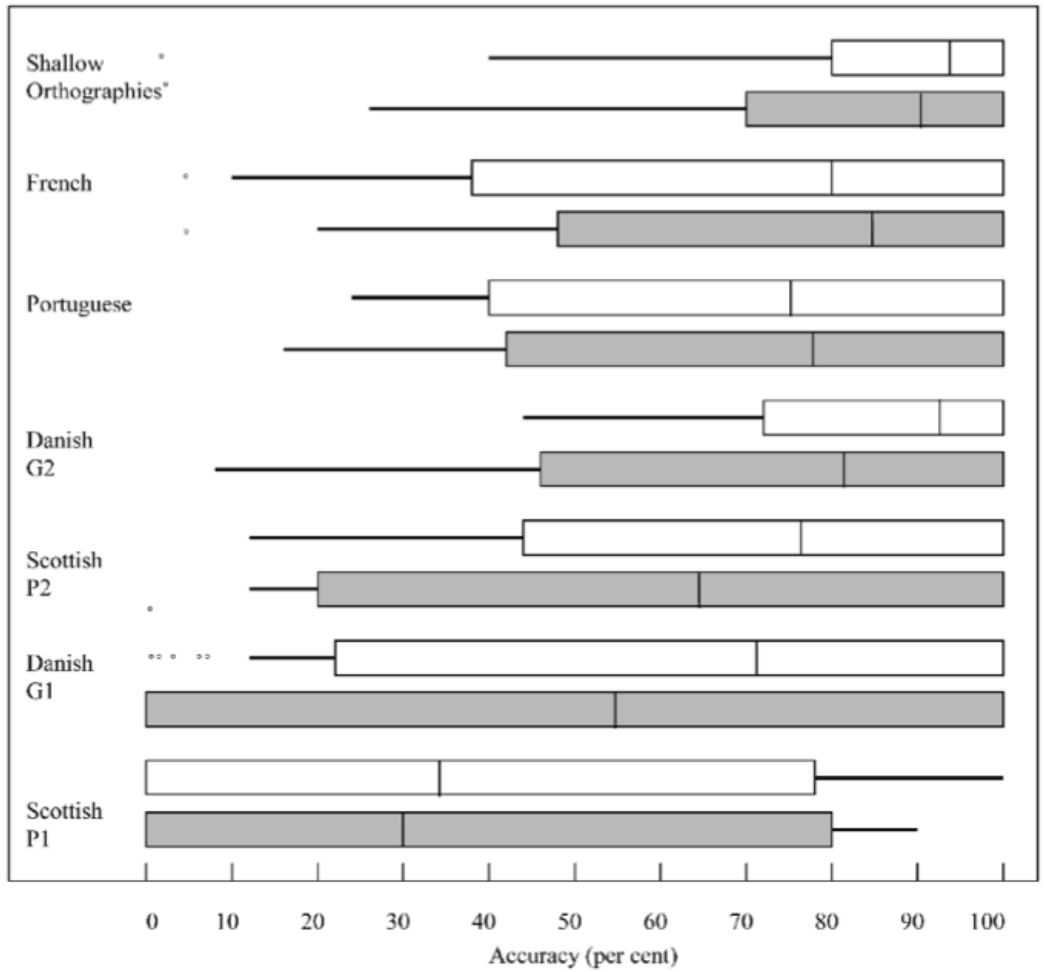
Paul Romer
NYU

September 14, 2018

# English is an Outlier

| | | Orthographic depth | | | | |
|---|---|---|---|---|---|---|
| | | Shallow | | | | Deep |
| Syllabic structure | Simple | Finnish | Greek<br>Italian<br>Spanish | Portuguese | French | |
| | Complex | | German<br>Norwegian<br>Icelandic | Dutch<br>Swedish | Danish | English |

# Orthographic Depth Increases Variance



# Rationalize English Spelling?

- Benefit - Lower inequality
  - Average child will learn to read English more quickly - "meh"
  - Lower incidence of functional illiteracy: $10\% \longrightarrow 5\%$ - "wow"
  - Would mean $5\% * 10$ billion $= 500$ million included as potential readers of English as a second language

# Introduce for People who Learn English as a Second Language

- Cost - Trivial
  - Perfect "plug-in" translation in any digital application
  - "sign" $\leftrightarrow$ "sine"
  - Teachers would need to learn the new spelling
  - This cost can be reduced by choosing new spelling that is close to traditional

# Parallel Systems

- Everyone can talk with each other
- With plug-in translation, everyone can correspond
- Gradually, failed readers in the Algosphere could switch to the new spelling
- In 20 or 30 years, as rationalized spelling is de-stigmatized, usage in the Anglosphere (400 - 500 million) will converge to rest of world (10 billion)

# Designing for a Complex Context

- An orthography is one part of a very complex dynamical system
  - Brain changes as a person learns
  - People interact with each other
- Design an orthography as we designed airplanes using
  - Toy models
  - Wind tunnels (both analog and digital)
  - Test flights
  - Using Field Data
- Conjecture: Machine learning can be a wind tunnel for human learning

# An Aside on Terminolgy

- A "design" interacts with a "context."
- We can study this interaction
  - In a toy model
  - In a wind tunnel
  - On test flights
  - Field data

# Abstraction versus Detail

- Most abstract: A toy model captures interaction of a stylized design with an artificial context–e.g. fluid flow around an infinitely long cylinder
- Intermediate: A wind tunnel accommodates more detailed designs in more realistic contexts–e.g. a scale model of an airplane inside a tube with fans
- A test flight exposes a fully specified design to a subset of the conditions encountered in the field–e.g. a prototype flies in good weather carrying little weight

# An Example from Economics

- Toy Model: A pencil and paper, representative agent, general equilibrium model of labor supply with a linear tax on labor
- Wind Tunnel: The "TPC Microsimulation Model" that the Tax-Policy Center of the Urban Institute uses to evaluate proposed changes to tax law
- Flight Test: Rarely tried, but would take the form of a laboratory or field experiment that exposes subjects to the full complexity of tax law and the details of enforcement.

# The Tax-Policy Wind Tunnel

- Accommodates a far more detailed version of the tax code than a toy model, but must still abstract away many of the details of the law and how it is enforced
- In aeronautical design, a fan-in-a-tube wind tunnel was a special purpose analog computer designed to do a specific set of complex calculations about a model
- Today, we can use a digital computer in place of such analog computers
- Whether analog or digital, a wind tunnel does calculations that are too complex for any human brain, on a model that still abstracts away much of the complexity that an instantiated design encounters out in the field.

# Education

- Toy Models: Described below
- Wind Tunnel: None
- Flight Tests: $>$ 10k trails at the What Works Clearinghouse
- Field Data: Internationally comparable tests: NAEP, PISA, TIMMS, ...

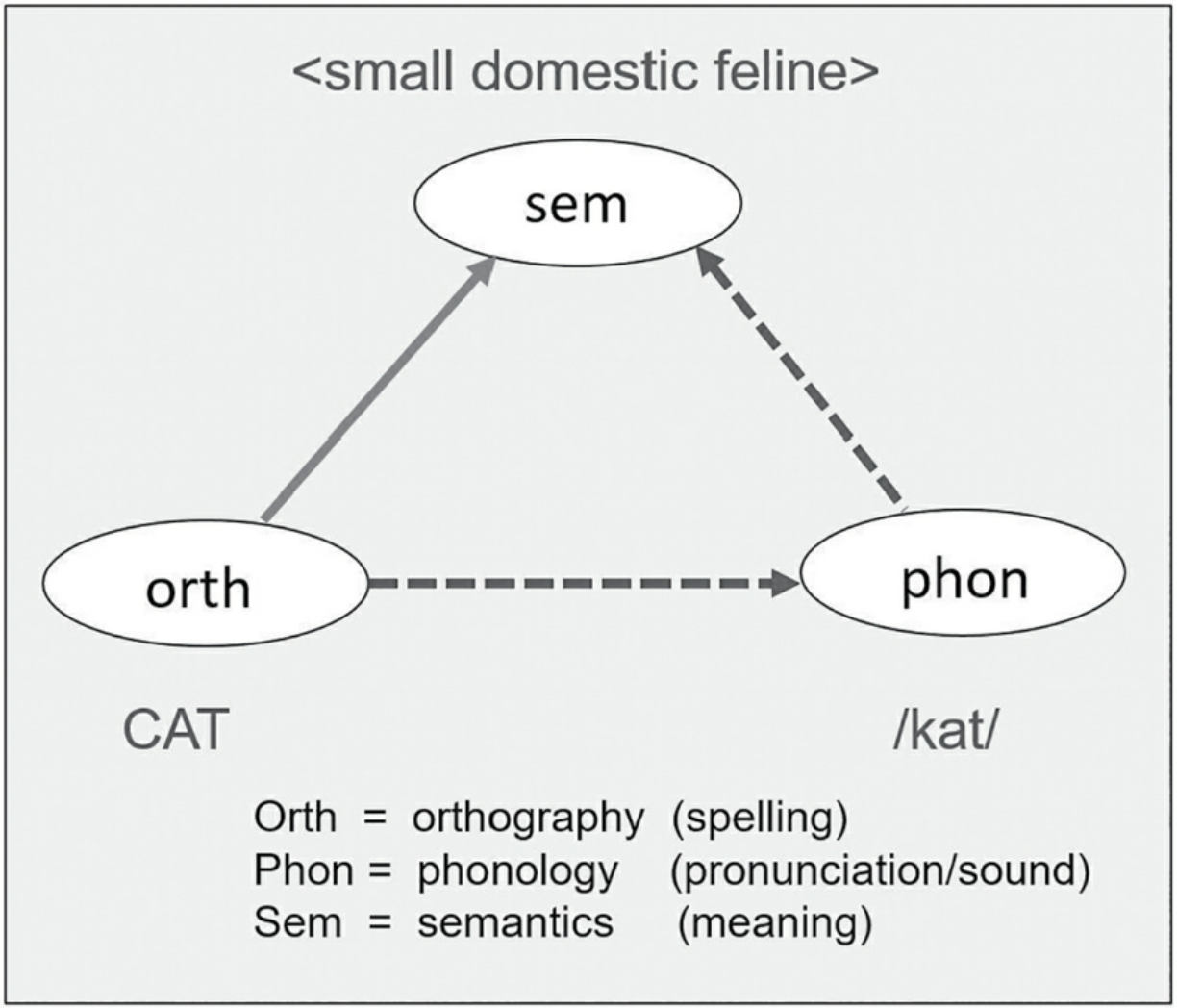
In North America, there is a huge gap about how to teach reading between academic users of toy models and the practitioners who cherry pick from flight tests

# To Make Progress

- Insist on consistency between toy models, wind tunnels, flight tests, field data
- Adjust methods in parallel as evidence accumulates
- Experience with a wind tunnel yields insights about designs and about ways to improve the model that is built into the design of the wind tunnel
- The most valuable output from work on English orthography might to demonstrate that ML can generate artificial learners that we can use as a wind tunnel that vets teaching methods

# Toy Model of Reading

- There already exists a toy "box and arrow" model

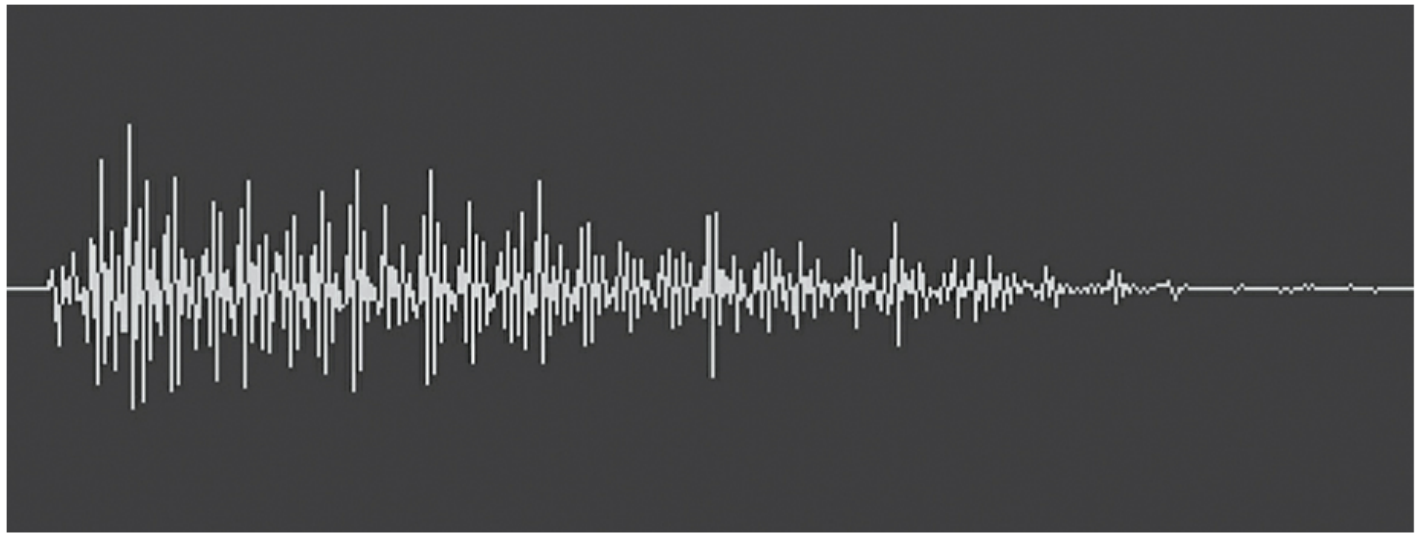


# Toy Connectionist Model of Reading

- "Connectionist" models of a brain that learns the mapping between these three domains
  - One or two hidden layers
  - Hundreds or thousands of nodes
  - Thousands or words to learn
- As far as I can tell, none of these models take advantage of recent progress in deep learning

# Phonemes

- Sound of “bat”



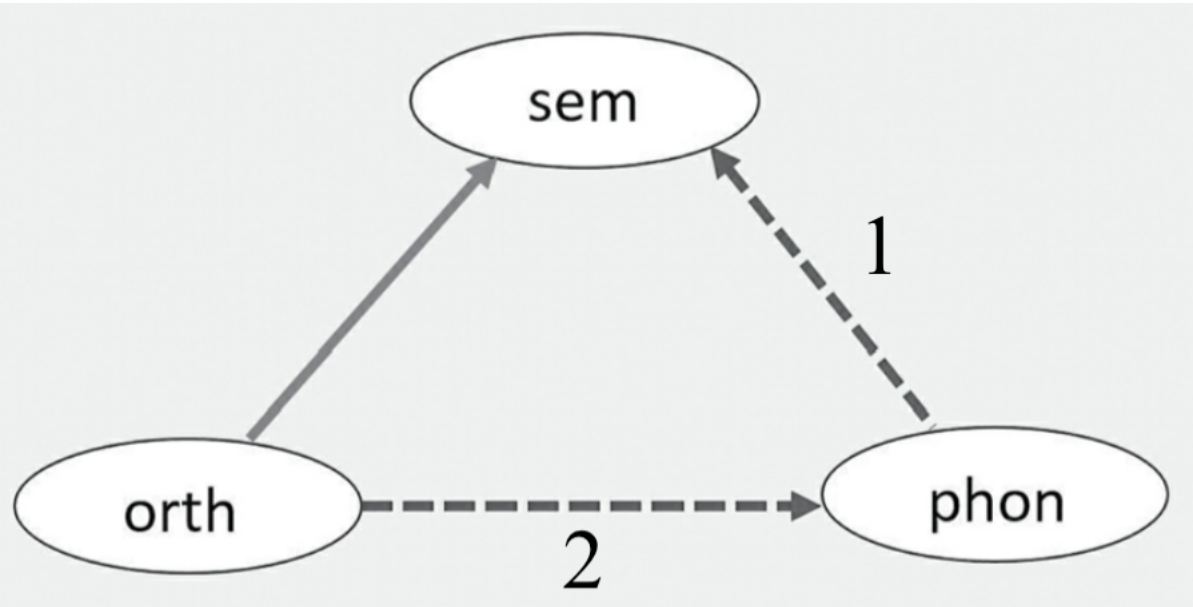
- Division into three phonemes "b", "a", and "t" is an abstraction that children learn with exposure to orthography

# Phonemic Awareness

- Ask a pre-reader out loud: "Does 'cat' end with the same sound as 'bat' "
- "If you take the first sound of 'hit' and add it to the ending sound of 'bat', what do you get?"
- People who are illiterate find it very difficult to answer these questions
- Pre-readers who can't answer them are much more likely to fail at reading

# Dynamics

1. Everybody learns to speak; that is, to associate semantics and phonology for at least a small number of words $N$
2. Children who learn to sound out printed words can read all $N$: orthography $\rightarrow$ phonology $\rightarrow$ semantics



# Dynamics

- Then reading causes explosive growth in $N$ via unsupervised learning:
  - Semantics from context, sound from orthography $\rightarrow$ phonology
- As reading age increases,
  - less reliance on orthography $\rightarrow$ phonology $\rightarrow$ semantics
  - stronger direct association orthography $\leftrightarrow$ semantics
- Resistance to phonics instruction comes from confusion about end state versus transition dynamics

# First Step in a New "Netflix Contest"?

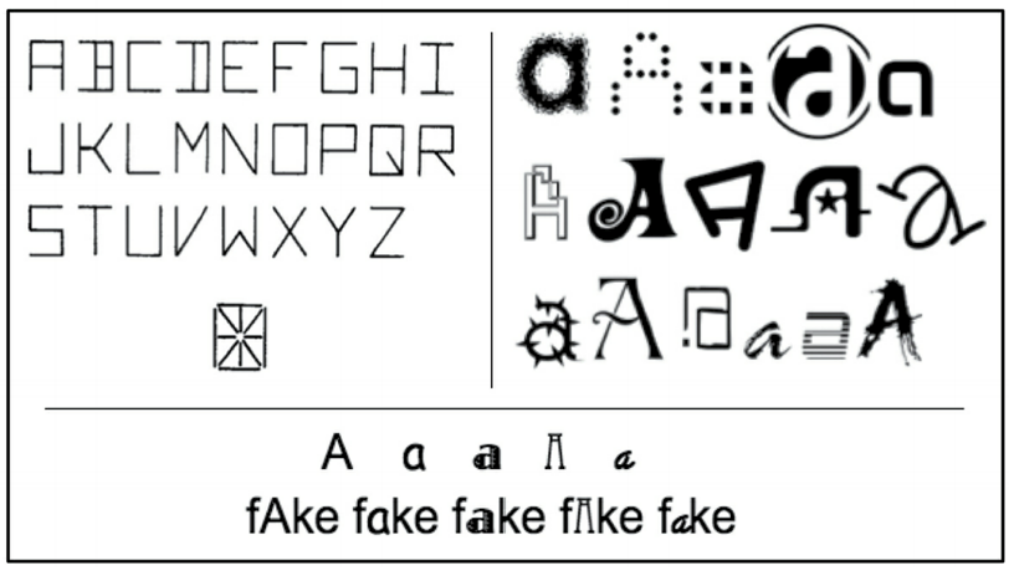
- Assemble a training set (to be specified below)
- Solicit models
- Require that each model include a base architecture and a parametric variants that generate a distribution of individual instances
- Need learning time for the artificial learners to be orders of magnitude faster than humans; simulate 5 or 10 human years in minutes or hours

# Elements in the Training Set for Toy Connectionist Models

- "cat" as code for orthography
- "/kat/" as code for the phonology
- attributes: "mammal", "has tail", "singular", ... to code semantics

# Richer Possibilities for the Training Set

- Animations with speech and narration to present phonology and semantics
- Initially, keep using digital strings to present orthography
  - Eventual extension - text as bit maps



# Next Step

- Test the predictions of the models against evidence on human readers
- Use evidence from quick experiments related to reading
  - Takes judgment about the aspects of human behavior that are important for learning to read
  - For example, there is detailed data about "jerky" eye of a person who reads a text
  - These details of vision seem not to be a first order concern for models of the interaction between orthography and reading
- Based on accuracy of predictions, adjust the set of models
- On a longer time horizon, adjust the training set too

# At Each Stage

- For each model try different orthographies and measure
  - Average time to learn to read
  - Dispersion in the time to learn to read
- Train models with traditional spelling then measure
  - Average time to learn to read with new orthography
  - Dispersion in the time to learn to read with new orthograpy
- Identify an efficiency frontier
- See how robust are the relative positions of orthographies in this space are robust across the collection of models

# Iterate

- Over time, this process should converge to
  - a collection of artificial models, each of which has a distribution of learners
  - an efficiency frontier that captures the trade offs between faster average learning, less variation in student outcomes and how difficult it is for people trained on the old orthography to learn the new one

# Reactions

1. "Only cranks and crazies support spelling reform"
   - They were right about the enormous potential benefit from spelling reform
   - Restating the benefits in a louder and more commanding tone of voice did not turn out to be a working strategy for implementing a reform

# Reactions (cont.)

2. "If you make this proposal, you will be perceived as a crank or a crazy"
   - True

# Reactions (cont.)

3. "The invisible hand tells us that the prevailing equilibrium must be optimal"
   - Seriously?
   - In science, a fact always beats a theory

# Reactions (cont.)

4. "Who could take the lead and establish the new standard"
   - Tech industry; Python as an analogy?
   - Chinese government
     - Chinese will presumably choose a new spelling that is easier for new learners, hence harder for a traditional speller to learn

# Reactions (cont.)

5. "It's impossible to change spelling"
   - False, many languages have reformed their spelling
   - It does require an actor who can solve a coordination problem
   - English is the obvious laggard
   - It is a cruel accident that a language with one of the worst orthographies is becoming the universal second language

# Reactions (cont.)

6. "It's impossible to change now"
   - What evidence generated your posterior?
   - Give me a range of certainty around your estimate of a zero probability that rationalized spelling could enter and survive
   - Tell me the implied expected return to investment in rationalized spelling

# Reactions (cont.)

7. "With proper instruction and enough student effort, the 5% you want to help can learn to read English with no change to its spelling"
   - Probably true, but is this the efficient solution?
8. "So we should just reform schools"
   - Read the exchanges in the phonics versus whole word reading wars
   - Describe your posterior for the probability that the school reform you have in mind will succeed
   - Give me your estimate of me expected return to effort devoted to school reform
   - THen we can compare it to the expected return to spelling reform

# Reactions (cont.)

8. "The solution to problems with literacy is to pay for universal preschool"
   - Investment in preschool has a high expected return, but it is likely to have the same effect on inequality as universal primary and secondary education,
   - which is to amplify small differences in ability

# Reactions (cont.)

10. "The solution to inequality is to give everyone a guaranteed income and put the failed readers out to pasture"
    - Have you asked the failed readers how they feel about this solution?
    - Or the people who will keep working and pay the taxes?
        - If low wage workers see themselves as the "deserving poor",
        - and see those who do not work as the "undeserving" poor,
        - will a system of taxes and transfers that compresses the income differential between those who work and those who do not build the sense of social solidarity needed for the taxes and transfers to be a stable political equilibrium?

# Reactions (cont.)

11. "I don't have time to think about a policy that might reduce inequality in 20 or 30 years. I'm too busy sounding the alarm about the inequality that AI will create in 20 or 30 years."
    - Hmmm

# Van Orden Effect

- Moreover, the orthography $\rightarrow$ phonology $\rightarrow$ semantics pathway remains
- Give skilled readers a category and ask if a string of characters corresponds to a word that belongs

# Van Orden Effect

- category: animals
- string: “horse”

# Van Orden Effect

- category: animals
- string: "hoarse"

# Van Orden Effect

- category: clothing
- string: "sote"

# Van Orden Effect

- category: clothing
- string: "sute"

# Van Orden Effect

- Homophones such as “hoarse”, “sute” induce errors
- Misspellings such as “hourse”, “sote” do not

# PISA Scores, Girls

| | PISA Reading Girls, Percent | | | |
|---|---|---|---|---|
| | Below Level 2 | | At or Below Level 2 | |
| | | Anglo Average | | Anglo Average |
| Australia | 13.3 | | 33.3 | |
| Canada | 7.5 | | 24.5 | |
| Ireland | 8.0 | | 27.3 | |
| New Zealand | 12.4 | | 32.0 | |
| United Kingdom | 14.8 | | 37.5 | |
| United States | 15.3 | | 38.0 | |
| Singapore | 8.5 | | 24.7 | |
| | | 11.4 | | 31.1 |
| Finland | 5.7 | | 18.9 | |

# PISA Scores, Boys

| | PISA Reading Boys, Percent | | | |
|---|---|---|---|---|
| | Below Level 2 | | At or Below Level 2 | |
| | | Anglo Average | | Anglo Average |
| Australia | 22.8 | | 45.4 | |
| Canada | 13.9 | | 34.8 | |
| Ireland | 12.3 | | 34.8 | |
| New Zealand | 22.2 | | 43.7 | |
| United Kingdom | 20.9 | | 46.6 | |
| United States | 22.6 | | 45.7 | |
| Singapore | 13.7 | | 31.3 | |
| | | 18.3 | | 40.3 |
| Finland | 16.1 | | 37.8 | |

# Tertiary Degrees, Finland and Various Anglo Countries

Females, Percent of Tertiary Graduates

| | | |
|---|---|---|
| Finland | 2016 | 59.3 |
| Italy | 2015 | 59.0 |
| United States of America | 2015 | 58.4 |
| United Kingdom of Great Britain and Northern Ireland | 2015 | 57.1 |

# Tertiary Degrees, Other Countries

Females, Percent of Tertiary Graduates

| | | |
|---|---|---|
| Portugal | 2015 | 59.3 |
| Puerto Rico | 2015 | 61.0 |
| Syrian Arab Republic | 2016 | 59.1 |
| Cuba | 2016 | 56.8 |
| Qatar | 2016 | 61.9 |
| Chile | 2015 | 56.0 |
| Brazil | 2015 | 61.1 |
| Tunisia | 2016 | 65.1 |