



Computational Education: A Data Opportunity?

Rakesh Agrawal

Microsoft Research

Mountain View, California

October 14, 2013

Thinking About Education

Three key questions:

- What is being taught
 - Curriculum, syllabus, educational material
- How it is being delivered
 - teachers, classes, assessments
- How it is funded
 - business models

Emergent Perfect Storm

- **Electronic textbooks**
 - Fast adoption of cloud-connected electronic devices (worldwide)
 - Open content (e.g. OpenStax, ck12.org, NCERT)
- **Internet-based classes**
 - MOOCs (e.g. Coursera, EdX, Udacity, Khan, TED-Ed)
 - Small virtual classes (e.g. Shankar Mahadevan Academy)
 - Electronic certification (e.g. Mozilla's OpenBadges)
- **New models of funding education**
 - Recipients give back to the seed fund for future recipients at their pace (e.g. Dakshana)
 - Market for options on future earnings

Data Mining for Enriching Electronic Textbooks

Diagnostic tools for identifying weaknesses in textbooks

Within section deficiencies

Syntactic complexity of writing and dispersion of key concepts in the section [AGK+11a]

Across sections deficiencies

Comprehension burden due to non-sequential presentation of concepts [ACG+12]

Algorithmic enhancement of textbooks for enriching reading experience

References to selective web content

Links to authoritative articles [AGK+10], images [AGK+11b] and videos [ACG+13] based on the focus of the section

References to prerequisites

Links to concepts necessary for understanding the present section, derived using a model of a how students read textbooks [AGK+13]

- Validation on textbooks from U.S.A and India, on different subjects, across grades
- Prototypes and research papers (see [References](#))

Some Data-Centric Research Questions

- Inferring learning units and dependence between them from current educational material (knowledge graph)
- Improvement in educational material based on data on student interactions with the material
- Personalized learning plans
- Dynamic formation of classes and study groups
- Performance evaluation methodologies and benchmarks

Meta Question

- Will we play or cede the space to others?

Data & Education: A Historical Perspective

- Readability Formulas (starting [Lorge 1939])
 - Coefficients of regression equations (e.g. over McCall-Crabbs Standard Test Lessons)

Flesch Reading Ease Score [17]	206.835	–	84.6	×	S/W	–	1.015	×	W/T
Flesch-Kincaid Grade Level [31]	–15.59	+	11.8	×	S/W	+	0.39	×	W/T
Dale-Chall Grade Level [14]	14.862	–	11.42	×	D/W	+	0.0512	×	W/T
Gunning Fog Index [23]			40	×	C/W	+	0.4	×	W/T
SMOG Index [37]	3.0	+	$\sqrt{30}$	×	$\sqrt{C/T}$				
Coleman-Liau Index [10]	–15.8	+	5.88	×	L/W	–	29.59	×	T/W
Automated Readability Index [46]	–21.43	+	4.71	×	L/W	+	0.50	×	W/T

C	=	Number of words with three syllables or more
D	=	Number of words on the Dale Long List
L	=	Number of letters
S	=	Number of syllables
T	=	Number of sentences
W	=	Number of words

- Item Response Theory (starting 1950s, in use in ETS)
 - 1PL Model: $P(\theta_i) = \exp(-(\theta_i - \beta_j)) / (1 + \exp(-(\theta_i - \beta_j)))$
- Intelligent Tutoring Systems (starting [Pressy 1924])
 - Adapt tutoring strategies based on student actions
 - Biennial ITS conferences starting 1988, Also AIED, EDM Confs.

References

(cited in [Enriching Electronic Textbooks slide](#))

[AGK+10] Rakesh Agrawal, Sreenivas Gollapudi, Krishnaram Kenthapadi, Nitish Srivastava, Raja Velu. "[Enriching Textbooks Through Data Mining](#)". [DEV 2010](#).

[AGK+11a] Rakesh Agrawal, Sreenivas Gollapudi, Anitha Kannan, Krishnaram Kenthapadi. "[Identifying Enrichment Candidates in Textbooks](#)". [WWW 2011](#).

[AGK+11b] Rakesh Agrawal, Sreenivas Gollapudi, Anitha Kannan, Krishnaram Kenthapadi. "[Enriching Textbooks With Images](#)". [CIKM 2011](#).

[ACG+12] Rakesh Agrawal, Sunandan Chakraborty, Sreenivas Gollapudi, Anitha Kannan, Krishnaram Kenthapadi. "[Empowering Authors to Diagnose Comprehension Burden in Textbooks](#)". [KDD 2012](#).

[AGK+13] Rakesh Agrawal, Sreenivas Gollapudi, Anitha Kannan, Krishnaram Kenthapadi. "[Studying from Electronic Textbooks](#)". [CIKM 2013](#).

[ACG+13] Rakesh Agrawal, Maria Christoforaki, Sreenivas Gollapudi, Anitha Kannan, Krishnaram Kenthapadi, Adith Swaminathan. "[Augmenting Textbooks with Videos](#)". Working Paper. 2013.