# Big Education in the Era of Big Data

#### Irwin King

Associate Dean (Education), Faculty of Engineering, CUHK Director, Shenzhen Rich Media and Big Data Analytics and Application Key Lab, SZRI, CUHK PI, The Knowledge and Education Exchange Platform (KEEP), CUHK

> Department of Computer Science and Engineering The Chinese University of Hong Kong

> > king@cse.cuhk.edu.hk http://www.cse.cuhk.edu.hk/~king

©2014 Irwin King. All rights reserved.

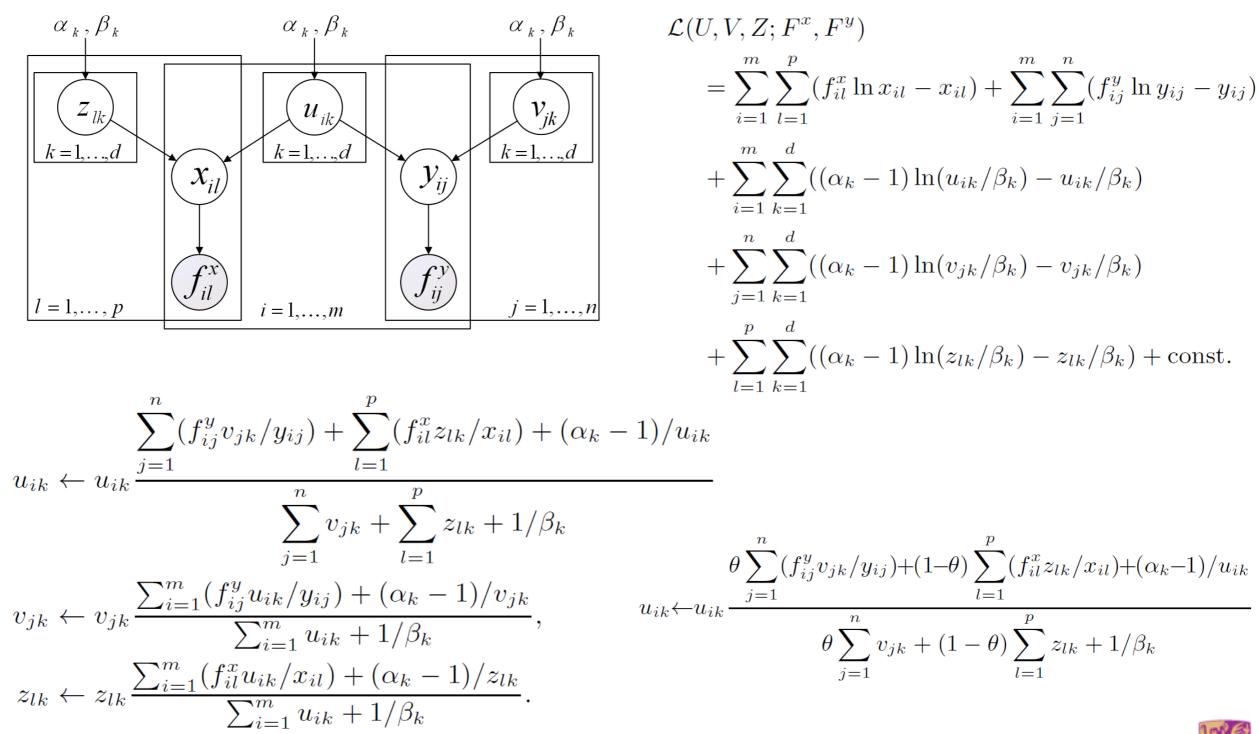


#### Wlodzislaw Duch

#### Under-Secretary of State Ministry of Science & Higher Education



### **Collective Probabilistic Factor Model**





The grass is greener on the other side... Be inspired!

### Stories and more stories...

### Be informed!

The devil is in the details...

Be challenged!



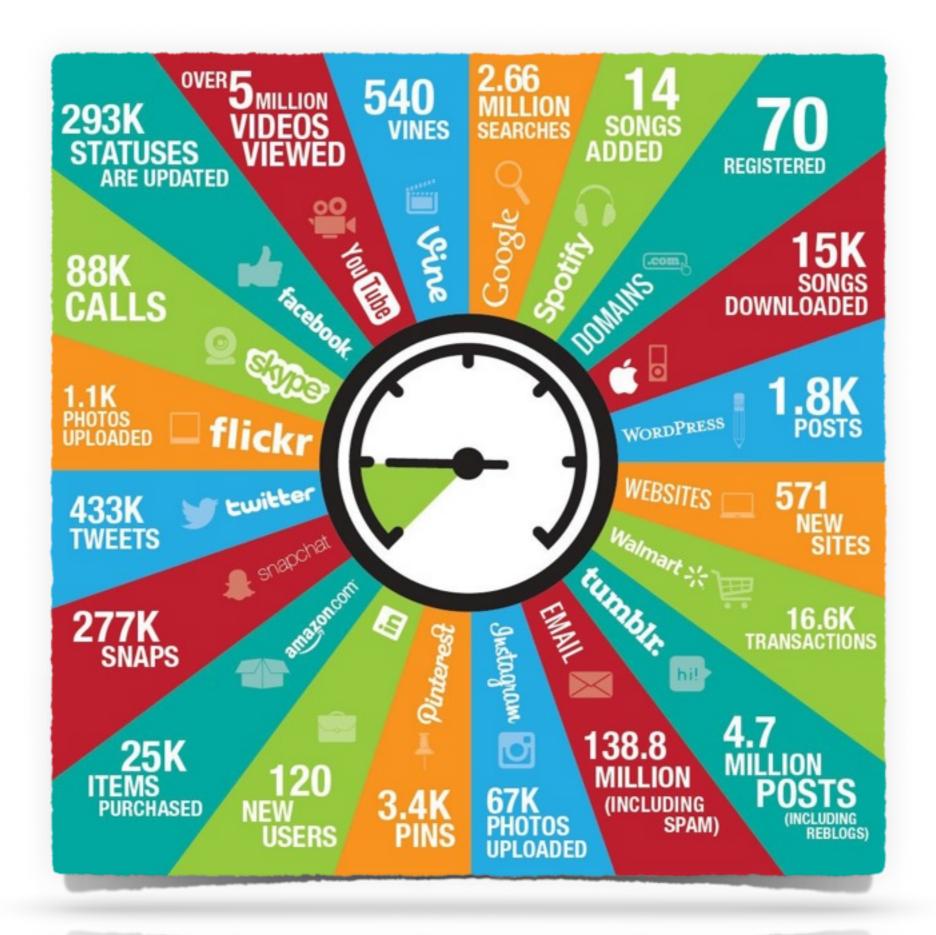
### Words of Wisdom

#### The **BEST** universities focus on **EDUCATION**!

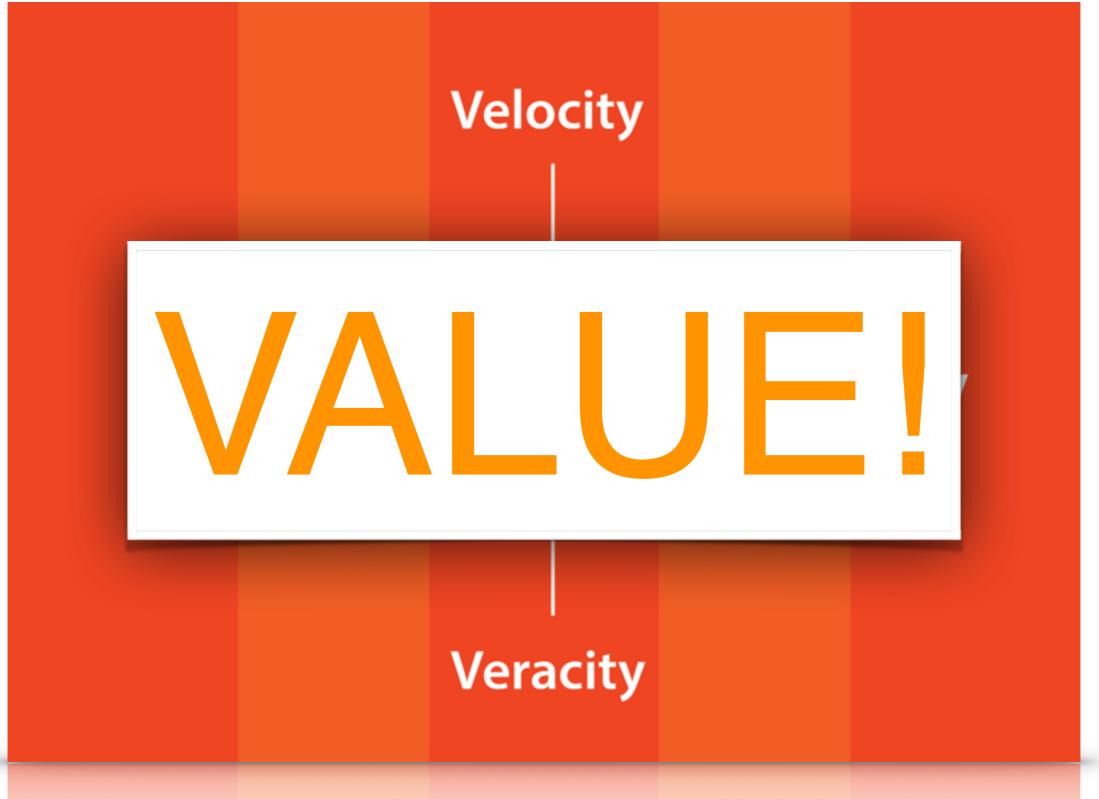
# The **BETTER** universities focus on citation numbers and impact factors...

# The **GOOD** universities focus on counting the number of publications...





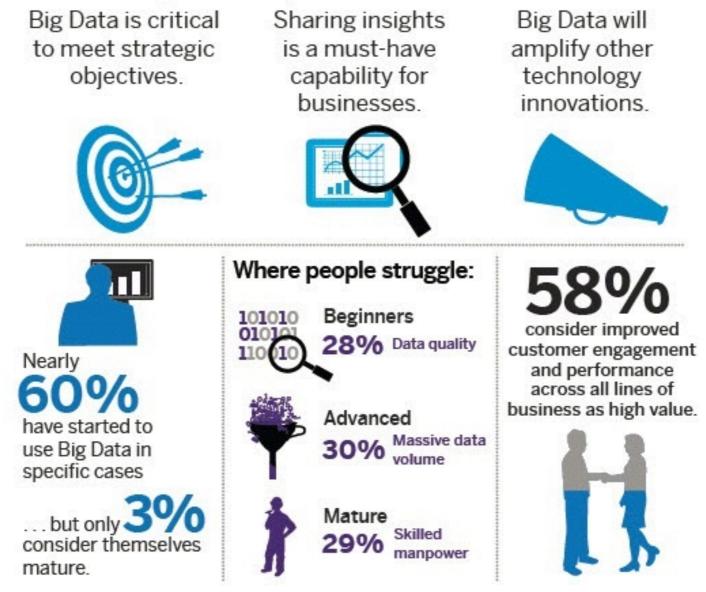






# The Value of Big Data

Over 80% of organizations say:

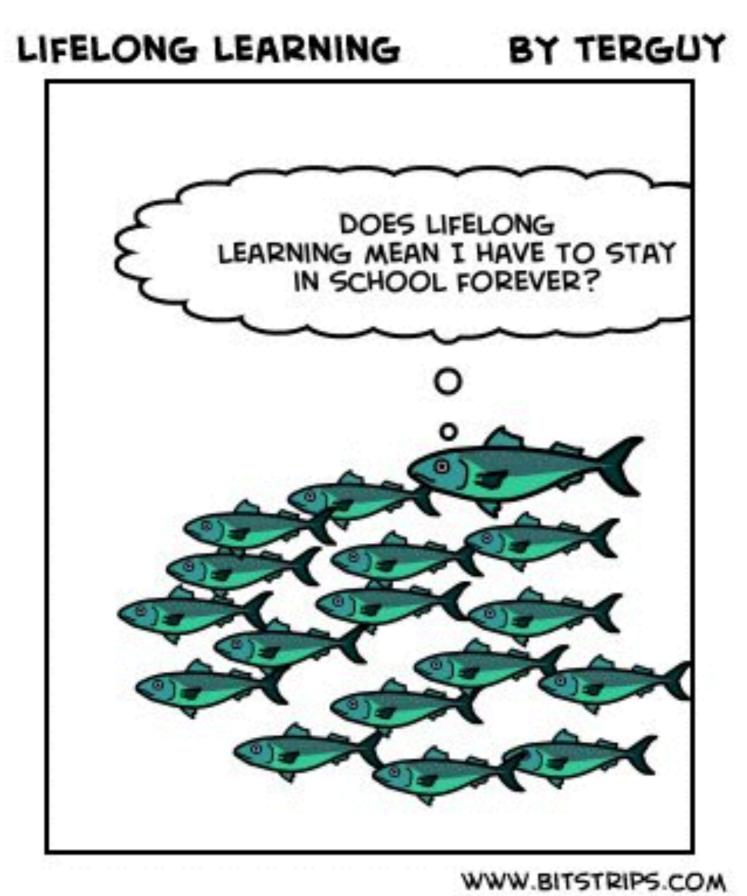




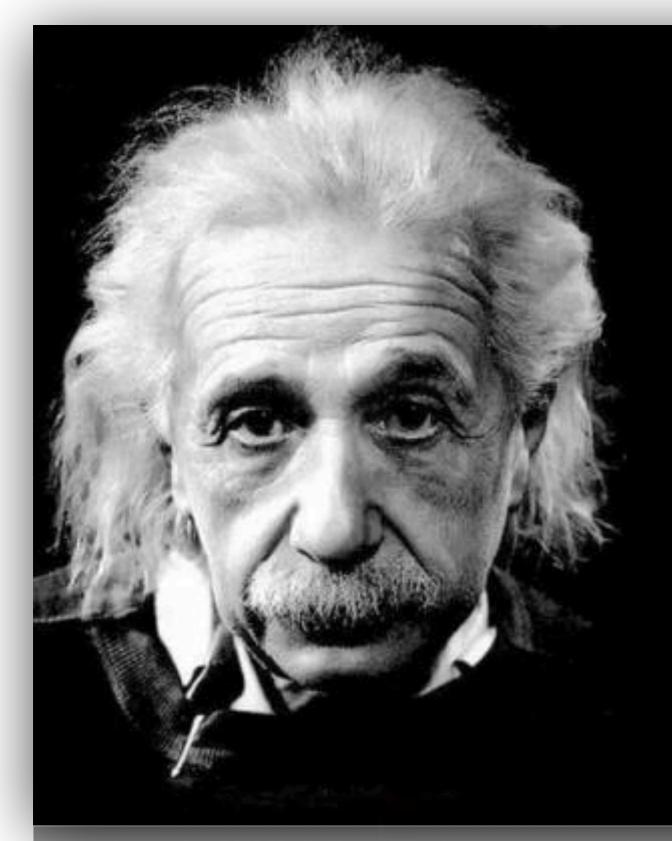
## **Big Education on Lifelong Learning**











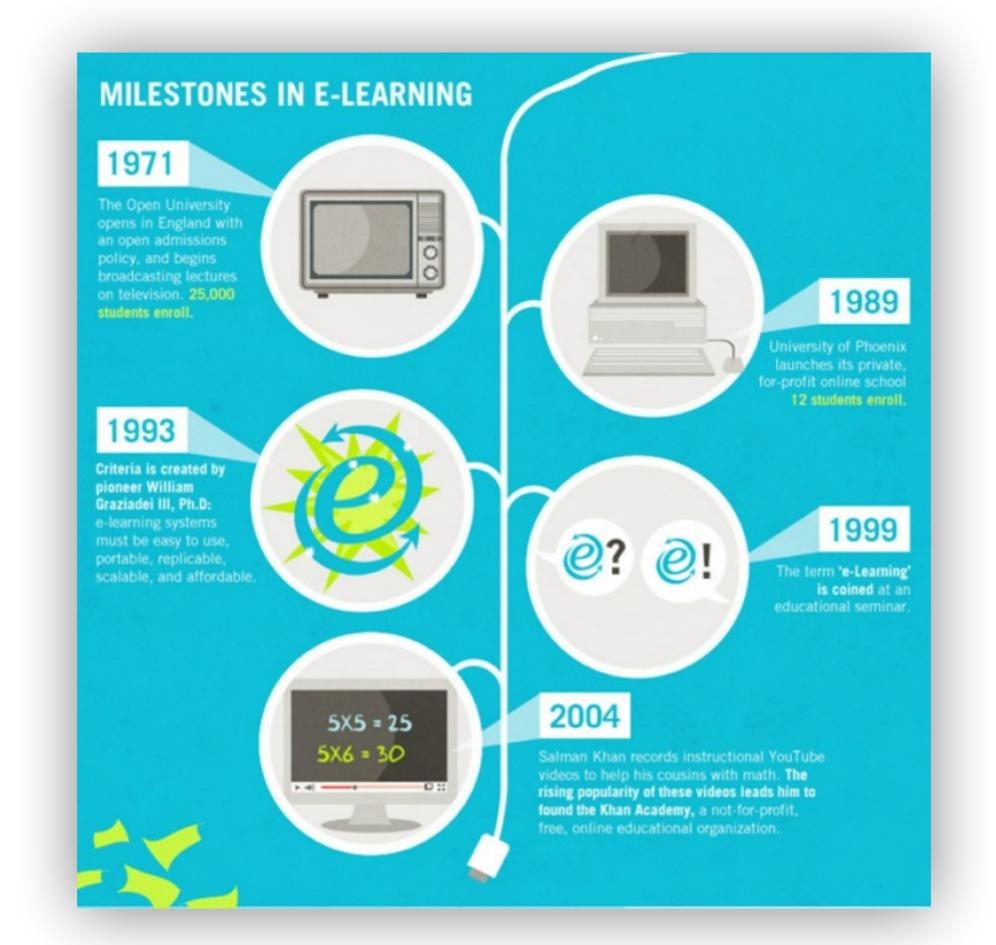
### Once you stop learning, you start dying...

Albert Einstein

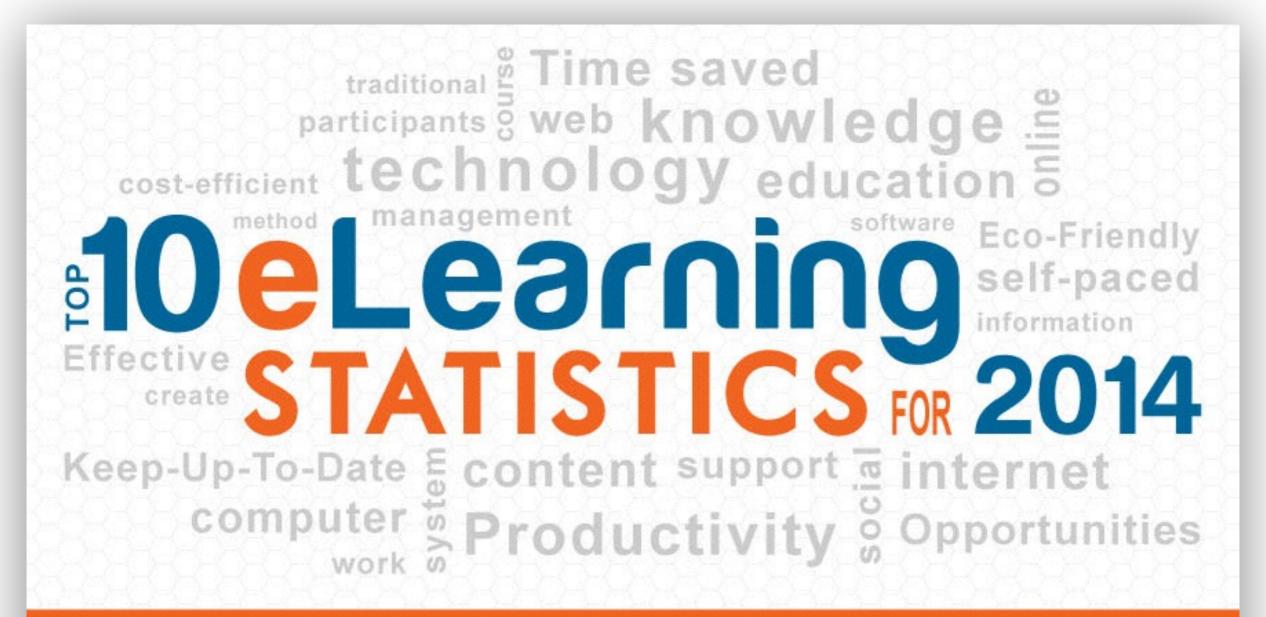






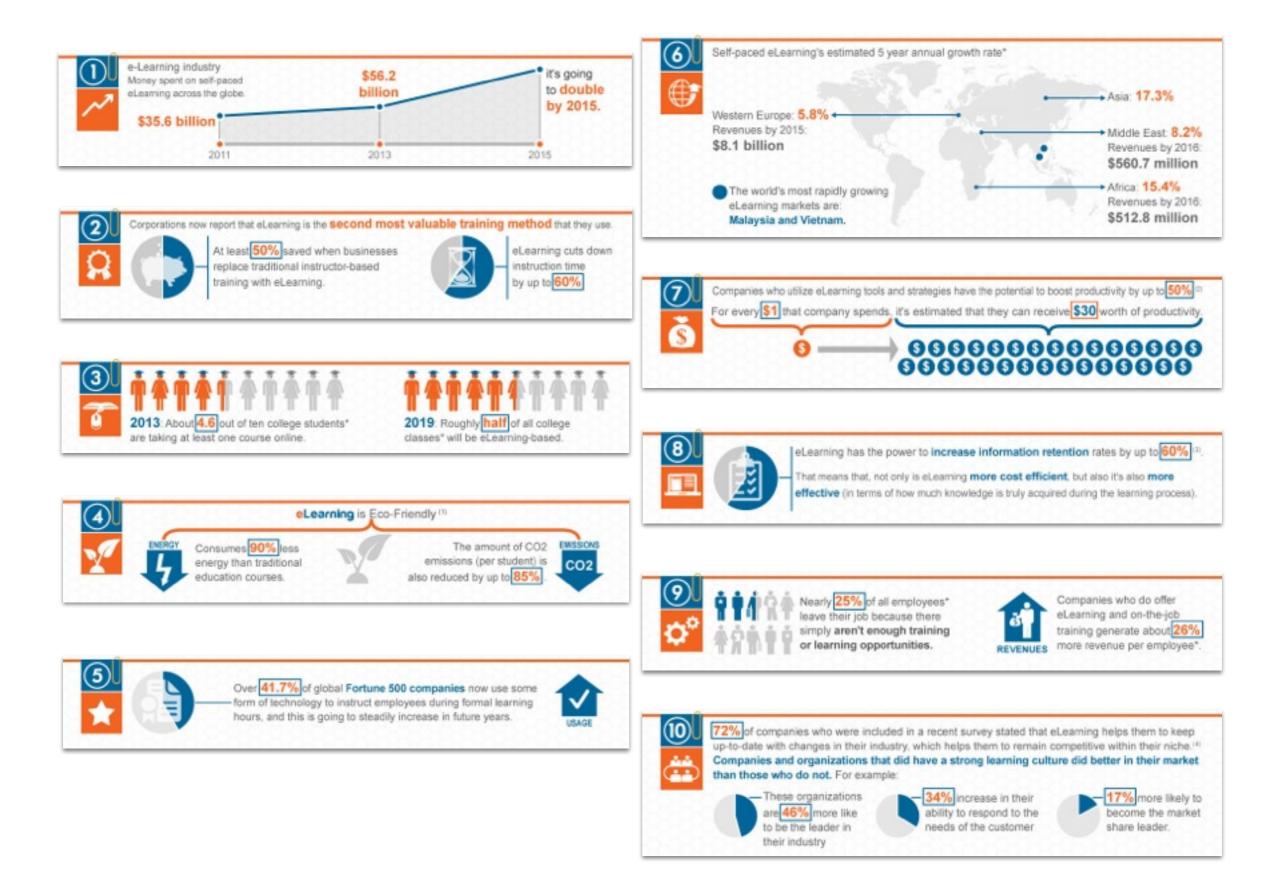






The rise in eLearning's popularity isn't showing any signs of slowing. In fact, judging by the following Top 10 eLearning statistics for 2014, the future of the eLearning Industry is brighter than ever:





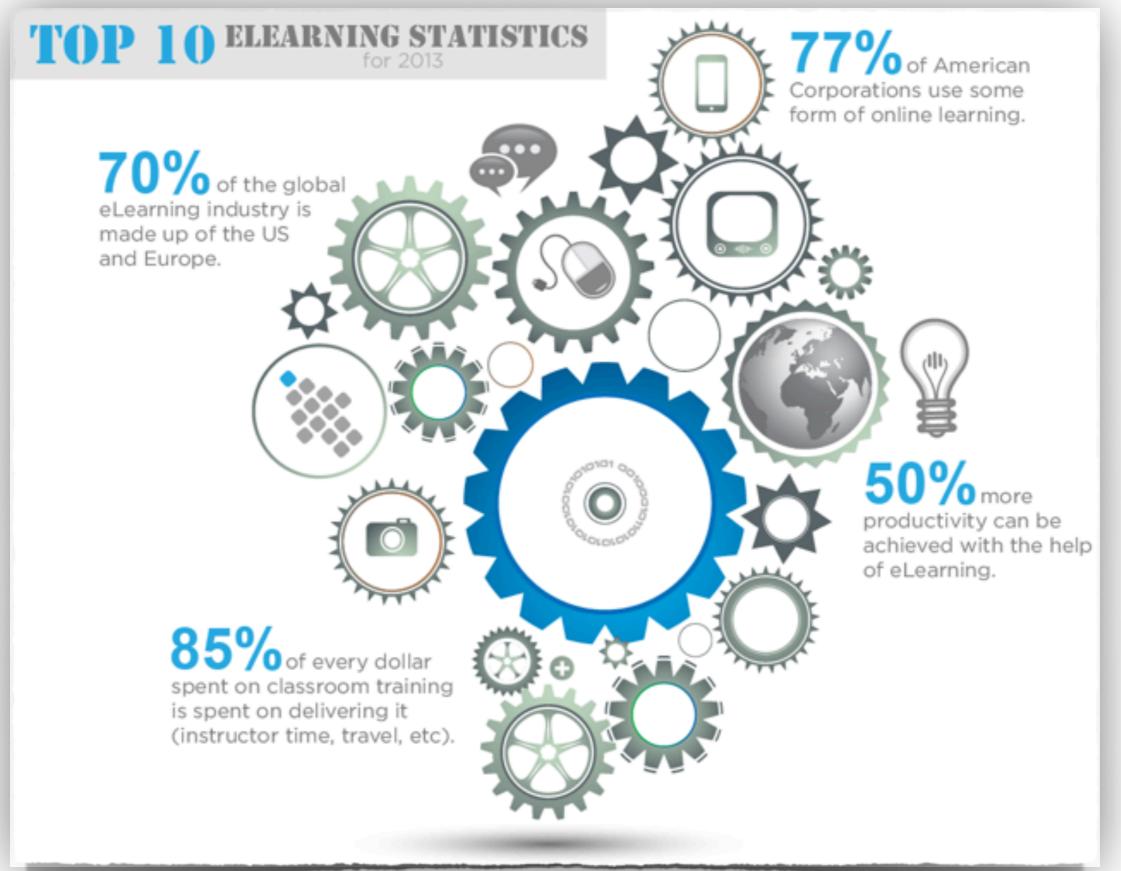








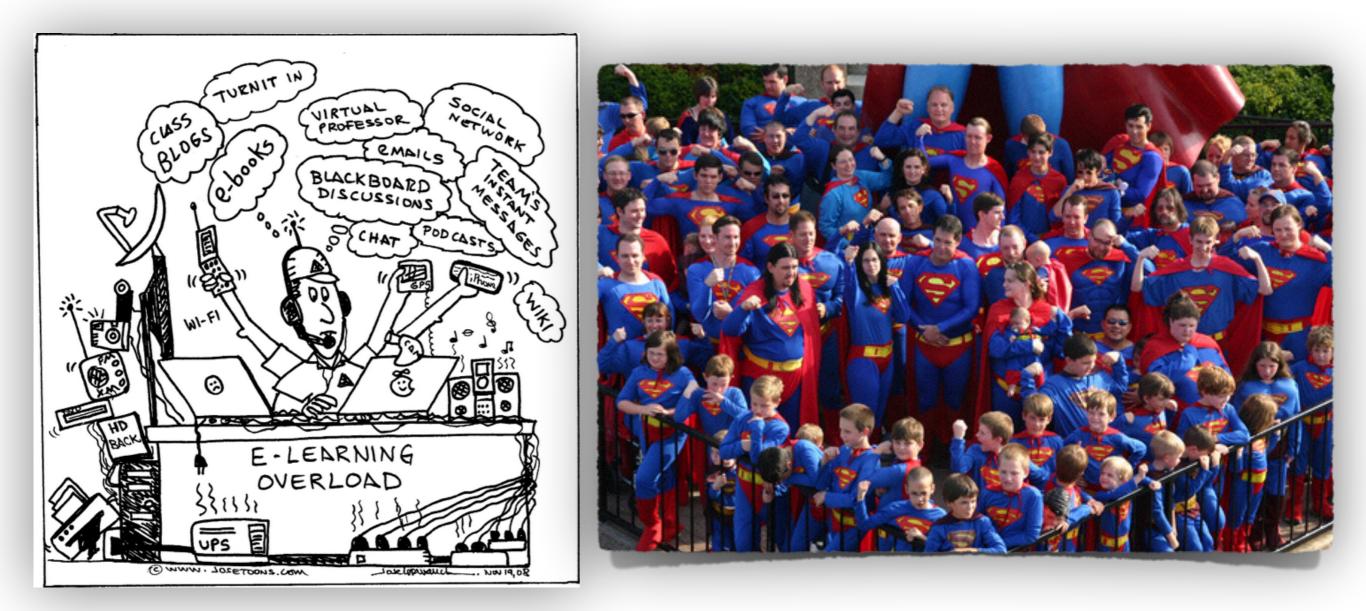




Big Education in the Era of Big Data @ FedCSIS 2014, September 7-10, 2014, Warsaw, Poland







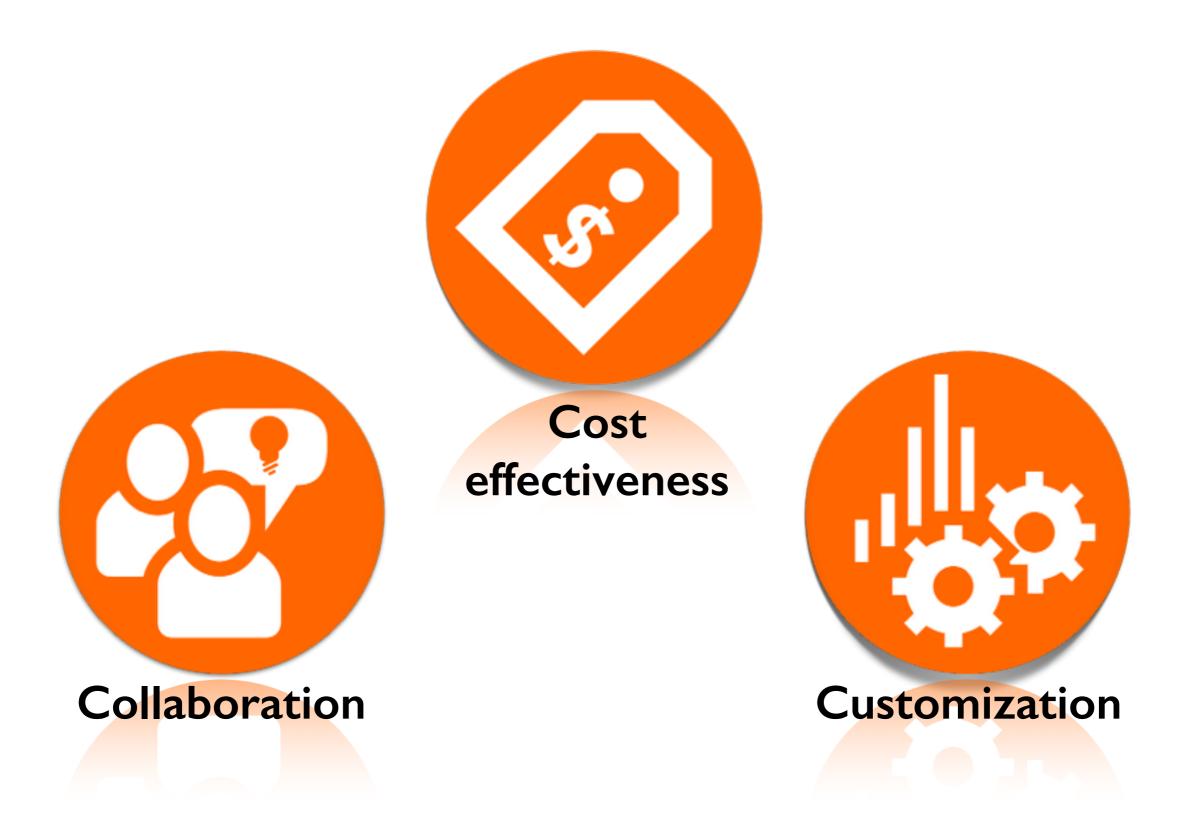
# The task of the modern educator is not to cut down jungles, but to irrigate deserts.

#### C.S. Lewis



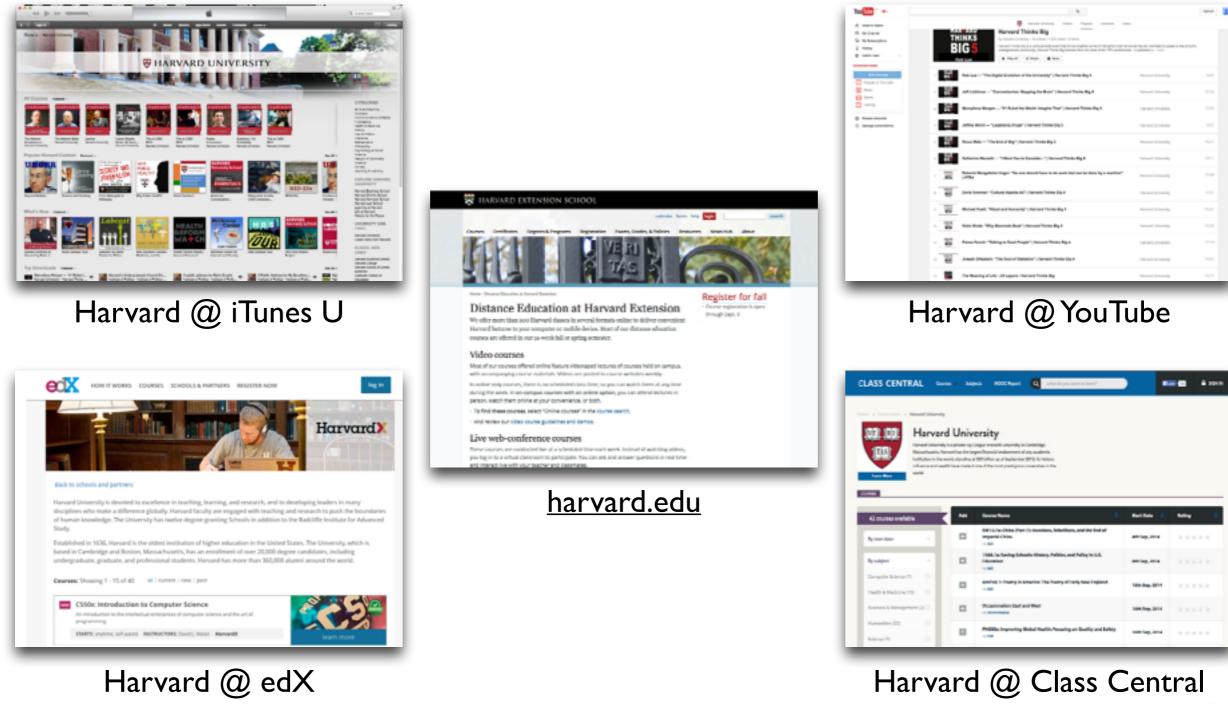
# Trends in Big Education







### Multimodal Learning

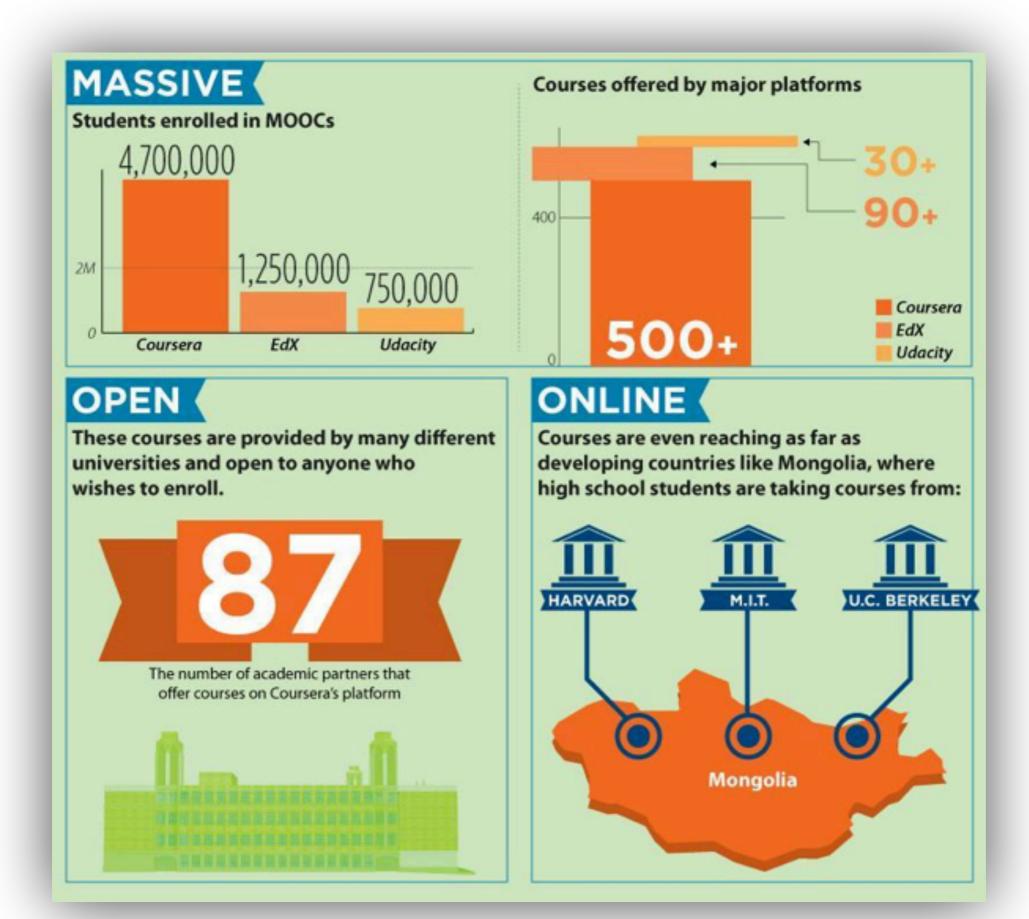




# MOOC

# Massive Open Online Course

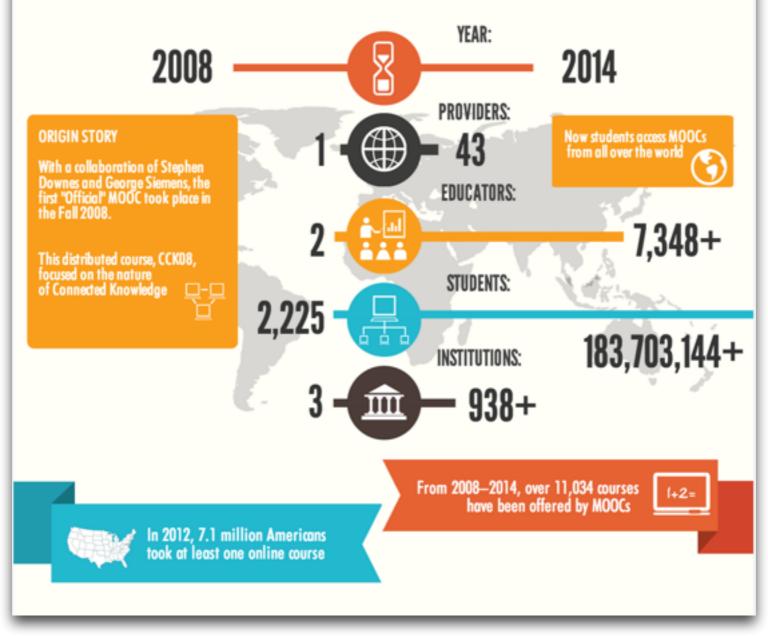


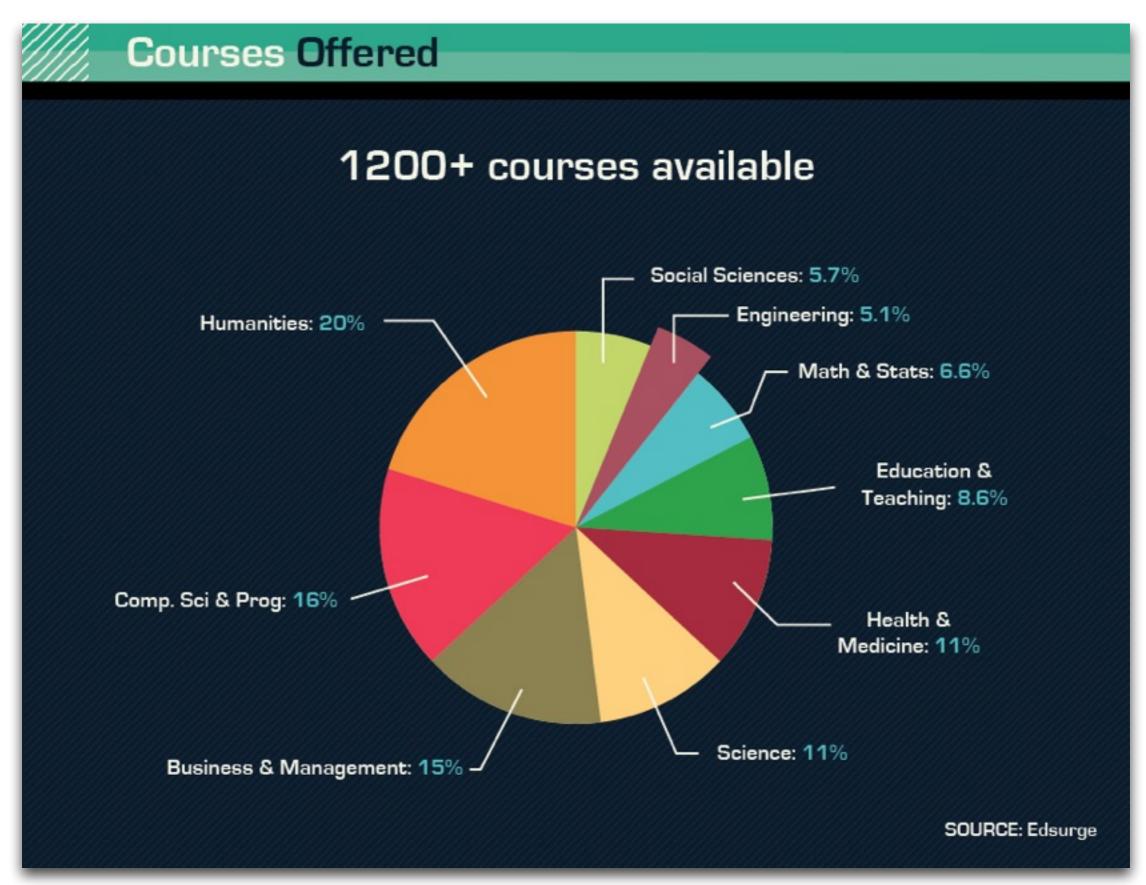






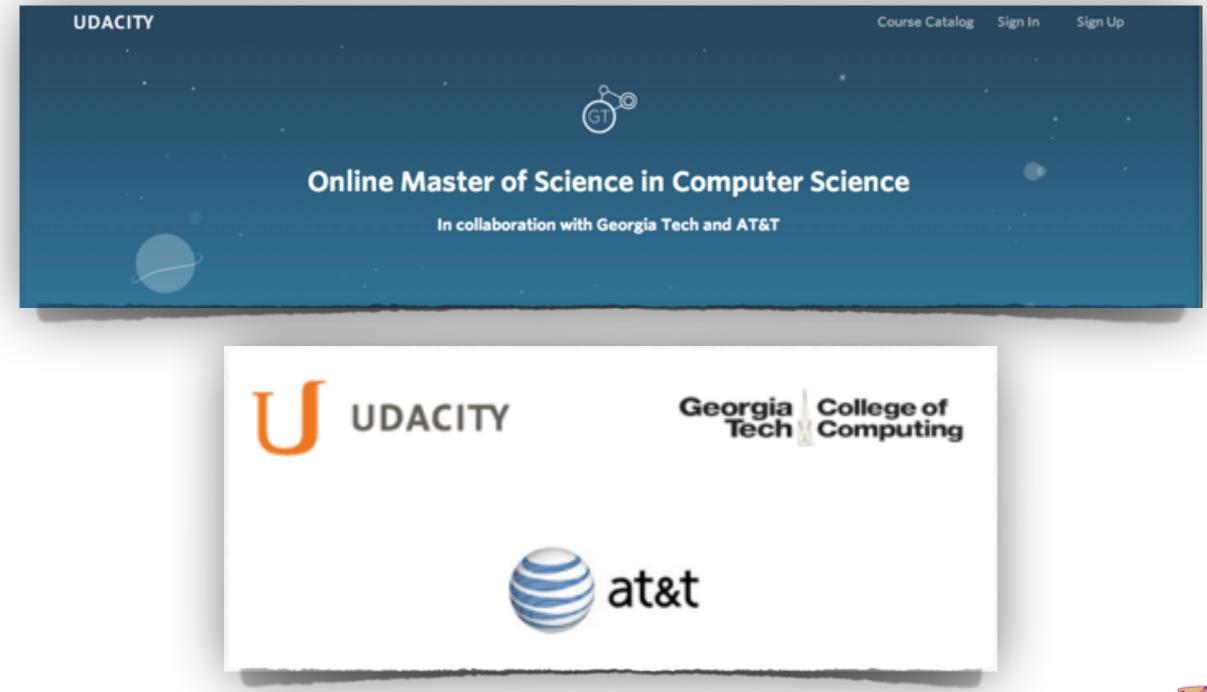
#### OVERVIEW OF MASSIVE OPEN ONLINE COURSES'





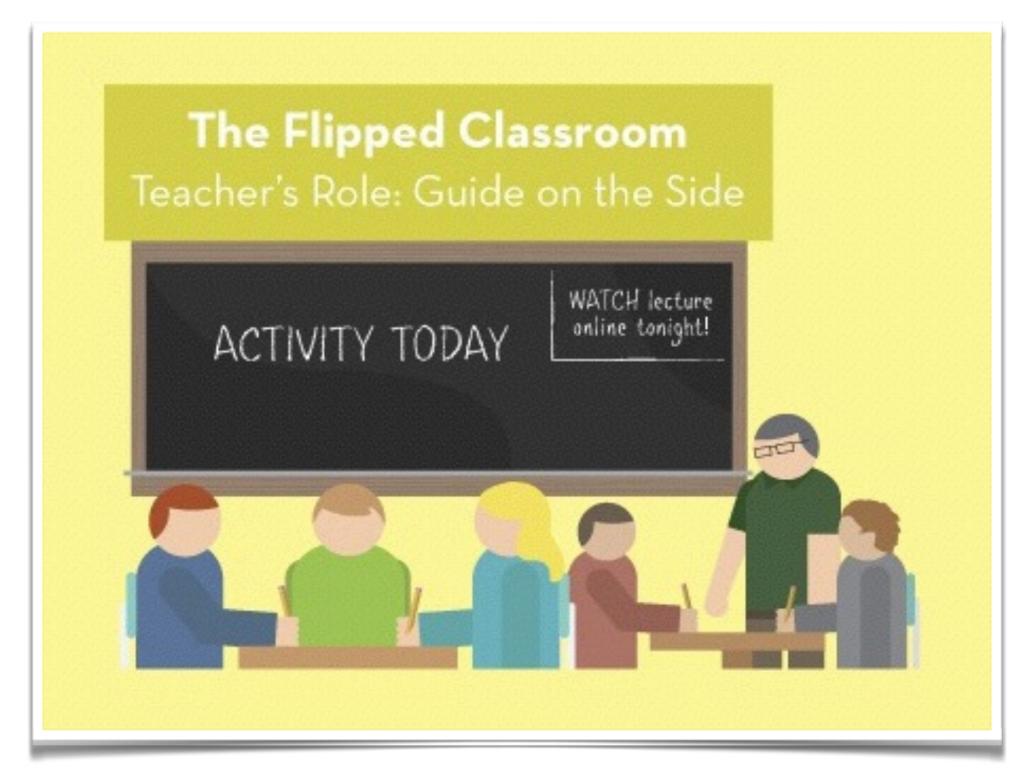


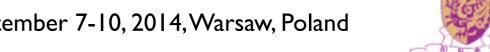
# Small Private Online Course (SPOC) with Degree





### Flipped Classroom

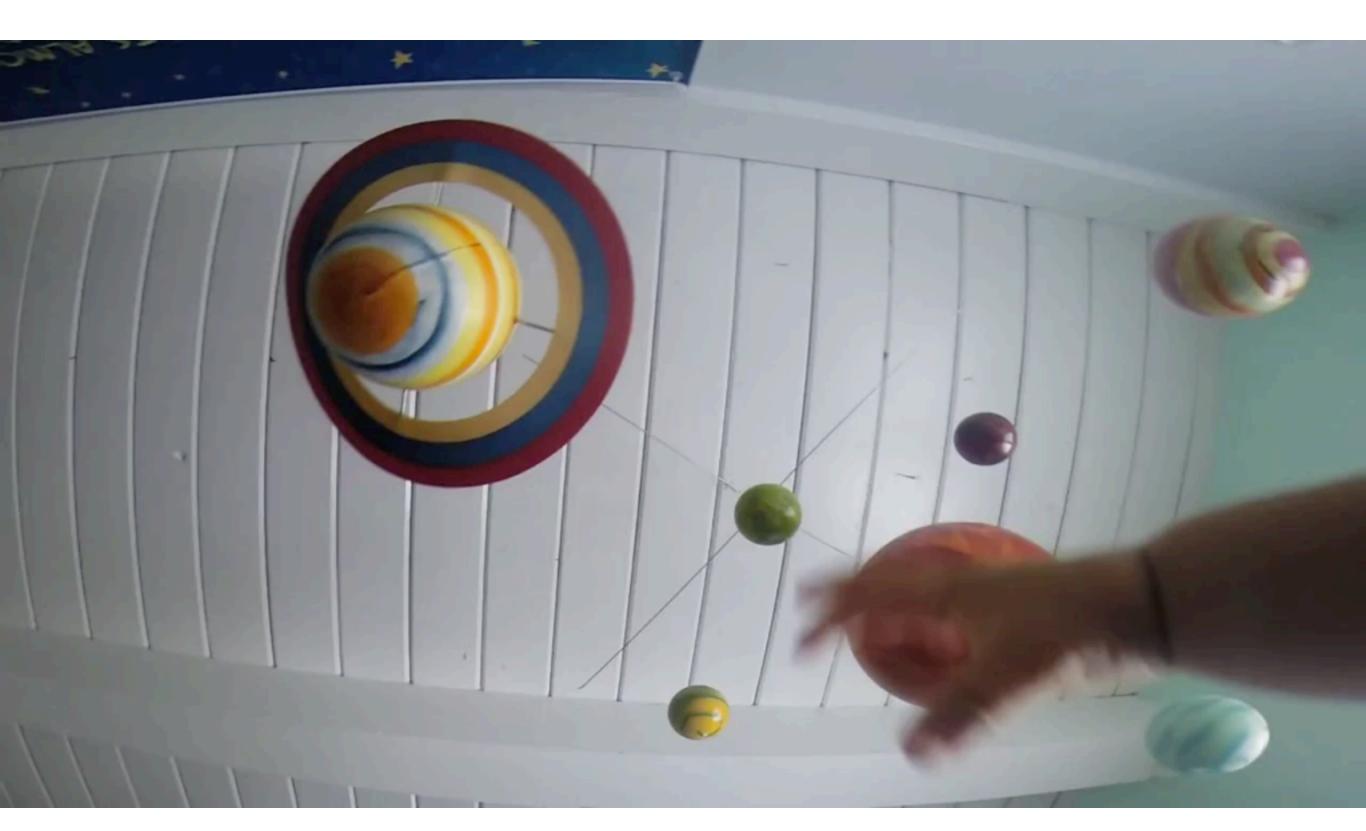




### Microlearning

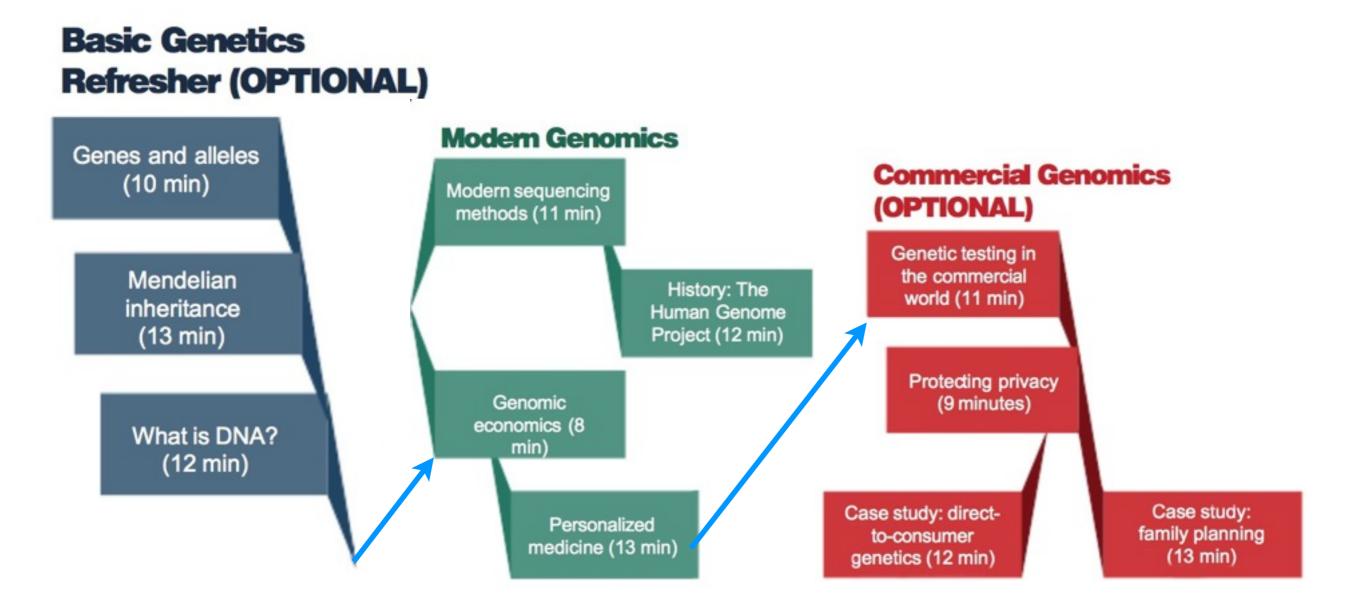
KHANACADEMY Subject	t: Computer pro 👻	Coach	About	Donate	Q. Search for subje	cts, skills, ar	nd videos	Log in Sign up
COMPUTER PROGRAMMING			ALL COM	NTENT IN "INT	TRO TO JS: DRAWING & A	NIMATION"		
Intro to JS: Drawing & Animation			Intro to programming			What is Programming?		
			If you've never been here before, check out this introductory video first. Then get coding!			A Tour of Programming on Khan Academy		
			Drawing basics			P	Intro to Drawing	
+ Create Program			We'll show you the basics of programming and how to draw shapes.		¢	Challenge: H for Hopper		
					¢	More Drawing!		
					¢	Challenge: Simple Shapes!		
? Help Requests						()	Challenge: CRAZY Face	
Project Evaluations			Coloring			P	Intro to Coloring	
Community Questions			We'll show you how to color and outline your shapes!		Ģ	Challenge: Ice Cream Code		
						¢	Challenge: It's a Beautiful Day	
					Ø	The Power of the Docs		
						9	Project: What's for Dinner?	
			Variab	les		Ð	Intro to Variables	
			We'll co	ver how to u	se variables to hold			





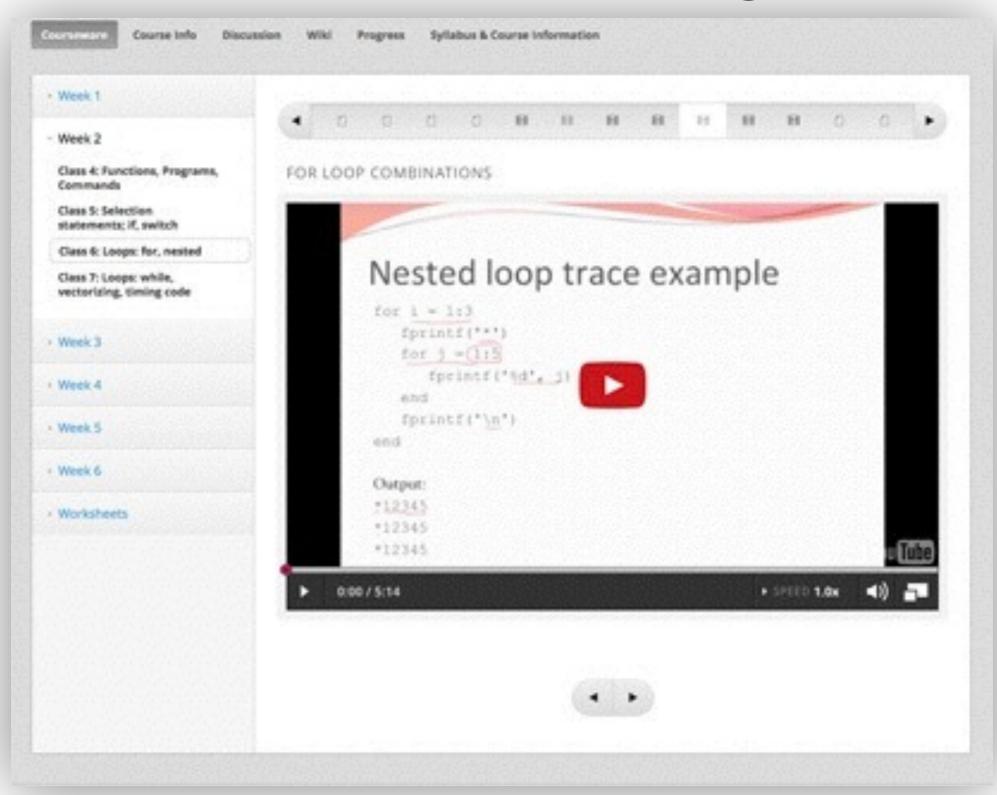


### Personalized Learning



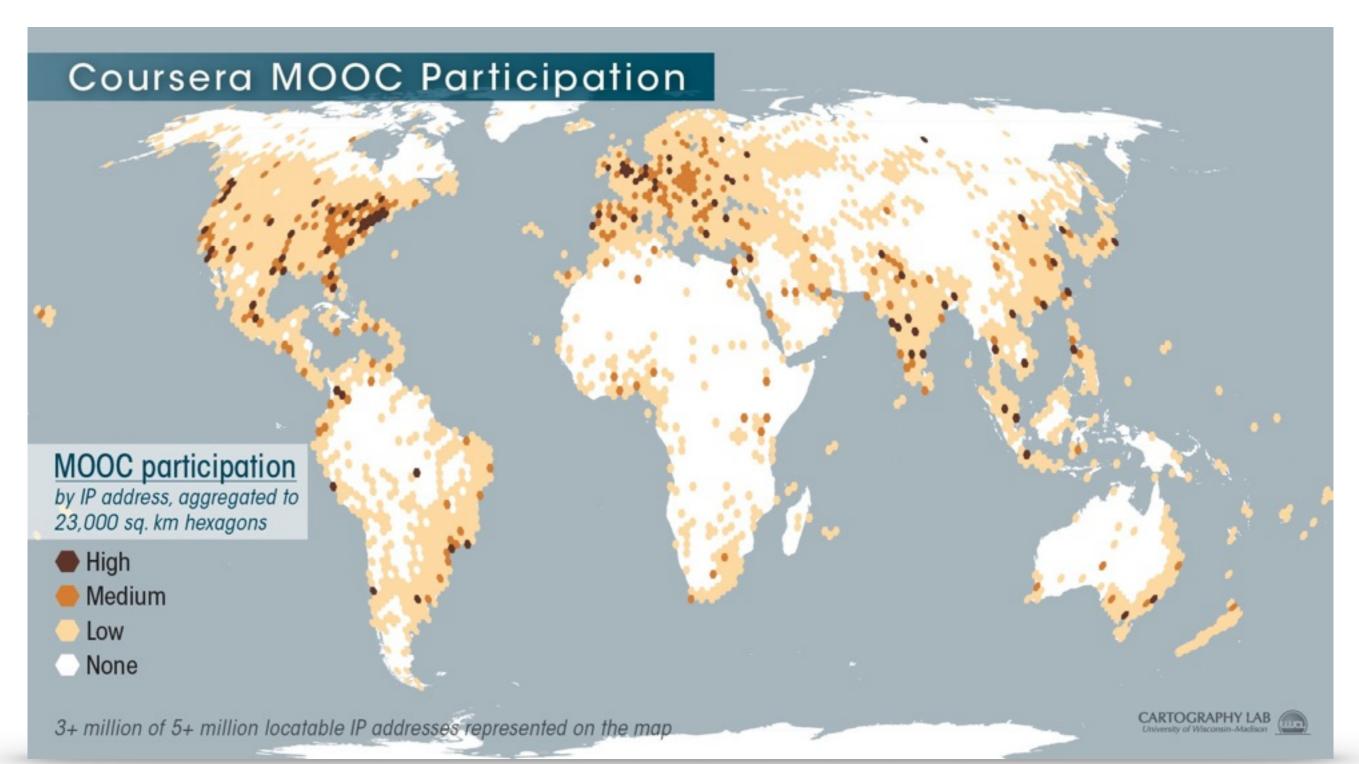


### Active Learning





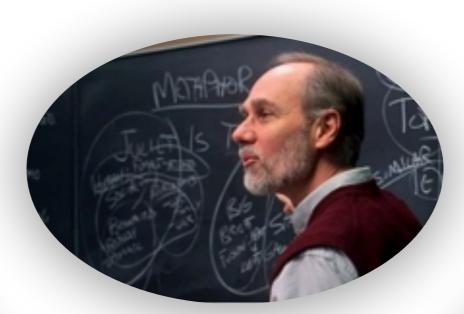
## Peer Learning





# Who Really Cares?







Instructors

Big Education Stakeholders



#### Administrators











Save Money





Keep Focused



Test Mastery

### Students



Get Quick Feedback

Big Education in the Era of Big Data @ FedCSIS 2014, September 7-10, 2014, Warsaw, Poland



Interactive and Collaboration



Access on Multiple Devices



Access from Anywhere





Release heavy teaching workload



Focus on interaction with students in the classroom

### Instructors

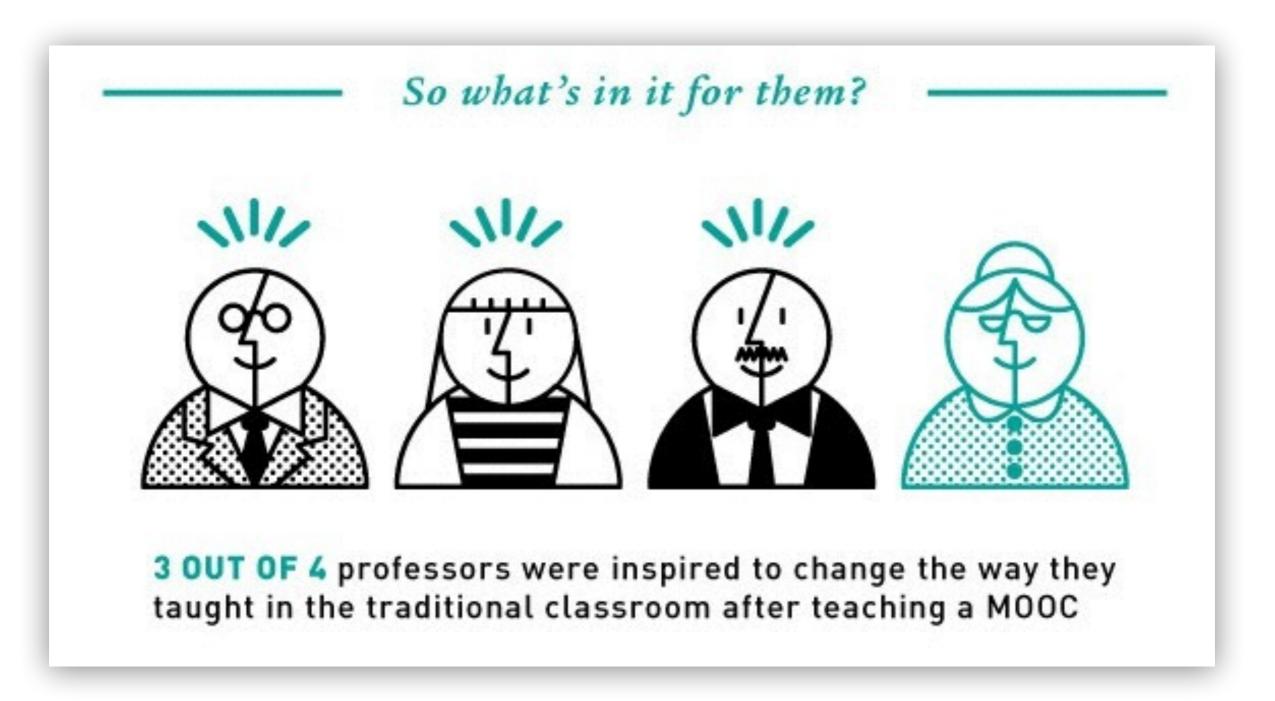




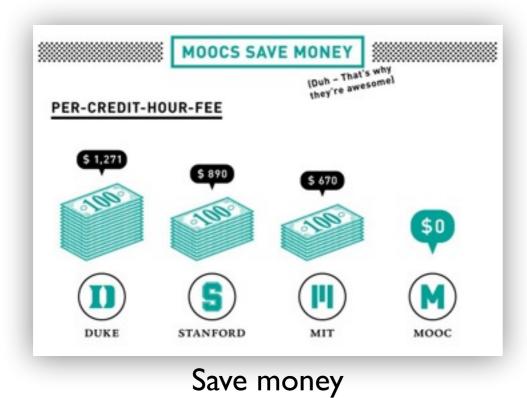
Provide personalized helpTrack student performanceBig Education in the Era of Big Data @ FedCSIS 2014, September 7-10, 2014, Warsaw, Poland

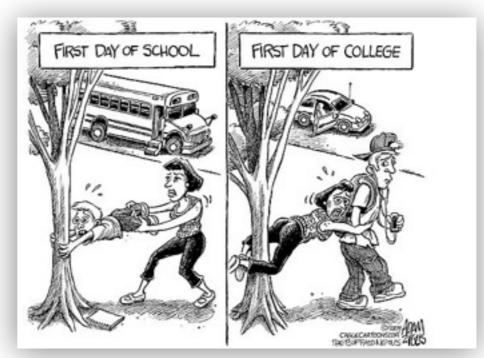


## Faculty





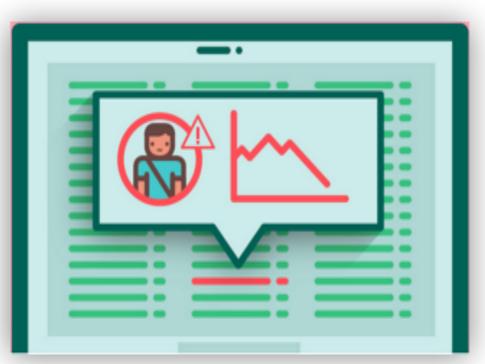




Improve engagement rate

### Administrators





Reduce failure rateTrack student performanceBig Education in the Era of Big Data @ FedCSIS 2014, September 7-10, 2014, Warsaw, Poland



## Case Studies Predict Student Performance





## Case Study I: Cortez P, Silva A M G. Using data mining to predict secondary school student performance[J]. 2008.

• Input

demographic information	sex, age, school, address, habit, health status, parents' education, job, family size, income
social information	romantic relationship, free time after school, going out with friends, weekend/workday alcohol consumption
school related information	study time, past failure course, extra paid support, family support, free time after school, previous course grade

#### • Output

- Binary classification (pass/fail)
- 5-Level classification (from I very good / excellent to V insufficient)
- Regression, with a numeric output that range between zero (0%) and twenty (100%)



Input	Mathematics					Portuguese				
Setup	NV	NN	SVM	DT	$\mathbf{RF}$	NV	NN	SVM	DT	$\mathbf{RF}$
Α	$91.9^{+1.0}$	$88.3_{\pm 0.7}$	$86.3_{\pm 0.6}$	$90.7_{\pm 0.3}$	$91.2_{\pm 0.2}$	$89.7_{\pm 0.0}$	$90.7_{\pm 0.5}$	91.4±0.2	93.0 <sup>†</sup> =0.3	92.6±0.1
в	83.81±0.0	$81.3_{\pm 0.5}$	$80.5_{\pm 0.5}$	$83.1_{\pm 0.5}$	$83.0_{\pm 0.4}$	$87.5_{\pm 0.0}$	$87.6_{\pm 0.4}$	88.0±0.3	$88.4_{\pm 0.3}$	$90.1^{+\pm0.2}$
$\mathbf{C}$	67.1±0.0	$66.3_{\pm 1.0}$	$70.6_{\pm 0.4}$	$65.3_{\pm 0.8}$	$70.5_{\pm 0.5}$	$84.6_{\pm 0.0}$	$83.4_{\pm 0.5}$	$84.8_{\pm 0.3}$	$84.4 \pm 0.4$	$85.0 \cdot \pm 0.2$

Table 3: Binary classification results (PCC values, in %; <u>underline</u> – best model; **bold** – best within the input setup)

<sup>†</sup> – statistical significance under pairwise comparisons with other methods.

\* - statistical significance under a pairwise comparison with NV.

Table 4: Five-level classification results (PCC values, in %; <u>underline</u> – best model; **bold** – best within the input setup)

Input	Mathematics					Portuguese				
Setup	INV	NN	SVM	DT	$\mathbf{RF}$	NV	NN	SVM	DT	$\mathbf{RF}$
А	$78.5^{+\pm0.0}$	$60.3_{\pm 1.6}$	$59.6_{\pm 0.9}$	$76.7_{\pm 0.4}$	$72.4_{\pm 0.4}$	72.9±0.0	65.1±0.9	$64.5_{\pm 0.6}$	<u>76.1</u> ±0.1	$73.5_{\pm 0.2}$
в	$69.5_{\pm 0.0}$	$49.8_{\pm 1.2}$	$47.9_{\pm 0.7}$	$57.5 \pm 0.8$	$52.7_{\pm 0.6}$	58.7±0.0	$52.0_{\pm 0.6}$	$51.7_{\pm 0.6}$	$\textbf{62.9}_{\pm 0.2}$	$55.3_{\pm 0.4}$
С	$32.9 \pm 0.0$	$30.4_{\pm 1.0}$	$31.0{\scriptstyle \pm 0.7}$	$31.5_{\pm 0.6}$	$33.5_{\pm 0.6}$	$31.0_{\pm 0.0}$	$33.7_{\pm 0.6}$	$34.9_{\pm 0.5}$	$32.8_{\pm 0.6}$	$36.7^{\dagger}_{\pm0.6}$

 $\dagger$  – statistical significance under pairwise comparisons with other methods.

Table 5: Regression results (RMSE values; <u>underline</u> – best model; **bold** – best within the input setup)

I.		N	Iathemat	ics		Portuguese				
S.	NV	NN	SVM	DT	RF	NV	NN	SVM	DT	RF
Α	$2.01_{\pm 0.00}$	$2.05_{\pm 0.02}$	$2.09_{\pm 0.02}$	$1.94_{\pm 0.04}$	$1.75^{+0.01}$	$1.32_{\pm 0.00}$	$1.36_{\pm 0.04}$	$1.35_{\pm 0.01}$	$1.46 \pm 0.03$	$1.32_{\pm 0.00}$
В	$2.80_{\pm 0.00}$	$2.82_{\pm 0.02}$	$2.90_{\pm 0.02}$	$2.67_{\pm 0.04}$	$2.46^{\dagger}{\scriptstyle \pm 0.01}$	$1.89 \pm 0.00$	$1.88 \pm 0.02$	$1.87_{\pm 0.01}$	$1.78 \cdot \pm 0.03$	$1.79_{\pm 0.01}$
$\mathbf{C}$	$4.59_{\pm 0.00}$	$4.41_{\pm 0.03}$	$4.37_{\pm 0.03}$	$4.46{\scriptstyle \pm 0.04}$	$\mathbf{3.90^{\dagger}}_{\pm 0.01}$	$3.23 \pm 0.00$	$2.79_{\pm 0.02}$	$2.76_{\pm 0.02}$	$2.93_{\pm 0.02}$	$2.67^{+\pm0.01}$

<sup>†</sup> – statistical significance under pairwise comparisons with other methods.

\* - statistical significance under a pairwise comparison with NV.

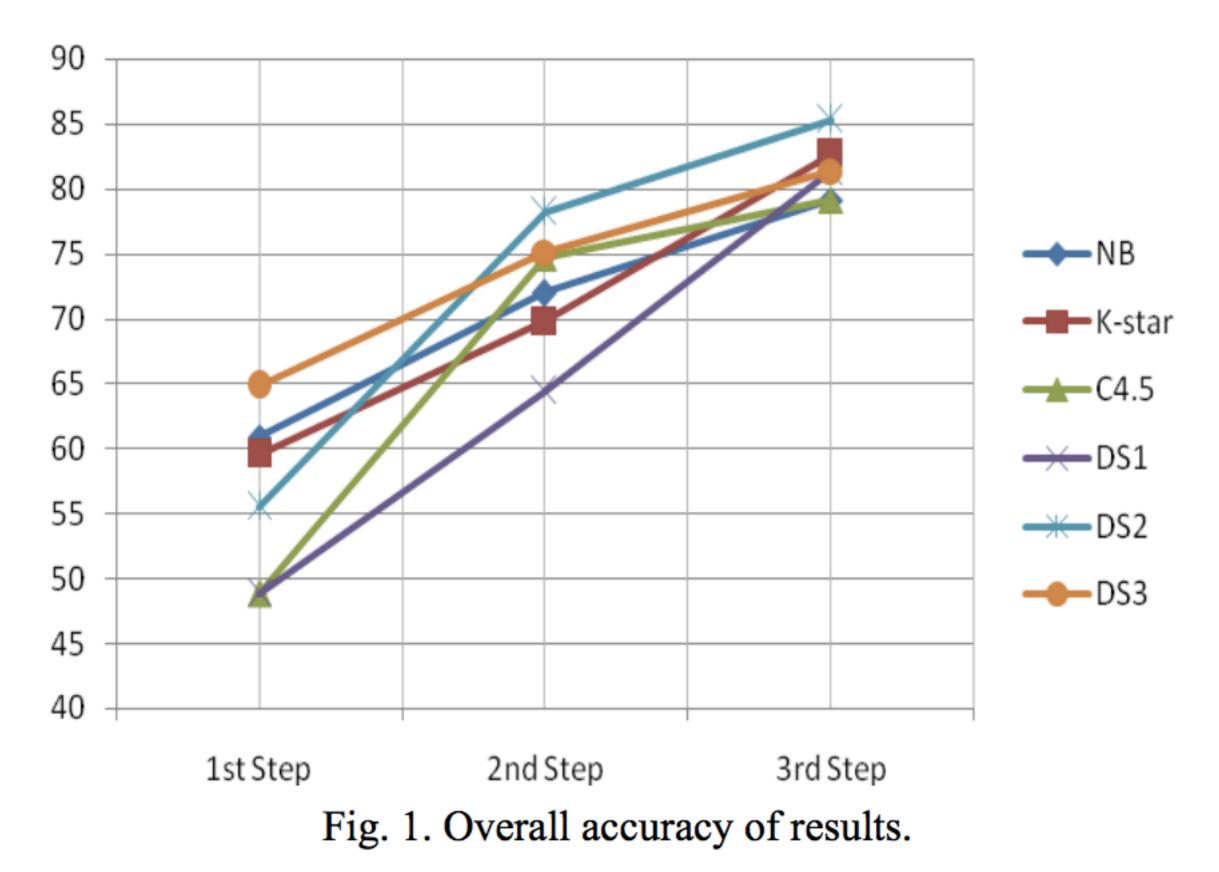


Case Study 3: Erkan E. Identifying At-Risk Students Using Machine Learning Techniques: A Case Study with IS 100. 2012.

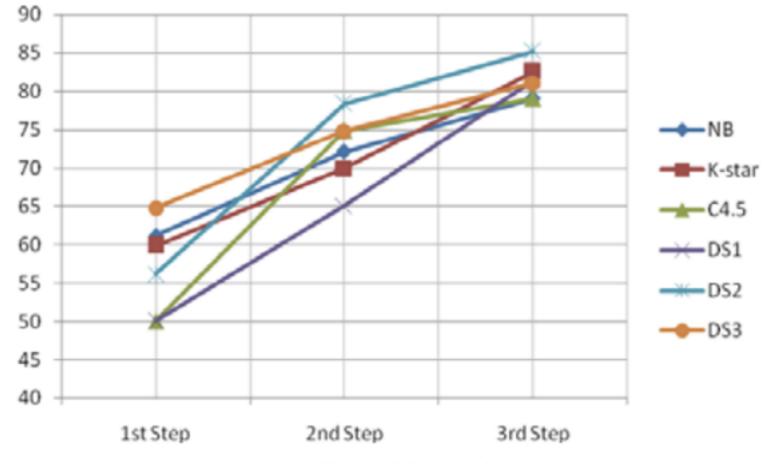
- Use three machine learning algorithms(instance-based learning Classifier, Decision Tree, and Naive Bayes) to predict students' performance
  - Three steps
    - Ist step: Attendance information for first four weeks, grade of Ist assignment,
    - 2nd step:Attendance information for first seven weeks, grade of Ist, 2nd assignments, midterm grade
    - 3rd step: Attendance information for first ten weeks, grade of 1st,
      2nd and 3rd assignments, final exam grade, midterm grade.

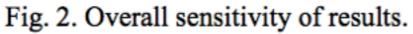
Case Study 3: Erkan E. Identifying At-Risk Students Using Machine Learning Techniques: A Case Study with IS 100. 2012.

- Algorithms
  - K-Star
    - One of the instance-based classifiers
  - C4.5
    - An extension to ID3 algorithm
  - Naive Bayes
  - Three decision schemas
    - DSI: if at least one of the algorithms classifies student as a failure than this student will be considered as failure
    - DS2: if at least two algorithms classify student as a failure than this student will be considered as failure
    - DS3: if all three algorithms classify student as a failure than this student will be considered as failure









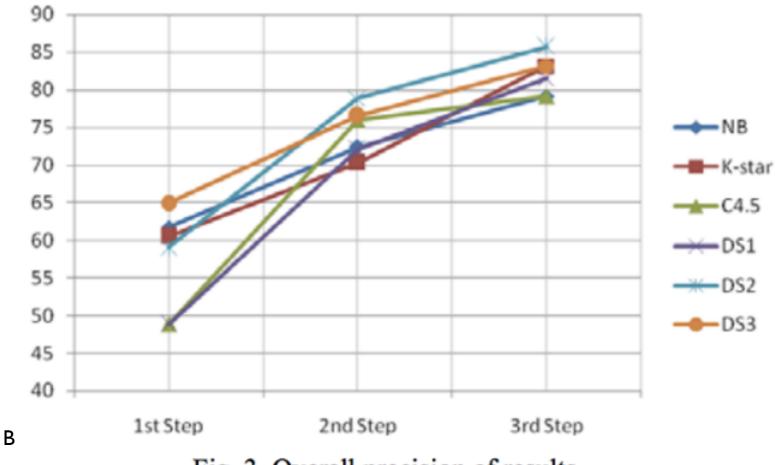


Fig. 3. Overall precision of results.

d

# KEEP Knowledge and Education Exchange Platform



## **KEEP Education Cloud**

- Educational resources to anyone, anytime, anywhere, on any device
- An education cloud platform to provide aggregated eLearning resources for teachers and students
- Big Data analytics for education
- Knowledge aggregation and technology integration!
- Multi-year, multi-discipline, and cross-institutional project with strong partners and alliances

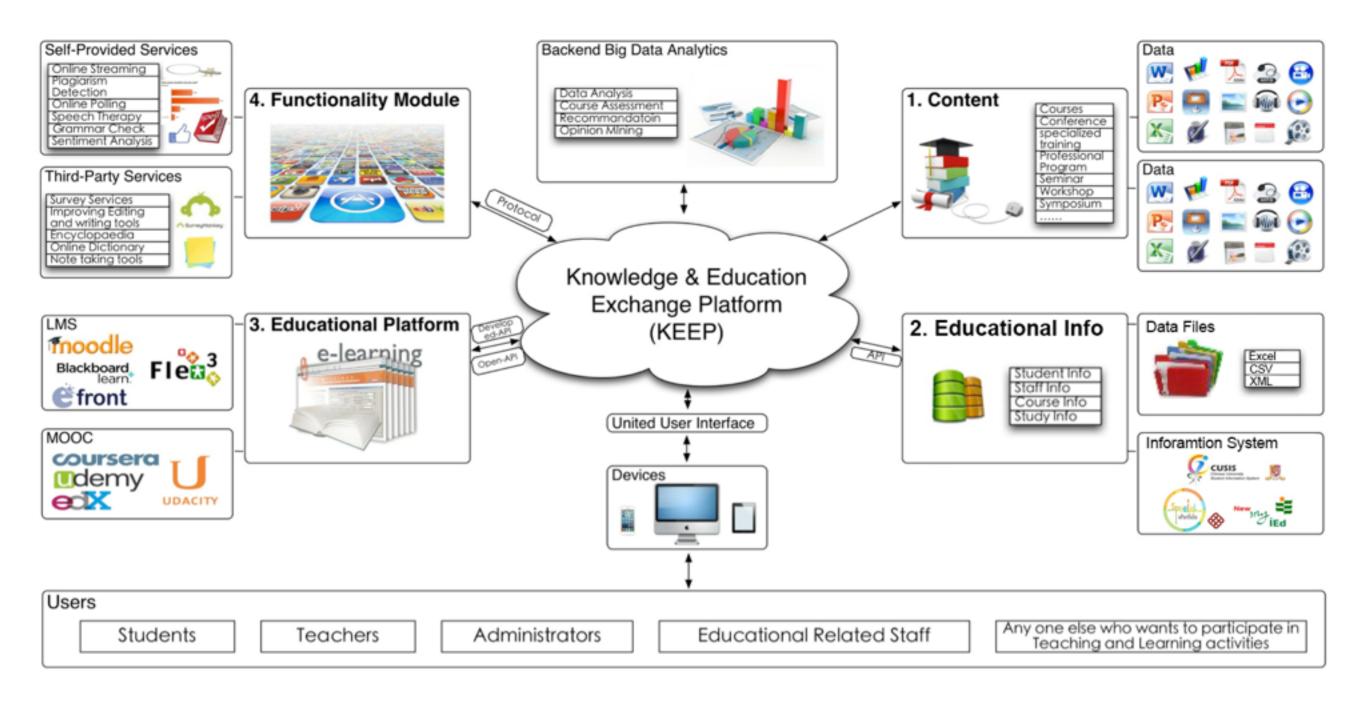


### Some Partners and Alliances





## The KEEP Education Cloud

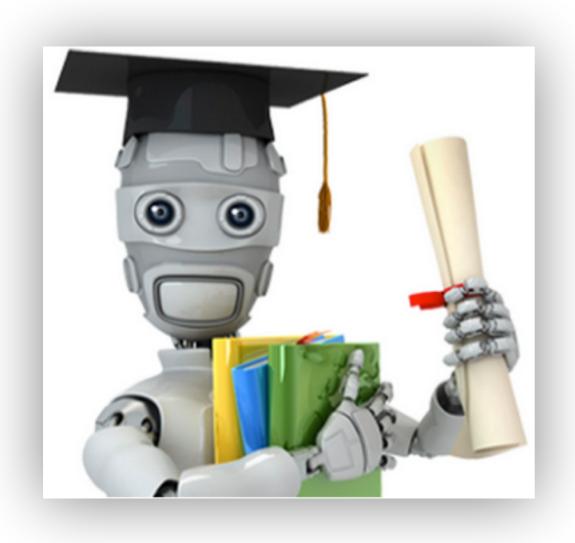




# Work in Progress & Future Works



## Natural Language Processing



- Text and semantic analysis
- Summarization
- Sentiment analysis
- Automated grading
- Q&A systems



### Recommendations



- Personalized learning
- Courses, tutors, peering learning partners, etc.
- Learning resources, time allocation, etc.
- Career planning



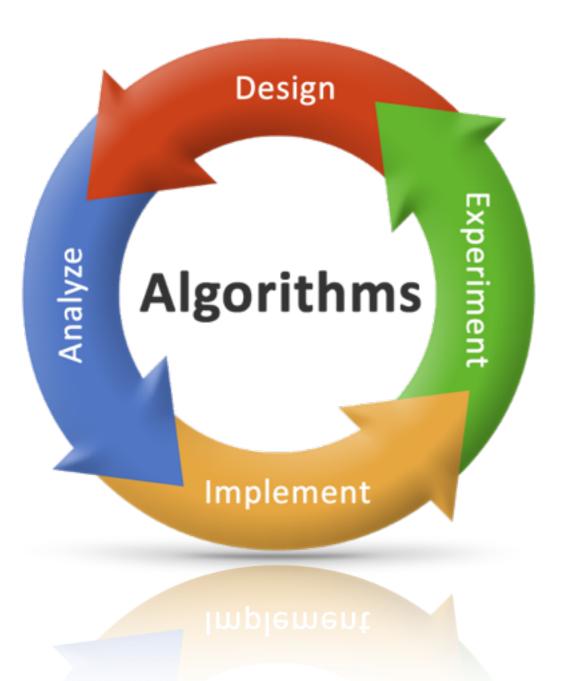
## Knowledge Map



- Explore topics
- Track topic changes
  - Make topic comparisons and inferences
- Better search on concepts



## Algorithms & Techniques



- Machine learning
- Data analytics
- Social computing
- Web intelligence
- Multimedia information processing



### Get Involved









"I think you'll find that mine is bigger ... "



## **Concluding Remarks**

### • Be Inspired

- Big Education is the focus!
- Be Informed
  - Big Data in Education is the VALUE proposition!
- Be Challenged
  - Use technologies to transform education in the Big Data Era!





#### CALL FOR PAPERS (TXT)

Researchers worldwide are currently producing more and more scholarly data of various types such as papers, books, patents, etc. Such data are big data by nature. For example, the DBLP Computer Science Bibliography provides bibliographic information on major computer science journals and proceedings. Additionally, DBLP indexes more than 2.3 million articles records containing title, pages, years and authors' information, etc. Concurrently, scholars are associated with various academic activities such as conferences, workshops, congresses, peer review, and so on. Such scenarios have motivated us to explore the Web of Scholars in the context of big scholarly data on a global scale. It is imperative and vital for researchers to drive their knowledge towards the innovative generation of values from Big Scholarly Data. The emerging worldwide Web of Scholars demands a re-evaluation of existing techniques, such as data mining, recommender systems, and social network analysis. Furthermore, there is a demand for novel ways of developing algorithms, methods and techniques to foster the analysis and interpretation of social environments such as academic collaboration networks.

In this workshop, we will explore promising areas of research in big scholarly data, with a focus on the rapidly emerging field of the Web of Scholars. This workshop also seeks to answer noteworthy research questions such as:

- How to connect scholars on the web?
- How to facilitate collaboration among scholars?
- How to find the experts in a particular field?

Researchers are welcome to submit their papers that address these questions above and other topics below which may include, but are not limited to:

- Academic social network analysis
- Scientific recommendation
- Methods and tools for analyzing big scholarly data
- Indexing, searching, and mining scholarly data
- Connecting scholars using a Web approach
- Platforms and services for the Web of Scholars
- Web tools and techniques for big scholarly data
- Paradigms to promote scientific collaboration
- Scientific trends prediction
- Applications, use cases, and evaluations of big scholarly data

#### IMPORTANT DATES

Paper Submission Deadline: Jan 14, 2014 Jan 28, 2014 Author Notification: Feb 4, 2014 Final Manuscript: Feb 12, 2014

#### WWW2014 Workshop on Web-based Education Technologies (WebET 2014) April 9, 2014, Seoul, Korea

Home About Committee Authors Attendees Program



#### WebET 2014

The Web has long been recognized as a powerful platform for teaching and learning. The educational community was among the early adopters of the technology and has contributed to its evolution. We are at this point at a major inflection point for Webbased Education Technologies. The convergence ("a perfect storm") of new technologies supporting search, social media, semantics, data mining (Big Data), and others along with current interest to distributed educational pedagogies such as connectivism, behaviorism, and "the flipped classroom" promises to dramatically change Web-based Education Technologies in the near future. The interest in Massive Open Online Courses (MOOCs) has been described as "a tsunami in education" and has re-kindled valuable discussions regarding the role of WebET. Search ...

SEARCH

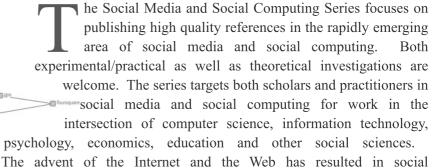
#### Call for Papers

Important Dates

- Submission Deadline: Jan. 14, 2014
- Author Notification: Feb. 4, 2014
- Final Manuscript: Feb. 12, 2014



# **SOCIAL MEDIA & CALL FOR BOOKS!**



interactions and behaviors through the use of technologies and web services, e.g., hardware devices such as smart phones, tablets, RFID, etc., software services such as wikis, blogs, micro-blogs, social network sites, recommender systems, social bookmarking, social

news, multimedia sharing sites, etc. Analyzing these technologicallyenabled interactions in their social context will benefit information providers and information consumers. However, the large volume and scale of user-generated contents require effective modeling methods and efficient algorithms to handle these chalenging problems.

#### Series Editor:

Irwin King



Prof. King is Associate Editor of the IEEE Transactions on Neural Networks (TNN) and IEEE Computational Intelligence Magazine (CIM). He is a senior member of IEEE and a member of ACM, International Neural Network Society (INNS), and VP & Governing Board Member of the Asian Pacific Neural Network Assembly (APNNA). He serves the Neural Network Technical Committee (NNTC) and the Data Mining Technical Committee under the IEEE Computational Intelligence Society.

https://www.cse.cuhk.edu.hk/irwin.king/home kin@cse.cuhk.edu.hk

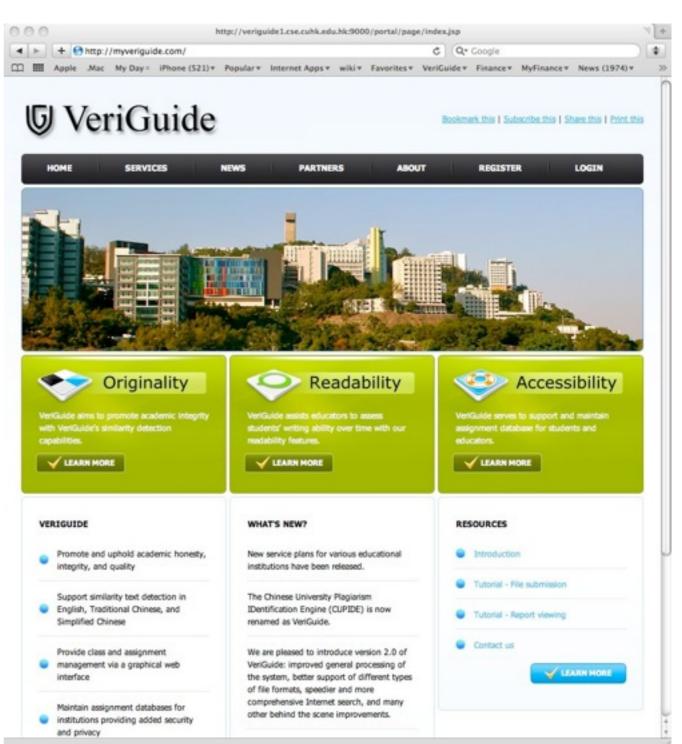


Pleas send Proposals to either the Series Editor or Directly to : LEONG Li-Ming Editor, CRC Press 240 Macpherson Road, #08-01, Pines Industrial Building, S'pore 348574 li.ming@tandf.com.sg Tel: (65) 67415166 x 115



**⑤** VeriGuide

- Similarity text detection system
- Developed at CUHK
- Promote and uphold academic honesty, integrity, and quality
- Support English, Traditional and Simplified Chinese
- Handle .doc, .txt, .pdf, .html, etc. file formats
- Generate detailed originality report including readability





## Acknowledgments

- Shouyuan Chan (Ph.D.)
- Xixian Chen (Ph.D.)
- Chen Cheng (Ph.D.)
- Junjie Hu (Ph.D.)
- Baichuan Li (Ph.D.)
- Guang Ling (Ph.D.)
- Haiqin Yang (Postdoc)
- Connie Yuen (Ph.D.)

- Hongyi Zhang (Ph.D.)
- Shenglin Zhao (Ph.D.)
- Tong Zhao (Ph.D.)

 Looking for more PhD students working on machine learning, Big Data, social computing,



. . .

## Acknowledgments

- Roger Cheung
- Byron Lai
- Daisy Lau
- Patrick Lau
- Sophia Man
- Lin Tsang
- Junfeng Yang
- Raymond Yuen

 Looking for more engineers, programmers, system analysts, etc. to work on KEEP...





### The Chinese University of Hong Kong



## CUHK Excellence

 The only university in Hong Kong having Nobel Laureates as faculty with five Distinguished Professorsat-Large



Prof. Yang Chen-Ning, Nobel Laureate in Physics



Prof. Charles Kao Nobel Laureate in Physics



Prof. Sir James A. Mirrlees, Nobel Laureate in Economic Sciences



Prof.Yau Shing-Tung, Fields Medalist



Prof. Andrew Yao, Turing Award

 Nine academicians of Chinese Academy of Sciences and Chinese Academy of Engineering







#### Reference

Cortez P, Silva A M G. Using data mining to predict secondary school student performance[]]. 2008. Silver D, Saunders M, Zarate E. What factors predict high school graduation in the Los Angeles Unified School District[]]. Policy Brief, 2008, 14. http://blog.gmee.com/online-in-60-seconds-infographic-a-year-later/ http://stor-age.zdnet.com.cn/stor-age/2012/1114/2130609.shtml http://www.bigdatanetwork.net/sites/default/files/field/image/value big data infographic.jpg http://www.today.mccombs.utexas.edu/sites/default/files/Open-Cover-spring-2012-Big-Data\_web.jpg http://www.carolsim.com/designworks/archive/CEO/Training-Period.gif http://www.mbain.eu/sites/default/files/Lifelong%20Learning.jpg?1398341082 http://graphs.net/wp-content/uploads/2013/01/Milestones-in-E-learning.jpg http://elearningindustry.com/top-10-e-learning-statistics-for-2014-you-need-to-know http://athentica.com/wp-content/uploads/2013/10/Online-Learning-Landscape-Oct-2013.jpg http://degreeoffreedom.org/mooc-infographics/ http://instructordiploma.com/wp-content/uploads/2010/05/teaching-e1273735852117.jpg http://www.allbusinessschools.com/asd-images/abs/database-administrator.jpg http://www.lipolitics.com/wp-content/uploads/media/2011/06/2012/06/3rdParty-1.png http://ims201juliadib.files.wordpress.com/2013/05/free-elearning-udacity-coursera-mitz-edx.png http://3.bp.blogspot.com/-RSsESvlbHRo/UW5LEKRXR0I/AAAAAAABDM/0r-g3Ra6h5s/s1600/save+money.jpg http://www.onefortraining.org/sites/default/files/core/SelfPacedLearningButton.jpg http://add-assets.com/asset/17.jpg http://www.smallbiztechnology.com/wp-content/uploads/2014/06/customer-feedback.jpg http://www.voiceyourmedia.com/wp-content/uploads/2012/10/Web-Accessibility-ADA-Section-508.png http://www.tsupportllc.com/wp-content/uploads/2013/12/600px-Baobab\_responsive\_imac\_macbook\_ipad\_iphone.png http://www.cambridgedigital.com/userfiles/collaborative-learning.jpg http://blog.themsls.org/wp-content/uploads/2014/04/Shutterstock\_Questions\_Image.jpg\http://static.guim.co.uk/sys-images/Guardian/Pix/pictures/2008/08/30/Stressedteacher-460x276.jpg http://extras.mnginteractive.com/live/media/site36/2014/0104/20140104 20140105 BI CDXXTEACHEREVALS(1)~p1.jpg http://tctechcrunch2011.files.wordpress.com/2013/11/user-timeline.png?w=400 http://www.ed.gov/blog/wp-content/uploads/2013/02/ADEdgewoodDCPrep-300.jpg http://elearninginfographics.com/wp-content/uploads/The-Minds-Behind-The-MOOCs.jpg http://elearninginfographics.com/the-minds-behind-the-moocs-infographic/ http://kimdalferes.com/wp-content/uploads/2013/07/First-day-of-college-cartoon.jpg http://www.pinnaclecare.com/cms/assets/parents-can-work-with-medical-providers-to-help-children-with-asd 1500 650964 0 14006637 500-300x212.jpg http://e459.http.cdn.softlayer.net/80E459/people4smartercities.com/sites/default/files/education smart moves visual.jpg https://dlz850dzhxs7de.cloudfront.net/topics/ml/large-icon.png http://www.excella.com/blog/wp-content/uploads/2013/02/Code-review\_11935373\_m.jpg http://www.agendaitalia.it/wp-content/uploads/2013/02/Google-Knowledge-Graph.png http://www.bjcbranding.com/wp-content/uploads/2013/05/search-engine.jpg

