**COGNITIVE SCIENCES AND HUMAN DEVELOPMENT**

# Using Entropy to Measure Text Readability in Bahasa Malaysia for Year One Students

**Mohamad Hardyman Barawi\*[1], Siti Nabilah Mohamed Osman[1], Noor Fazilla Abd Yusof[2], Ebuka Ibeke[3] & Muhibuddin Fadhli[4]**

[1]Faculty of Cognitive Sciences and Human Development, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia
[2]Faculty of Information Communication Technology, Universiti Teknikal Malaysia Melaka, 76100 Durian Tunggal, Melaka, Malaysia
[3]School of Creative and Cultural Business, Robert Gordon University, AB10 7AQ, United Kingdom
[4]Education Technology Department, Faculty of Education, Universitas Negeri Malang, 65145 East Java, Indonesia

## ABSTRACT

Text readability is essential for effective learning and communication, especially for beginner readers. However, there are no known measures to calculate the readability of Bahasa Malaysia, the national language of Malaysia. This research proposes a new method based on entropy, a measure of information and uncertainty, to assess the readability of Bahasa Malaysia texts for Year One students. An experiment was conducted with six Year One students to determine the relationship between entropy and readability. The results indicated a positive correlation, suggesting that higher entropy values corresponded with lower readability for this age group. This study also revealed the need for beginner readers to focus on the text difficulty level to enhance learning.

**Keywords:** readability, reading, text analysis, text difficulty

# 1    INTRODUCTION

One of the ways humans gain information about the world is through reading. Reading has been described as a combination of skills of decoding information from written materials into a mental representation of letters and words (Just & Carpenter, 1980). The relations between components of texts have led to the revision of text readability measures. The primary purpose is to make texts more accessible (McNamara & Kintsch, 1996). Individuals with low literacy skills struggle to comprehend written material (Snowling, 2013). At their initial stages of achievements, they focus more on processing the meaning rather than the form of the written texts (Farrokhi et al., 2008).

A text with more frequent words and shorter sentences is more appropriate for beginner readers (Crossley et al., 2007a). In Malaysia, most teachers rely on textbooks to convey information and as instructional guides in teaching. Thus, textbooks play an essential role in the education of students (Ball & Cohen, 1996; Kulm et al., 1999; Rockinson-Szapkiw et al., 2013). The definition of a textbook, as put forth by Brammer (1967), describes a textbook as a book that contains established principles in a specific subject and is primarily designed for use in classroom instruction or student-book-teacher scenarios. Given the widespread use of textbooks in the classroom, it is essential to carefully select textbooks to ensure they are effective for all students (Bruhn & Hasselbring, 2013). A good textbook must meet benchmarks and standards (Kulm et al., 1999). This helps ensure that the textbook's material is relevant, appropriate, and high-quality for students. Educators can create a positive learning environment by selecting textbooks that meet these criteria and help students achieve their full potential.

According to Chall's Stages of Reading Development (Chall, 1983), children aged seven to eight years are at the second stage, the confirmation and fluency stage. At this stage, children can only read simple and familiar texts. To help children develop strong reading skills by the end of this stage, it is essential to focus on building their basic decoding abilities, expanding their sight vocabulary, and enhancing their understanding of context when reading familiar stories. A solid foundation in these elements is essential for compelling reading and will empower children to become confident and skilled readers. By focusing on these critical areas, educators and parents can support children in developing the skills they need to become successful readers. At the end of this stage, children should be able to read about 3000 words and understand about 9000 words when listening (Chall, 1983).

Despite the problems and criticisms levelled at conventional readability formulas, they remain popular and are still used in research. However, significant technological advances in the last two decades have enabled the streamlining and automation of traditional readability formulas and developed more modern methods for measuring text difficulty (Crossley et al., 2019). In addition, researchers, administrators, and policymakers in the education field may require guidance on which methods are helpful in research studies and classrooms.

Since the rapid introduction of newly designed approaches for analysing text difficulty and matching readers and texts, educational researchers may become overwhelmed by the numerous options or be tempted to stick with the conventional methods used for decades. In contrast,