New Perspectives in Science Education
International Conference
NEW PERSPECTIVES in SCIENCE EDUCATION


Text Mining and Visual Analysis in the Context of Informatics Textbooks

Guglielmo Abbruzzese[1], Teresa Celestino[2]

Department of Mathematics, Tor Vergata University, Rome, Italy[1]
CAFIS, Roma Tre University, Rome, Italy[2]

Abstract

Textbooks' analysis is a general area of interest in understanding educational trends. This paper deals with informatics textbooks, where the term "informatics" includes both computer science and ICT, considering their transversal and interdisciplinary aspects. Given the pervasiveness of informatics and its impact on school systems, it is particularly important to study the structure of its textbooks. In this paper, a carefully selected sample of informatics textbooks used in Italy and in the International Baccalaureate Organisation secondary schools is analysed in order to assess the degree of interdisciplinary openness, with particular attention to ethical, historical, philosophical and cultural issues in a broad sense. The analysis was carried out by manual text mining and evaluation of graphic elements, considering quantitative and qualitative factors Italian textbooks are generally more attractive from a graphic point of view, but flattened on procedural contents, paying little attention to the historical-epistemological roots of informatics as well as to interdisciplinary aspects.

Keywords: *science textbooks; informatics; text mining; interdisciplinarity; computer science education.*

1. Introduction

The impact of computer science (CS) and information & communication technology (ICT) has significantly transformed individual and societal habits, making basic training essential for all citizens starting from the school curriculum [1]. Unlike what happens in some foreign countries, in Italy CS has yet to secure its fully recognised role within the educational system [2].
Before discussing the "computer revolution", it is important to distinguish between CS and ICT. CS involves the representation, organisation, and automatic processing of information, while ICT covers the broader study of acquiring, processing, storing, and communicating information with computer technology. The term "informatics" includes both CS and ICT [3], making it a more general concept. Thus, it is possible to speak not only of *computational thinking*, but also of *informatical thinking*, including the transversal and interdisciplinary aspects of computer science and its applications in a broad sense [4].
Only in recent years it has become clear how the computer revolution has profound implications for learning and teaching processes; teachers must therefore offer students the mental tools to best manage the possibilities offered by new technologies. This involves a reflection on the very nature of informatics at an epistemological level and on its relationships with other disciplines, not only in a merely applicative sense, but also in a historical, philosophical and ethical sense.
Informatics is not ethically neutral [5]. While it has enabled significant progress in fields such as education, health and environment, it has also had negative consequences. For example, Spitzer [6] discusses the harmful effects of smartphones, and Floridi [7] examines the ethical challenges and opportunities of AI. Because good teaching materials for CS are very relevant [8], this work aims to examine the current state of teaching interdisciplinary informatics through the analysis of a traditional educational tool: the textbook.

2. Informatics in Italian and IB Schools

This work examines two types of textbooks, used in Italian and International Baccalaureate Organisation (IB) upper secondary schools.
The IB is a non-profit international foundation based in Geneva. IB schools, public or private, offer access to universities around the world. Table 1 shows how the IB levels correspond to the Italian education system.

Education system	School level	School level name	Students' ages
Italian	1	Pre-primary school	3÷6
	2	Primary school	6÷11
	3	First level secondary school	11÷14
	4	Second level secondary school	14÷19
IB	1	Primary years program (PYP)	3÷12
	2	Middle years program (PYP)	11÷16
	3	Diploma program (DP)	16÷19
	4	Career-related program	16÷19

Table 1. Italian and IB education system.

2.1 Informatics in Italian Schools

Italian high school lasts five years, attended by students aged 14 to 19; thus, 4th-5th year Italian students are 18-19 years old, while their IB counterparts in the final two years (11th-12th grade) are 17-18 years old.
In Italian high schools, informatics is taught independently in technical institutes specialising in economics ("Istituto Tecnico Economico", ITE) and technology ("Istituto Tecnico Tecnologico", ITT). In the ITT focused on "Computer Science and Telecommunications", the subjects taught include:

- "Tecnologie e Progettazione di Sistemi Informatici e di Telecomunicazioni" (TPSIT), that is "Technologies and Design of Information and Telecommunications Systems";
- "Gestione Progetto e Organizzazione Impresa" (GPOI), that is "Project Management and Business Organisation";
- "Sistemi e Reti" (SR), that is "Systems and Networks".

There are several types of Liceo, an Italian upper secondary school where historical-philosophical subjects are studied in depth. The "Liceo scientifico – opzione scienze applicate" (LS-SA), offers more hours in biology, geology, chemistry and computer science. CS is a separate subject only in LS-SA (it is taught two hours a week). A third type of Italian upper secondary school is the professional institute, which emphasises practical education and teaches only ICT, not CS.

2.2 Informatics in IB Schools

Both CS and ICT are included in the IB syllabus, with ICT referred to as ITGS (Information and Communication Technology in a Global Society), in line with IB philosophy. The course has been renamed Digital Society, but this work retains the term ITGS to reflect the textbooks used up to recent times. ITGS is not focused on programming or computer architecture; its main goal is to explore the relationship between technology and people, placing it within the social sciences.
Given this context, the presence of Theory of Knowledge (ToK) in the DP curriculum, also in CS and ITGS, is understandable. ITGS textbooks frequently reference ToK often dedicating sections to it, encouraging readers to consider different perspectives and pose thought-provoking questions. ToK addresses questions about the origins and structure of knowledge. The aim is to promote discussions between students and teachers, not to find definitive answers, but to inspire ongoing inquiry.

3. Textbooks' Classification Criteria

The textbooks have been analysed in order to evaluate informatics teaching within Italian and IB system in order to deduce, for example, their level of interdisciplinarity (which in turn can provide indications about the methodological approaches generally used). The analysis was carried out by manual text mining [9] and evaluation of graphic elements.
The selected textbooks are divided according to the age groups of the students when possible. CS e ICT Italian textbooks analysed are used in the following upper secondary schools:

- LS-SA (students' age: 14-15/16-19);
- ITT specialising in Computer Science and Telecommunications (students'age: 14-15/16-19).

Concerning IB-DP textbooks, ITGS and CS (SL/HL) have been considered (age of students: 16-19).
To evaluate the selected textbooks some indicators have been chosen, even taking into account the criteria formulated by Fourez [10] for scientific literacy from a mainly socio-constructivist perspective, according the following areas:

- Graphic design

- Economic/political/social axis
- Historical/philosophical axis
- Cultural axis (aesthetic, communicative, corporal, pragmatic)
- Project Management (PM)/Business area
- Environmental field
- Other (technical content)

3.1 Selected Textbooks

In the preliminary phase, many textbooks were viewed beyond those analysed in this document.
The initial sample of Italian textbooks included those adopted in the largest number of schools until the 2022/2023 school year, according to freely accessible databases. For IB textbooks the choice was simpler, given the very small number of titles due to the need to standardise teaching tools as much as possible at a global level. The number of publishers involved is also limited, since IB books are never published by local publishing houses.
The selected texts were purchased in paper or digital format, or borrowed from school libraries or some teachers. The edition of some texts found in the libraries is not recent, but the numerous Italian teachers interviewed (via e-mail or in person) assured that the differences with the texts currently adopted are minimal and do not concern the contents. The changes made from edition to edition respond above all to sales strategies, in order to minimise the purchase of used texts (for example by changing the page numbers and the numerical order of the exercises, so that the used text becomes uncomfortable for the students). IB teachers were also interviewed for indications on the most representative texts to examine; their addresses were found among those provided by the teachers themselves during previous research on IB and Italian schools [11]. The texts were also selected because they have been on the market for many years. This also motivated the choice of a non-recent year of publication, approximately in the middle of a time interval within which the same text has been used, even in various editions: this means that the text's structure (more or less constant) has influenced various generations of students, before and after the year considered; this excluded the choice of editorial novelties. The books reported in the following two paragraphs were selected, as they were considered sufficiently representative in their respective categories. To simplify the numerous citations, every book has been labelled with an unambiguous alias.

3.1.1 Italian Selected Textbooks

In addition to the textbooks of the first two years, those of the three-year specialisation of the ITT were analysed: Technologies and Design of Information and Telecommunications Systems (TPSIT), Systems and Networks (SR), Project Management and Business Organisation (GPOI). The books relating to the last two subjects (SR and GPOI) were not classified for each parameter selected for the comparative analysis, either for the strong typographical analogy with the TPSIT books, or for the greater specificity of the contents, which would have made the comparisons less homogeneous.
The selected textbooks are listed in Table 2.

ITT	Code	Title	Authors, year
First two-year period	ITT-CS1B-BV	“Il nuovo Dal bit al web”	Barbero & Vaschetto, 2014 [12]
	ITT-CS1B-BLP	“Mastermind – Pensare Programmare Condividere”	Boscaini, Lunghezzani & Princivalle, 2016 [13]
ITT	**Code**	**Title**	**Authors**
Last three-year period	ITT-CS2B-GS	“Cloud: Informatica – Secondo biennio. Istituti Tecnici – settore tecnologico indirizzo Informatica e Telecomunicazioni” Vol. A&B	Gallo & Salerno, 2012 [14]
	ITT-SR35-LB	“Sistemi e reti – Per l’articolazione informatica degli istituti tecnici settore tecnologico” Vol. 1÷3	Lo Russo & Bianchi, 2012 [15]
	ITT-TP35-CN	“Tecnologie e progettazione di sistemi informatici e di telecomunicazioni” Vol. 1÷3	Camagni & Nikolassy, 2012 [16]
	ITT-PM5-DD	“Il project management nella scuola superiore (Gestione Progetto e Organizzazione	Dell’Anna & Dell’Anna, 2015 [17]

		Impresa)"	
LS-SA	**Code**	**Title**	**Authors**
First two-year period	LS-CS1B-LG	"Informatica: Per licei scientifici Scienze Applicate"	Govoni, 2015a [18]
	LS-CS1B-GS	"Informatica app"	Gallo & Sirsi, 2018 [19]
LS-SA	**Code**	**Title**	**Authors**
Last three-year period	LS-CS2B-LG	"Informatica: Per licei scientifici Scienze Applicate"	Govoni, 2015b [20]
	LS-CS5A-LG	"Informatica: Per licei scientifici Scienze Applicate"	Govoni, 2015c [21]

Table 2. Italian informatics textbooks selected.

3.1.2 IB Selected Textbooks

The IB textbooks are generally intended for an international audience, not necessarily native speakers. They are written by teachers with many years of experience in the international circuit, preferably trained in different fields of knowledge (the author Oliver Kim, for example, has taught Biology, ToK and ITGS in the IB-DP courses). The selected textbooks are listed in Table 3.

IB	Code	Title	Authors
Diploma Program	IB-ITGS-KIM	"Information Technology in a Global Society"	Kim, 2011 [22]
	IB-ITGS-SG	"Information Technology in a Global Society"	Gray, 2011 [23]
IB	**Code**	**Title**	**Authors**
Diploma Program	IB-CCS-DH	"Core Computer Science"	Dimitriou & Hatzitaskos, 2015 [24]
	IB-ACS-DH	"Advanced Computer Science"	Dimitriou & Hatzitaskos, 2016 [25]

Table 3. IB informatics textbooks selected.

4. Graphic Design: Data Collection

Modern school textbooks pay great attention to graphic design. The presence of images can trivially respond to mere aesthetic criteria or to specific didactic needs for key topics learning. Colours, size and position of the images influence the style of the text page, providing important food for thought. For example, IB textbooks are usually printed in black and white. In Italian textbooks, each piece of information, its location and its function are characterised in order to be presented in a specific typographical guise.
The use of areas to highlight texts (with different colours according to their function), note elements on diagrams and images, graphs, tables, portions of code with coloured text, can represent a visual shock if analysed immediately after the books of the IB circuit.
To analytically formulate this characteristic, some data extrapolated from the listed texts using precise parameters are provided:
- text/subject identifier;
- usable space size of a page (with reference to A4 sheet);
- number of usable pages (i.e. excluding indexes, appendices, etc.);
- type of elements different from pure text:
    - Text Box (this is still text, but highlighted within coloured areas).
    - Images (photographs, infographics, or drawings of situations or subjects).
    - Diagrams (explanatory images of flows or processes, or photos accompanied by notes).
    - Graphs.
    - Tables.
    - Code (portions of programming code inserted in coloured areas).

For each of the elements listed, the occurrences were counted and classified both on the basis of the size in relation to the page and on the basis of the use of colours.
An example of analysis is shown in Table 4 for Kim [22]; the same type of analysis was carried out for all selected textbooks.

Book	IB-ITGS-KIM
Pages	402
Size	21.3 cm x 28.0 cm
Page area	596.4 cm^2 (95% A4)
Book area	239752 cm^2

	Small	**Medium**	**Large**	**Full**		%
Text Box	171	11	1	0		3.68
Image	189	4	1	0		3.75
Scheme	44	4	1	0		1.03
Graph	11	0	0	0		0.21
Table	15	7	2	0		0.66
Code	2	0	0	0		0.04
tot	432	26	5	0	**tot**	%
Area (cm^2)	19440	2340	675	0	22455	**9.37**

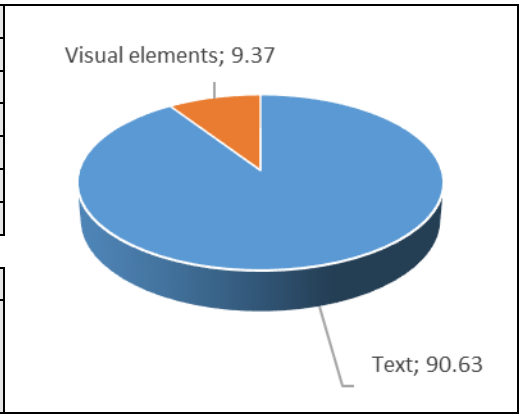

Table 4. Analysis of Kim (2011).

The overall results are shown in Figure 1.

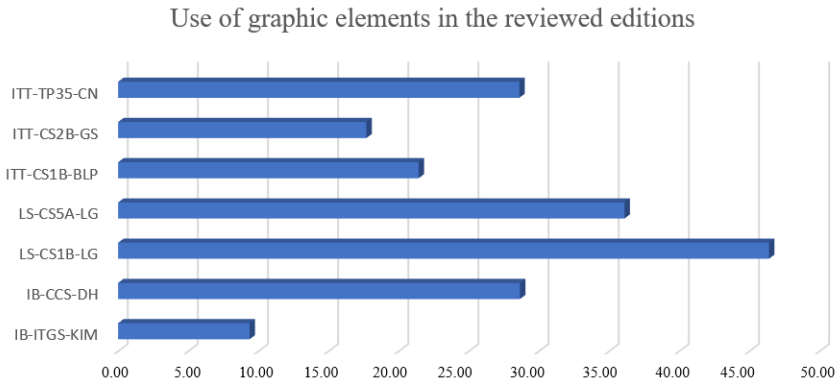

Fig. 1. Graphical representation of the overall graphic analysis

The main differences are as follows:

a. The manuals with the greatest use of graphic elements, also considering the size of the page and the relative margins, are intended for the Liceo.
b. IB texts are those that use images and colours the least.
c. ITT books – especially those of the three-year period – are those with the greatest graphic affinity with those of the IB.
d. The books of the first two-year period have a greater graphic component than those of the three-year period.

5. Interdisciplinary Character: Data Collection

Below are some indications about the operational choices in relation to textbook classification according to the aforementioned fields of knowledge. The classification of the pages has been realised with an approach as rigorous as possible, paying attention to infographics, inserts, case studies, insights, exercises, boxes of various kinds. The topics proposed in the various textbooks are generally similar (for example computer architecture, operating systems, web, security, networks), even if the same topics are treated with a different point of view; for example, we can appreciate the differences between the IB textbook of Kim [22] and the textbook of Barbero&Vaschetto [12], a very used text in Italian LS-SA (Figure 2).

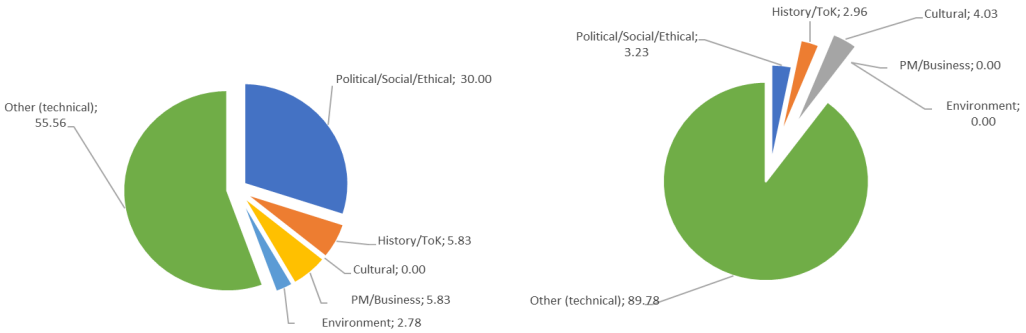

Fig. 2. Classification of Kim (2011) and Barbero&Vaschetto 2014) according to the fields of knowledge.

The diagram in Figure 3 shows the overall data for all the axes examined. IB textbooks, with a generally sober graphic design, show strong attention to ethical, social, and political themes, particularly in relation to ToK, in line with the IB curriculum. This is not the case for Italian technical institute and Liceo textbooks, despite national guidelines emphasising these themes, especially for Liceo. The LS-SA and ITT textbooks share some similarities. Although Liceo books should emphasise historical and epistemological aspects, LS-SA texts are disappointing in this regard, focusing instead on high graphic liveliness; owever, this graphic focus appears to be more aimed at usability rather than improving learning outcomes. ITT textbooks, though specialised, show some attention to the general and interdisciplinary aspects of computer science.

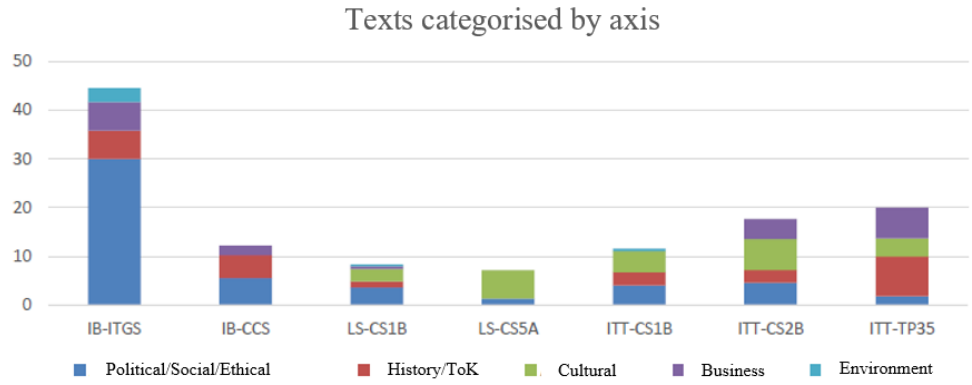

Fig. 3. Graphical representation of interdisciplinary character

5. Conclusion

Starting from IB and Italian school systems, a comparative study of selected textbooks on CS and ICT was conducted using quantitative and qualitative indicators. Textbooks play an important role in IB student success, since they are shaped by continuously updated curricula resulting from educational research carried out within the IBO. IB textbooks support broad-ranging, interdisciplinary teaching, foster critical thinking; they are characterised by simple graphic design and focused on textual content. Italian Liceo textbooks, on the other hand, lack historical, epistemological, and socio-political depth, focusing on procedural content with lively graphics. The textbooks of the technical institute are more balanced, thanks to a long tradition of subject teaching and to the longer school time provided.
While the quality of texts is not directly related to student performance, they reflect in various ways current educational trends. Although the IB system is not without flaws, IB textbooks' high quality raises important questions about the critical use of Italian ones, which since the 1960s have shifted to prioritising graphics over content. Greater attention from Italian teachers to the quality of texts and educational research in the field of computer science could begin to trigger a real process of change.

REFERENCES

[1] Nardelli, E. "La rivoluzione informatica. Conoscenza, consapevolezza E potere nella società digitale", Roma, Themis, 2022.
[2] Abbruzzese, G., Celestino, T., De Vitis, A. "Epistemologia, etica ed estetica dell'informatica: insegnamento tra tecnologie e mondo digitale". Nuova Secondaria Ricerca, 2023, 2, 431-450.

[3] Gorn, S. “Informatics (Computer and information science)”, Knowledge, 1982, 4(2), 173-198. https://doi.org/10.1177/107554708200400203
[4] Lodi, M. “Informatical Thinking”, Olympiads in Informatics, 2020, 14, 113–132. https://doi.org/10.15388/ioi.2020.09
[5] Gotterbarn, D. “Informatics and professional responsibility”, Science and Engineering Ethics, 2001, 7(2), 221-230. https://doi.org/10.1007/s11948-001-0043-5
[6] Spitzer, M. “Die Smartphone-Epidemie. Gefahren für Gesundheit, Bildung und Gesellschaft”, Stuttgart, Klett-Cotta, 2018.
[7] Floridi, L. “The ethics of artificial intelligence: Principles, challenges, and opportunities”, Oxford, Oxford University Press, 2023.
[8] Cerar, Š., Lokar, M., Anželj, G., Brodnik, A., Nančovska Šerbec, I. “Textbooks and Materials for Teaching Computer Science in Slovenia”. In: Bollin, A., Futschek, G. (Eds) “Informatics in Schools. A Step Beyond Digital Education”. ISSEP 2022. Lecture Notes in Computer Science, vol 13488, Cham, Springer, 2022. https://doi.org/10.1007/978-3-031-15851-3_12
[9] Brock, R., Tsourakis, N., Kampourakis, K. “Using Text Mining to Identify Teleological Explanations in Physics and Biology Textbooks: An Exploratory Study”, Science & Education, 2024. https://doi.org/10.1007/s11191-024-00513-3
[10] Fourez, G. “Le mouvement Sciences, technologies et société (STS) et l’enseignement des sciences”, Perspectives, 1995, 25, 27-41.
[11] Celestino, T., Marchetti, F. “Surveying Italian and International Baccalaureate Teachers to Compare their Opinions on System Concept and Interdisciplinary Approaches in Chemistry Education”, Journal of Chemical Education, 2020, 97(10), 3575-3587. https://doi.org/10.1021/acs.jchemed.9b00293
[12] Barbero, A., Vaschetto, F. “Il nuovo Dal bit al web”, Milano-Torino, Pearson Italia, 2014.
[13] Boscaini, M., Lunghezzani, F., Princivalle, D. “Mastermind: Pensare Programmare Condividere”, Milano, Hoepli, 2016.
[14] Gallo, P., Salerno, F. “Cloud: Informatica – Secondo biennio. Istituti Tecnici – settore tecnologico indirizzo Informatica e Telecomunicazioni (Vol. A, B)” Milano, Mondadori, 2012.
[15] Lo Russo, L., Bianchi, E. “Sistemi e reti: Per l’articolazione informatica degli istituti tecnici settore tecnologico (Vol. 1÷3)”, Milano, Hoepli, 2012.
[16] Camagni, P., Nikolassy, R. “Tecnologie e progettazione di sistemi informatici e di telecomunicazioni (Vol. 1÷3)”. Milano, Hoepli, 2012.
[17] Dell’Anna, A., Dell’Anna, M. “Il project management nella scuola superiore”, online e-book on Matematicamente.it, 2015, CC-BY-ND-3.0. Available online at the webpage: https://projectmanagement.matematicamente.it/attachments/category/13/Gestione_Progetto_Libro_V2_02_ebook.pdf (retrieved January 20, 2025).
[18] Govoni, L. “Informatica: Per licei scientifici scienze applicate (Vol. 1)”, Bergamo, Atlas, 2015a.
[19] Gallo, P., Sirsi, P. “Informatica app (Vol. 1)”, Milano, Mondadori, 2018.
[20] Govoni, L. “Informatica: Per licei scientifici scienze applicate (Vol. 2)”, Bergamo, Atlas, 2015b.
[21] Govoni, L. “Informatica: Per licei scientifici scienze applicate (Vol. 3)”, Bergamo, Atlas, 2015c.
[22] Kim, O. (2011). Information Technology in a Global Society. Sunbury, AU: IBID Press.
[23] Gray, S. “Information Technology in a Global Society”, Scotts Valley, CA, USA, CreateSpace Independent Publishing Platform, 2011.
[24] Dimitriou, K., Hatzitaskos, M. “Core Computer Science”, Berkshire, UK, Express Publishing, 2015.
[25] Dimitriou, K., Hatzitaskos, M. “Advanced Computer Science”, Berkshire, UK, Express Publishing, 2016.