

Quality of Textbooks: An Empirical Study

R. Agrawal S. Chakraborty S. Gollapudi A. Kannan K. Kenthapadi
Search Labs, Microsoft Research, Mountain View, CA, USA

Textbooks are the educational input most consistently associated with gains in student learning. Particularly in developing countries, textbooks are the primary conduits for delivering content knowledge to the students and the teachers base their lesson plans on the material given in textbooks. Abstracting from the education literature, we propose that well-written textbooks exhibit the following properties:

FOCUS. Each section explains very few concepts.

UNITY. For each concept, there is a unique section that best explains the concept.

SEQUENTIALITY. Concepts are discussed in a sequential fashion: a concept is explained prior to occurrences of this concept or any related concept. Further, the tie for precedence in presentation between two mutually related concepts is broken in favor of the more significant of the two.

Quantification of these properties in a textbook requires identification of: (i) concepts presented in the book, and (ii) the key section for every concept. We define concepts to be terminological noun phrases. We denote by $\lambda(c, i)$ the significance score of concept c in section i for understanding other concepts in the book and algorithmically compute it as a function of the frequency of c in i and the number of concepts related to c . We set the key section for a concept to be the section with the largest significance score for the concept.

We computed the extent to which these properties are followed in high school textbooks published by the Indian National Council of Educational Research and Training. We report the findings for three books: Grade XII History, Grade XII Sociology, and Grade X Science.

FOCUS. We measure FOCUS by computing the average number of concepts explained in a section (closer to 1 is better) and the deviation of the number of concepts explained in each section from this average. The average number of concepts explained per section is respectively 2.8 (Grade XII History), 5.3 (Grade XII Sociology), and 3.9 (Grade X Science), and the standard deviation across all sections is 0.9 times the respective average for all three books. Fig. 1 shows the distribution of the number of concepts explained in different sections for Grade XII Sociology book. We see that some sections explain more than 10 concepts (one of them as many as 21), making these sections hard to comprehend.

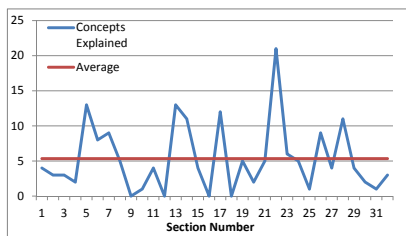


Figure 1: Adherence to FOCUS (XII Sociology)

UNITY. For a given concept c , find the two sections, k_c, k'_c , that have the largest significance scores, *i.e.*, $k_c := \arg \max_{1 \leq i \leq n} \lambda(c, i)$ and $k'_c := \arg \max_{1 \leq i \leq n, i \neq k_c} \lambda(c, i)$. Now, define the *unity index* for concept c as $\lambda(c, k_c) / \lambda(c, k'_c)$. The larger the index, the more the concept respects UNITY.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
ACM DEV'12, March 11–12, 2012, Atlanta, GA.
Copyright 2012 ACM 978-1-4503-1262-2/12/03 ...\$10.00.

Fig. 2 shows the normalized histogram of unity index values across the three books. X-axis corresponds to the (binned) values of the index and Y-axis corresponds to the fraction of the concepts with that index. For books exhibiting good unity, smaller fraction of total concepts should have unity index closer to one and larger fraction of concepts should have higher values. We see that the results are quite opposite. For Grade XII History, nearly 50% of concepts have unity index of one! Smaller number of, but still nearly 30% of the concepts have unity index of one for Grade X Science as well as Grade XII Sociology. Interestingly, much less fraction of concepts in Grade XII Sociology have index of one when compared to Grade XII History. This value is closer to the value for Grade X Science, though arguably the nature of material in Grade XII Sociology is closer to Grade XII History than Grade X Science.

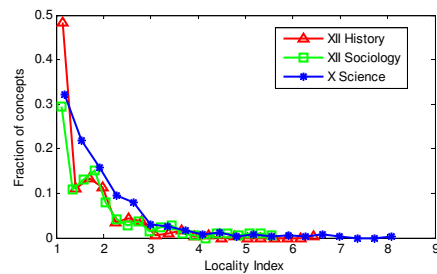


Figure 2: Adherence to UNITY

SEQUENTIALITY. For each concept c , we create a sorted list of sections ordered by section number in which the concept is present. We define the rank of the key section for concept c to be the position of the corresponding section in this list.

Fig. 3 shows the cumulative fraction of concepts that belong up to a certain rank for three books. Clearly, if the key section for every concept in a book had a rank of one, then the graph for the book would be a single point (1,1). However, due to related concepts, this ideal can only be approached but not achieved. Still there are differences between the books. We can see from the figure that Grade XII Sociology book explains about 90% of the concepts within the third section it is mentioned while Grade XII History book explains only shy of 80% of them. The graph for Grade XII Sociology is again more similar to Grade X Science than Grade XII History, indicating the influence of the organization on SEQUALITY irrespective of the subject.

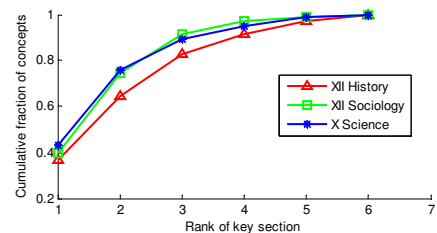


Figure 3: Adherence to SEQUALITY

This empirical study points out the need for a diagnostic tool to help authors identify sections and concepts that can benefit from reorganizing the material presented. We are in the midst of building such a tool.