

ANALISIS KETERBACAAN TEKS BAHASA INDONESIA DENGAN PROGRAM MICROSOFT WORD

Juliana Tirza

Universitas Pelita Harapan
juliana.tirza@uph.edu

Hananto

Universitas Pelita Harapan
hananto.fip@uph.edu

Abstract

One of the Indonesian government's efforts to improve children's literacy skills is by leveling schoolbooks created by the Ministry of Education, Culture, Research and Technology (Kemendikbudristek). Children are expected to grow their interest in reading-by-reading books that match their level. Unfortunately, the leveling of books by the government is subjective so it needs to be investigated whether the leveling is in accordance with the calculation of keterbacaan statistics which are objective/quantitative. This study aims to see whether the book leveling by the government is in accordance with the calculation of keterbacaan statistics using the Flesch-Kincaid Grade Level formula and the Flesch Reading Ease test available in the MS Word program. 5 children's story books with levels A, B1, B2, B3 and C (from lower to higher) were used as samples in this study. The calculation results showed that out of the five books, there were three (60%) books with inappropriate levels. It can be concluded that book leveling is subjectively less reliable. This paper provides recommendations that can be used to improve book grading by the government.

Keywords: words or phrases for search engine purposes

Abstrak

Salah satu usaha pemerintah Indonesia untuk meningkatkan kemampuan literasi anak adalah dengan membuat pelevelan buku sekolah yang dibuat oleh Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (Kemendikbudristek). Minat baca anak akan tumbuh sesuai level buku yang

mereka baca. Sayangnya pelevelan buku oleh pemerintah itu bersifat subjektif sehingga perlu diteliti apakah pelevelan itu tersebut memang sudah sesuai dengan perhitungan statistik keterbacaan teks yang bersifat objektif/kuantitatif. Penelitian ini bertujuan untuk melihat apakah pelevelan buku oleh pemerintah sudah sesuai dengan perhitungan statistik keterbacaan dengan menggunakan formula *Flesch-Kincaid Grade Level* dan tes *Flesch Reading Ease* yang tersedia di program MS Word. 5 buku cerita anak dengan evel A, B-1, B-2, B-3an C (dari yang rendah ke yang lebih tinggi) digunakan sebagai sampel dalam penelitian ini. Hasil perhitungan menunjukkan dari lima buku tersebut, ada tiga (60%) buku yang pelevelannya tidak sesuai. Dapat disimpulkan bahwa pelevelan buku secara subjektif kurang dapat diandalkan. Makalah ini memberikan rekomendasi yang dapat digunakan untuk memperbaiki pelevelan buku oleh pemerintah.

Kata Kunci: Keterbacaan teks, pelevelan buku, keterbacaan teks secara sintaktis

Pendahuluan

Kondisi literasi anak di Indonesia masih rendah. Indonesia masih berada di posisi bawah dalam hal kemampuan literasi membaca. Menurut data PISA tahun 2022, Indonesia berada pada urutan 71 dari 78 negara Kondisi ini memicu kebutuhan pendekatan yang lebih terarah dalam pemilihan bahan bacaan untuk meningkatkan minat baca anak. Pendekatan yang efektif untuk dilakukan ialah memilih bacaan dengan keterbacaan teks yang sesuai dengan level literasi membaca anak.

Pemerintah Indonesia telah mengeluarkan pedoman pelevelan buku dengan mengacu pada Undang-Undang nomor 3 tahun 2017 mengenai perbukuan, dan peraturan Menteri Pendidikan, Kebudayaan, Riset dan Teknologi, nomor 22 tahun 2022, tentang standar mutu buku. Pedoman ini dalam bentuk rubrik dengan beberapa kriteria yang perlu dicentang untuk menentukan keterbacaan teks dan jenjang buku tersebut. Aspek yang diidentifikasi dalam pelevelan buku ialah, jenis kata, tanda baca, huruf kapital, adanya kalimat tunggal, kalimat majemuk atau hanya klausa juga ada tidaknya paragraf dalam teks buku. Berikut adalah ringkasan pelevelan buku nonteks pelajaran dari Panduan Pelevelan Buku Nonteks Pelajaran, Badan Bahasa:

1. Pembaca Dini (usia 0–7): memerlukan perancah (*scaffolding*) untuk mendampingi anak membaca. Simbol Pembaca Dini adalah bintang berwarna merah dengan huruf A.
2. Pembaca Awal (usia 6–8 tahun). Simbol Pembaca Awal adalah lingkaran berwarna ungu dan kode B1.
3. Pembaca Awal (usia 7–9 tahun): memerlukan perancah (*scaffolding*) dan mampu membaca teks berupa kata/frasa dengan kombinasi huruf, klausa, kalimat sederhana, dan paragraf sederhana. Simbol Pembaca Awal B2 ditandai dengan lingkaran berwarna ungu dan kode B2.
4. Pembaca Awal (usia 8–10 tahun). Simbol Pembaca Awal B3 ditandai dengan lingkaran berwarna ungu dan kode B3.
5. Pembaca Semenjana (10–13 tahun): mampu membaca teks secara lancar berbentuk paragraf dalam satu wacana. Simbol Pembaca Semenjana ditandai dengan lingkaran berwarna biru dan kode C.
6. Pembaca Madya (13–15 tahun): mampu memahami beragam teks dengan tingkat kesukaran menengah. Simbol pembaca madya D ditandai dengan segitiga berwarna hijau dan kode D.
7. Pembaca Mahir (lebih dari 16 tahun): mampu membaca secara analisis dan kritis berbagai sumber bacaan untuk menyintesis pemikiran secara lebih baik. Simbol pembaca mahir ditandai dengan segi empat berwarna kuning dan kode E.

Pedoman ini bersifat subjektif dan masih berpeluang keliru dalam proses kurasi buku atau bacaan (Kemdikbud, 2023). Perkiraan kesetaraan usia di setiap jenjang adalah perkiraan yang tidak selalu tepat atau dapat digunakan, terutama untuk siswa dengan kemampuan membaca lebih rendah atau lebih tinggi pada usia yang sama.

Pada dasarnya ada dua cara untuk menentukan level/tingkat kesukaran sebuah teks/bacaan: (1) secara subjektif, dan (2) secara objektif. Pendekatan pertama yang dipilih oleh pemerintah karena relatif lebih mudah bila dibandingkan pendekatan kedua yang lebih kompleks bila dilakukan/dihitung secara manual.

Untungnya program MS Word menyediakan fitur yang dapat menghitung tingkat kesukaran bacaan secara objektif.

Perhitungan Tingkat Keterbacaan secara objektif

Ada dua pendekatan secara objektif yang biasa digunakan untuk menentukan tingkat kesukaran sebuah bacaan, yaitu dengan pendekatan sintaktik dan semantik. Pendekatan sintaktik menghitung jumlah huruf, kata, dan suku kata dalam kalimat (Giles & Still, 2005). Pentingnya kesederhanaan dan kompleksitas sintaksis sangat memengaruhi pemahaman teks, terutama bagi pembaca dengan tingkat keterampilan menengah ke bawah, karena struktur kalimat yang lebih sederhana cenderung lebih mudah dipahami (Eslami, 2014). Fitur sintaksis juga penting dalam mengadaptasi teks untuk berbagai tingkat Kemahiran membaca, seperti yang ditunjukkan dalam penelitian beberapa teks ujian formal (Varlamova & Safonkina, 2020).

Beberapa kelebihan pendekatan sintaksis antara lain, membantu menyederhanakan kalimat kompleks menjadi unit yang lebih mudah dipahami tanpa kehilangan informasi penting, meningkatkan keterbacaan dan kemampuan pemahaman teks (Salman et al., 2023). Di samping itu, ini dapat membantu pembaca mengenali struktur sintaksis, meskipun efektivitasnya bervariasi di antara individu (Funke & Sieger, 2012). Memilah-milah teks berdasarkan unit sintaksis juga dapat mengatasi tantangan dalam membaca dan meningkatkan kemahiran (Pitler & Nenkova, 2008; Rasinski, 1994). Dapat disimpulkan, pendekatan sintaksis memberikan kerangka kerja yang kuat untuk mengevaluasi dan meningkatkan keterbacaan teks. Hal ini menjadikannya pendekatan analisis keterbacaan teks yang efektif dalam konteks pendidikan dan evaluasi.

Microsoft Word menyediakan fitur Keterbacaan Statistik untuk menentukan tingkat keterbacaan teks bahasa Inggris. Landasan teori yang dipakai adalah bahwa kata dan/atau kalimat yang pendek/serhana lebih mudah dipahami bila dibandingkan dengan kata dan/atau kalimat yang panjang/kompleks, misalnya rumus *Flesch-Kincaid Grade Level* dan *Fog Index*. Kedua rumus keterbacaan ini mendasarkan perhitungannya pada jumlah rata-rata suku kata di setiap kata dan rata-

rata jumlah kata di setiap kalimat. Formula *Flesch Reading Ease* dan *Flesch-Kincaid Grade Level* adalah sebagai berikut:

- Rumus skor untuk *Flesch Reading Ease* adalah: $206.835 - (1.015 \times \text{ASL}) - (84.6 \times \text{ASW})$.
- Rumus skor untuk *Flesch-Kincaid Grade Level* adalah: $(.39 \times \text{ASL}) + (11.8 \times \text{ASW}) - 15.59$.

Keterangan:

ASL = *average sentence length* (jumlah kata dibagi jumlah kalimat)

ASW = *average number of syllables per word* (jumlah suku kata dibagi jumlah kata)

Flesch Reading Ease menilai teks pada skala 100 poin. Semakin tinggi skornya, semakin mudah untuk memahami dokumen tersebut. Untuk sebagian besar teks standar, target skornya di antara 60 dan 70. *Flesch-Kincaid Grade Level* menilai teks pada tingkat kelas sekolah AS. Misalnya, skor 8,0 berarti siswa kelas VIII dapat memahami dokumen tersebut. Untuk sebagian besar dokumen, target skor yang diinginkan di sekitar 7,0 hingga 8,0.

Berbeda dengan pendekatan sintaktik, pendekatan semantik menghitung jumlah/persentase kata yang sering muncul dan kata yang jarang dipakai dengan asumsi kata yang sering muncul lebih mudah dipahami daripada kata yang jarang dipakai, misalnya rumus Keterbacaan Dale-Chall (*Dale-Chall Keterbacaan Formula*) yang menggunakan daftar kata-kata umum/sering dipakai dan menghitung skor keterbacaan berdasarkan persentase kata-kata yang ada dalam teks. Pendekatan ini dinilai lebih akurat dibandingkan tes keterbacaan teks konvensional dengan menggunakan rumus sintaktik (Besharati, 2021). Metode semantik memberikan pemahaman yang lebih dalam karena mempertimbangkan makna dan hubungan antara kata dan frasa (Corley & Mihalcea, 2005; Vakulenko, 2023). Kelebihan pendekatan semantik, yang melibatkan pemahaman makna teks, menunjukkan bahwa faktor-faktor seperti memori, perhatian, dan kemampuan kognitif secara signifikan memengaruhi proses membaca anak (Kalugina et al., 2021).

Penelitian ini bertujuan untuk mengetahui ketepatan pelevelan buku secara subjektif dari pemerintah dengan perhitungan *keterbacaan statistic* di MS Word yang bersifat objektif dengan formula *Flesch Reading Ease* dan *Flesch-Kincaid Grade Level*.

Metode Penelitian atau Pendekatan Pembahasan

Ada lima buku cerita bergambar yang diambil dalam penelitian ini. Kelima buku tersebut mewakili 5 jenjang buku (1 buku per jenjang). Kelima buku tersebut adalah:

- (1) A_Traktor Mendapat Bantuan (Marutla, 2022)
- (2) B1_Aku Cinta Planetku (Reinchenbach, 2022)
- (3) B2_Siapa Kau (Hello English, 2022)
- (4) B3_Malam di Musim Dingin (Munshi Premchand, 2022a)
- (5). C_Sang Inspektur Garam (Munshi Premchand, 2022b)

Bagian depan menunjukkan jenjangnya (A, B1, B2, B3, dan C) diikuti oleh judul bukunya. Kelima buku tersebut disusun berdasarkan level terendah ke level yang lebih tinggi.

Prosedur Penelitian

Buku cerita tersebut dalam format buku bergambar elektronik berjenis PDF sehingga harus dikonversi dulu/ditulis ulang menjadi file teks agar bisa dilakukan perhitungan *Keterbacaan Statistics* di MS Word. Sebagai contoh, berikut adalah hasil tulis ulang dari buku berjudul “Traktor Mendapat Bantuan”:

Ini Lori. Ini Skuter. Lori dan Skuter berteman. Mereka bergegas menjenguk teman mereka yang sakit. Skuter melaju sangat cepat. Oh, tidak! Hati-hati! Halo, teman-teman. Aku sakit! Ayo, kita mencari obat! Kamu akan sehat lagi. Traktor sehat kembali. Semua teman traktor datang merayakan. Pesta yang meriah Terima kasih! Traktor sehat kembali.

Hasil dan Pembahasan

Teks yang dianalisis tersebut diketik ulang dalam format MS.Word. Dengan menggunakan opsi Review/Spelling and Grammar (Editor)/Insight (Document stats)

akan diperoleh hasil Keterbacaan Statistik. Berikut adalah satu contoh Keterbacaan Statistik dari buku “Traktor Mendapat Bantuan” tersebut:

Readability Statistics	
Counts	
Words	53
Characters	313
Paragraphs	16
Sentences	15
Averages	
Sentences per Paragraph	1,0
Words per Sentence	3,1
Characters per Word	5,4
Readability	
Flesch Reading Ease	11,0
Flesch-Kincaid Grade Level	12,4
Passive Sentences	0,0%

Gambar 1: hasil *perhitungan Keterbacaan Statistics* teks Traktor Mendapat Bantuan.

MS Word menghitung jumlah kata, huruf, paragraf, dan kalimat. MS juga memberikan perhitungan rata-rata jumlah kalimat dalam paragraf, rata-rata jumlah kata dalam kalimat, dan rata-rata jumlah huruf dalam kata. Bagian paling bawah memberikan informasi skor Keterbacaan menurut rumus *Flesch Reading Ease* (11.0), *Flesch-Kincaid Grade Level* (12,4), dan jumlah kalimat pasif (0.0%)

Tabel 1: Hasil perhitungan Keterbacaan Statistik

No	Buku	Level	Jumlah Kata	Jumlah Kalimat	Kata/ Kalimat	FRE	F-K GL
1	Traktor Mendapat Bantuan	A	53	15	3,1	11	12,4
2	Aku Cinta Planetku	B1	177	22	7,9	0	16,3
3	Siapa Kau	B2	617	91	6,6	9,2	13,6

ANALISIS KETERBACAAN TEKS BAHASA INDONESIA DENGAN
PROGRAM MICROSOFT WORD

Juliana Tirza, Hananto

4	Malam di Musim Dingin	B3	1,935	206	9,3	0	15,7
5	Sang Inspektur Garam	C	2,433	246	9,8	0	17,3

Kalau dilihat dari jumlah kata dan kalimat, pelevelan/level A sampai C sudah sesuai dalam arti ada urutan jumlah kata dan kalimat dari yang kecil/sedikit menjadi semakin besar/banyak, tetapi kalau dilihat dari *Words per Sentence* (rata-rata jumlah kata dalam kalimat), ada ketidaksesuaian antara level B-1 dan B-2, yaitu skor level B-1 (7,9) lebih tinggi dibandingkan level B-2 (6,6). Menurut teorinya, kalimat yang sederhana/pendek akan mudah dipahami bila dibandingkan kalimat yang kompleks/panjang. Jadi, seharusnya rata-rata jumlah kata per kalimat di B1 lebih rendah bila dibandingkan yang di B-2, bukan lebih tinggi.

Hasil perhitungan fi kolom FRE (*Flesch Reading Ease*) di Tabel 1 juga bermasalah. Di kolom FRE terdapat tiga perhitungan yang hasilnya 0 (di level B1, B3, dan C). Kalau dilakukan perhitungan secara manual, hasilnya minus (kurang dari nol). Jadi kalau tiga buku hasilnya 0 maka tidak tampak buku mana yang lebih sulit. Walaupun demikian, dari perhitungan FRE terlihat adanya masalah dengan pelevelan buku B2 dengan hasil 9,2 sedangkan buku B1 hasilnya nol yang artinya buku level B2 lebih mudah dari buku level B1. Hasil ini tidak sesuai dengan teorinya di mana seharusnya hasil FRE B1 lebih besar dari yang di B2.

Kolom terakhir (F-K GL) menunjukkan masalah yang lebih serius. Seharusnya hasil perhitungannya adalah urut dari kecil (mudah) ke yang lebih besar (sulit) seperti terlihat dalam Tabel 3. Berdasarkan perhitungan F-K GL, ada 3 buku (B1, B2, dan B3) yang urutannya salah.

Tabel 2: Perbandingan urutan hasil yang didapat dengan yg Ideal/seharusnya

No	Buku	Level	F-K GL	
			Hasil	Ideal
1	Traktor Mendapat Bantuan	A	12,4	12,4
2	Aku Cinta Planetku	B1	16,3	13,6
3	Siapa Kau	B2	13,6	15,7
4	Malam di Musim Dingin	B3	15,7	16,3
5	Sang Inspektur Garam	C	17,3	17,3

Perhitungan *Keterbacaan Statistics* dengan formula FRE dan/atau F-K GL tidak cocok digunakan untuk teks berbahasa Indonesia karena formula tersebut berdasarkan penelitian dengan teks bahasa Inggris sehingga menghasilkan *grade level* yang terlalu tinggi. Misalnya buku dengan level C ditujukan untuk anak usia 10-13 tahun (anak SD-SMP), tetapi menurut formula F-K GL buku itu cocok untuk anak dengan grade level 17 (mahasiswa).

Sangat disayangkan keterbacaan Statistics di MS Word tidak mencantumkan jumlah suku kata. Perhitungan jumlah suku kata biasanya dihitung berdasarkan jumlah bunyi hidup (*vowel*: a, e, i, o, u). Dua *vowel* yg berurutan (misalnya “oo” dan “ee”) akan dihitung sebagai 1 suku kata.

Kombinasi gabungan dua *vowel* yg berurutan dalam bahasa Inggris adalah “ee”, “oo”, “aa”, “ai”, “ou”, “ea”, “io”, “ea”, “ui”, “ae”. Kombinasi gabungan dua *vowel* yg berurutan dalam bahasa Indonesia adalah “ai”, “au”, “oi”, “ei” seperti dalam kata “santai”, “kerbau”, “amboi”, “survei”. Gabungan 2 *vowel* berurutan dalam bahasa Inggris dan bahasa Indonesia berbeda sehingga akan menghasilkan salah perhitungan yang cukup besar. Misalnya kalimat “Kau mau harimau?” di MS Word mungkin akan dihitung sebagai 8 suku kata, sedangkan sebenarnya hanya 5 suku kata.

Walaupun perbedaan perhitungan rata-rata jumlah suku kata dalam sebuah kata itu kecil, hal ini akan menimbulkan perbedaan perhitungan yang cukup besar karena dikalikan dengan 11.8 (lihat rumus skor untuk *Flesch-Kincaid Grade Level* (.39 x ASL) + (11.8 x ASW) – 15.59).

Kesimpulan

Walaupun formula FRE dan F-K GL tidak cocok digunakan untuk menentukan pelevelan buku teks berbahasa Indonesia, formula ini tetap dapat digunakan untuk melihat/menentukan teks mana yang lebih mudah/lebih sulit. Dari perhitungan berdasarkan kedua formula itu, maka dapat disimpulkan bahwa cukup banyak pelevelan buku oleh Kemendikbudristek yang dilakukan secara subjektif kurang dapat diandalkan. Dari lima sampel buku dalam penelitian ini, berdasarkan formula F-K GL, terdapat tiga buku atau 60% pelevelan yang tidak sesuai.

Saran/Rekomendasi

Dalam menentukan pelevelan buku, Kemendikbudristek sebaiknya juga menggunakan atau memperhatikan perhitungan statistik keterbacaan yang bersifat objektif. Penggabungan pendekatan subjektif dan objektif diharapkan akan menghasilkan pelevelan buku yang lebih baik sehingga anak akan dapat membaca buku yang benar-benar sesuai dengan level mereka sehingga mereka dapat tumbuh minat baca mereka.

Teknik atau cara untuk melakukan perhitungan statistik keterbacaan memang cukup sulit kalau dilakukan perhitungan suku kata, kata dan kalimat secara manual, tetapi dengan adanya program komputer, maka hal itu menjadi mudah dilakukan.

Demikian juga dengan konversi buku gambar menjadi teks akan cukup sulit kalau dilakukan dengan menulis/mengetik ulang teks buku tersebut, terutama kalau teksnya panjang. Terdapat program komputer yang dapat mempercepat/membantu konversi mencari gambar menjadi teks, misalnya dengan memanfaatkan program OCR (Optical Character Reader) atau Speech-to-Text yang dapat menuliskan isi buku yang dibacakan.

DAFTAR PUSTAKA

- Besharati, M. R. (2021). *DASTEX: a New Keterbacaan Formula based on Semantic Complexity of Text. March*. <https://doi.org/10.20944/preprints202103.0582.v1>
- Corley, C., & Mihalcea, R. (2005). Measuring the semantic similarity of texts. *EMSEE*

2005 - *Empirical Modeling of Semantic Equivalence and Entailment@ACL 2005, Proceedings of the Workshop, June, 13–18.*
<https://doi.org/10.3115/1631862.1631865>

Eslami, H. (2014). The effect of syntactic simplicity and complexity on the keterbacaan of the text. *Journal of Language Teaching and Research, 5*(5), 1185–1191. <https://doi.org/10.4304/jltr.5.5.1185-1191>

Funke, R., & Sieger, J. (2012). Continued access to syntactic information in reading. *Reading and Writing, 25*(7), 1769–1794. <https://doi.org/10.1007/s11145-011-9341-8>

Giles, T. D., & Still, B. (2005). A syntactic approach to keterbacaan. *Journal of Technical Writing and Communication, 35*(1), 47–70. <https://doi.org/10.2190/PHUC-GY8L-JRLE-VMNN>

Hello English. (2022). *Siapa Kau*. Badan Pengembangan dan Pembinaan Bahasa.

Kalugina, N., Larina, E., Garkusha, N., & Kuznetsova, D. (2021). The development of semantic reading among students of a defectological profile in the digitalization of education. *SHS Web of Conferences, 101*, 03032. <https://doi.org/10.1051/shsconf/202110103032>

Kemdikbud. (2023). *Pelevelan Buku untuk Meningkatkan Kecintaan Membaca*. <https://www.kemdikbud.go.id/main/blog/2023/07/pelevelan-buku-untuk-meningkatkan-kecintaan-membaca>

Marutla, M. (2022). *Traktor Mendapat Bantuan*. Badan Pengembangan dan Pembinaan Bahasa.

Munshi Premchand. (2022a). *Malam di Musim Dingin*. Badan Pengembangan dan Pembinaan Bahasa.

Munshi Premchand. (2022b). *Sang Inspektur Garam*. Badan Pengembangan dan Pembinaan Bahasa.

Pitler, E., & Nenkova, A. (2008). Revisiting keterbacaan: A unified framework for predicting text quality. *EMNLP 2008 - 2008 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference: A Meeting of SIGDAT, a Special Interest Group of the ACL, October, 186–195.*

- Rasinski, T. V. (1994). Developing Syntactic Sensitivity in Reading Through Phrased Texts. *Intervention in School and Clinic*, 29(3), 165–168. <https://doi.org/10.1177/105345129402900307>
- Reinchenbach, P. (2022). *Aku Cinta Planetku.pdf*. Badan Pengembangan dan Pembinaan Bahasa.
- Salman, M., Haller, A., & Méndez, S. J. R. (2023). *Syntactic Complexity Identification, Measurement, and Reduction Through Controlled Syntactic Simplification*. 1–12. <http://arxiv.org/abs/2304.07774>
- Vakulenko, M. O. (2023). Semantic comparison of texts by the metric approach. *Digital Scholarship in the Humanities*, 38(2), 766–771. <https://doi.org/10.1093/lc/fqac059>
- Varlamova, E., & Safonkina, O. (2020). Keterbacaan of syntactic constructions for transparency and reliability of input texts for Unified State exams in English. In S. Hidri (Ed.), *Changing Language Assessment: New Dimensions, New Challenges* (pp. 181–193). Springer International Publishing. https://doi.org/10.1007/978-3-030-42269-1_8